



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÕES



PÁTRIA AMADA
BRASIL
GOVERNO FEDERAL

sid.inpe.br/mtc-m21c/2020/08.12.17.37-TDI

PREDIÇÃO DE PROPRIEDADES MOLECULARES VIA APRENDIZADO DE MÁQUINA E REPRESENTAÇÃO SMILES

Gabriel Augusto Lins Leal Pinheiro

Dissertação de Mestrado do
Curso de Pós-Graduação em
Computação Aplicada, orientada
pelos Drs. Marcos Gonçalves
Quiles e Marinalva Dias Soares,
aprovada em 29 de julho de 2020.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34R/433H4PS>>

INPE
São José dos Campos
2020

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GBDIR)

Serviço de Informação e Documentação (SESID)

CEP 12.227-010

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/7348

E-mail: pubtc@inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE - CEPPII (PORTARIA Nº 176/2018/SEI-INPE):**Presidente:**

Dra. Marley Cavalcante de Lima Moscati - Centro de Previsão de Tempo e Estudos Climáticos (CGCPT)

Membros:

Dra. Carina Barros Mello - Coordenação de Laboratórios Associados (COCTE)

Dr. Alisson Dal Lago - Coordenação-Geral de Ciências Espaciais e Atmosféricas (CGCEA)

Dr. Evandro Albiach Branco - Centro de Ciência do Sistema Terrestre (COCST)

Dr. Evandro Marconi Rocco - Coordenação-Geral de Engenharia e Tecnologia Espacial (CGETE)

Dr. Hermann Johann Heinrich Kux - Coordenação-Geral de Observação da Terra (CGOBT)

Dra. Ieda Del Arco Sanches - Conselho de Pós-Graduação - (CPG)

Silvia Castro Marcelino - Serviço de Informação e Documentação (SESID)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon

Clayton Martins Pereira - Serviço de Informação e Documentação (SESID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Simone Angélica Del Ducca Barbedo - Serviço de Informação e Documentação (SESID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SESID)

EDITORAÇÃO ELETRÔNICA:

Ivone Martins - Serviço de Informação e Documentação (SESID)

Cauê Silva Fróes - Serviço de Informação e Documentação (SESID)



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÕES



PÁTRIA AMADA
BRASIL
GOVERNO FEDERAL

sid.inpe.br/mtc-m21c/2020/08.12.17.37-TDI

PREDIÇÃO DE PROPRIEDADES MOLECULARES VIA APRENDIZADO DE MÁQUINA E REPRESENTAÇÃO SMILES

Gabriel Augusto Lins Leal Pinheiro

Dissertação de Mestrado do
Curso de Pós-Graduação em
Computação Aplicada, orientada
pelos Drs. Marcos Gonçalves
Quiles e Marinalva Dias Soares,
aprovada em 29 de julho de 2020.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34R/433H4PS>>

INPE
São José dos Campos
2020

Dados Internacionais de Catalogação na Publicação (CIP)

Pinheiro, Gabriel Augusto Lins Leal.

P655p Predição de propriedades moleculares via aprendizado de máquina e representação SMILES / Gabriel Augusto Lins Leal Pinheiro. – São José dos Campos : INPE, 2020.
xx + 167 p. ; (sid.inpe.br/mtc-m21c/2020/08.12.17.37-TDI)

Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2020.

Orientadores : Drs. Marcos Gonçalves Quiles e Marinalva Dias Soares.

1. Aprendizado de máquina. 2. SMILES. 3. Redes neurais.
4. Ciência dos materiais. 5. Propriedades moleculares. I.Título.

CDU 004.85



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](#).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](#).

Aluno (a): **Gabriel Augusto Lins Leal Pinheiro**

Título: "PREDIÇÃO DE PROPRIEDADES MOLECULARES VIA APRENDIZADO DE MÁQUINA E REPRESENTAÇÃO SMILES"

Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de **Mestre** em

Computação Aplicada

Dr. Nandamudi Lankalapalli Vijaykumar

N L Vijaykumar

Presidente / INPE / SJC Campos - SP

☒ Participação por Video - Conferência

☒ Aprovado ☐ Reprovado

Dr. Marcos Gonçalves Quiles

N L Vijaykumar

Orientador(a) / UNIFESP / São José dos Campos - SP

☒ Participação por Video - Conferência

☒ Aprovado ☐ Reprovado

Dra. Marinalva Dias Soares

N L Vijaykumar

Orientador(a) / INPE / São José dos Campos - SP

☒ Participação por Video - Conferência

☒ Aprovado ☐ Reprovado

Dr. Rafael Duarte Coelho dos Santos

N L Vijaykumar

Membro da Banca / INPE / SJC Campos - SP

☒ Participação por Video - Conferência

☒ Aprovado ☐ Reprovado

Dr. Eudes Eterno Fileti

N L Vijaykumar

Convidado(a) / UNIFESP / São José dos Campos - SP

☒ Participação por Video - Conferência

☒ Aprovado ☐ Reprovado

Dr. Ronaldo Cristiano Prati

N L Vijaykumar

Convidado(a) / UFABC / Santo André - SP

☒ Participação por Video - Conferência

☒ Aprovado ☐ Reprovado

Este trabalho foi aprovado por:

☐ maioria simples

☒ unanimidade

São José dos Campos, 29 de julho de 2020

AGRADECIMENTOS

Primeiramente, agradeço a Deus pelo dom da vida. Agradeço aos meus pais (Roberto Carlos e Jurema Lins) e irmãos (Maria Teresa, Tiago Francisco, Daniel José e Rafael Alberto) pelo incentivo, encorajamento e apoio incondicional. Ao meu orientador, Prof. Dr. Marcos Gonçalves Quiles, pelo apoio, incentivo, disponibilidade, amizade e condução nesta minha trajetória pela pesquisa. À minha coorientadora, Dra. Marinalva Dias Soares, pela orientação, as boas conversas e risadas. Ao Prof. Dr. Juarez Lopes Ferreira da Silva e seu aluno Johnatan Mucelini, pela colaboração, apoio e supervisão. A todos os meus amigos que me deram suporte, em especial, Layla dos Santos, Pedro dos Santos e o Wu Chun. Institucionalmente, agradeço ao Laboratório Associado de Computação e Matemática Aplicada (LAC) por disponibilizar a infraestrutura utilizada durante a minha pesquisa. Ao CNPq pelo apoio financeiro.

RESUMO

O campo de ciência dos materiais possui caráter interdisciplinar, com aplicações nas mais diversas áreas do conhecimento que acabam influenciando vários aspectos da vida cotidiana. Pode-se citar como exemplo de aplicações pesquisas relacionadas à energia renovável, saúde, engenharia mecânica, aviação e espaço. Nesta área da ciência, o estudo das propriedades dos materiais e as relações com suas estruturas em escala atômica ou molecular torna-se imprescindível para a descoberta de novos materiais, requerendo assim, o conhecimento de como estes se comportam e reagem uns com os outros em diferentes ambientes. Na área espacial, por exemplo, o desenvolvimento de nanotecnologias para a composição das estruturas de um satélite pode permitir a descoberta de materiais mais leves e resistentes a danos de radiação. Contudo, abordagens tradicionais em ciência dos materiais, como a teoria do funcional da densidade, usadas no processo do cálculo das propriedades físico-químicas de um material, são computacionalmente custosas. Por outro lado, a quantidade de dados providos sobre cálculos e experimentos realizados na última década possibilitou a aplicação de algoritmos de aprendizado de máquina como alternativa aos métodos tradicionais, tendo por vantagem a otimização no tempo do cálculo. Para que o algoritmo de aprendizado de máquina realize um mapeamento adequado entre a entrada e saída, é necessário fazer a descrição da molécula, chamado de descritor ou atributo, de maneira a codificar informações significativas presentes no sistema molecular. Neste sentido, este trabalho visou explorar o uso de descritores baseados na representação *Simplified Molecular Input Line Entry System* (SMILES) para a predição de propriedades moleculares. Para isso, utilizou-se uma rede Perceptron de múltiplas camadas e cinco métodos de seleção de atributos. As melhores predições foram obtidas para as propriedades térmicas, que atingiram um erro absoluto médio próximo a 0.05 eV e portanto, bem próximo ao erro aceitável pela acurácia química.

Palavras-chave: Aprendizado de máquina. SMILES. Redes neurais. Ciência dos materiais. Propriedades moleculares.

MACHINE LEARNING PREDICTION OF MOLECULAR PROPERTIES BASED ON THE SMILES REPRESENTATION

ABSTRACT

The field of material science is characterized by its interdisciplinary in several scientific areas, which influences diverse aspects of daily life. For instance, one can cite material science applications in clean energy, health, mechanical engineer, and aircraft materials. In this realm, the study of materials properties and its relationship with the molecular structure at the atomic or molecular scale is crucial for the development of new materials, requiring an understanding of the behavior of the material to different environments and other materials. In aerospace applications, for instance, the development of nanotechnology for satellites structure can achieve materials more resistant to spatial radiation and weightless. Nevertheless, traditional approaches in material science (i.e., density functional theory) used to compute the physicochemical properties demand high computational resources. By contrast, the amount of available data provided from previous calculations and experiments in the last decade has allowed the application of a machine learning algorithm as an alternative approach for such computations with less time and computational resources. However, machine learning algorithms rely on a well suitable representation of the data to perform an accurate prediction. Thus, molecular features, or descriptors, are essential to encode meaningful information present in the molecules. In this sense, this work aims to explore molecular descriptors based in the *Simplified Molecular Input Line Entry System* (SMILES) representation to predict molecular properties. For that, it was used a multilayer perceptron and five feature selection methods. The best achievements in this work obtained an error, in terms of mean absolute error, of 0.05 eV for the thermodynamic properties, which is an error close to the chemical accuracy.

Keywords: Machine learning. SMILES. Neural networks. Materials science. Molecular properties.

LISTA DE FIGURAS

	<u>Pág.</u>
2.1 Diagrama das fases da descoberta de conhecimento em base de dados. Vale ressaltar que tais fases não necessariamente são executadas de maneira linear.	10
2.2 Interpretação da árvore de decisão.	15
2.3 Redução da variância usando métodos <i>ensemble</i> . (a), (b) e (c) ilustram a fronteira de decisão de 3 modelos bases. (d) A combinação dos resultados de cada modelo base. As linhas tracejadas representam as fronteiras de decisão dos modelos (a), (b) e (c), sendo que a linha vermelha representa o modelo (a), azul o modelo (b) e verde o modelo (c). (e) Fronteira de decisão do método <i>ensemble</i>	16
2.4 Exemplo do método <i>boosting</i> para um problema de classificação binária. (a), (b) e (c) ilustram a fronteira de decisão de 3 modelos bases. Os pesos das amostras foram ilustrados usando pontos de diferentes tamanhos. (d) Fronteira de decisão do método <i>boosting</i>	18
2.5 Método <i>bagging</i> versus RF. (b) A área destaca em verde representa a etapa em que seleção aleatória dos m atributos ocorre.	20
2.6 SQ dos erros em dado os parâmetros β_0 e β_1	22
2.7 Anatomia de um neurônio.	24
2.8 Neurônio artificial.	26
2.9 $\hat{y} = \varphi(x_1 w_1 + b)$ para $w_1 = 0.6$, $b = 6.2$ e $\varphi(v) = v$	27
2.10 Função de limiar.	28
2.11 Função sigmoide.	28
2.12 Função ReLU.	29
2.13 Rede Perceptron de múltiplas camadas.	30
2.14 Representação gráfica do ajuste de curva do polinômio para diferentes ordens. (a) ajuste de uma reta. (b) ajuste de uma função quadrática. (c) ajuste de uma função polinomial do nono grau.	35
2.15 Representação gráfica do <i>underfitting</i> e <i>overfitting</i> por meio do erro de treino e teste.	36
2.16 Divisão do conjunto de treino em k partições, no qual a partição com a cor verde ilustra o conjunto de validação na k -ésima iteração e as partições em branco, o conjunto de treino.	38

2.17	Exemplo de atributo relevante, redundante e irrelevante. (a) x_1 é considerado como atributo relevante já que consegue discriminar bem as duas classes do problema. (b) x_2 é dito como atributo redundante devido a correlação alta com x_1 . (c) x_3 é categorizado como atributo irrelevante já que a separação das classes é feita completamente pelo x_1	39
2.18	Exemplo de espaço de soluções de atributos considerando $d = 4$. Uma seleção <i>forward</i> é ilustrada percorrendo o diagrama da esquerda para a direita, um percurso ao contrário do citado anteriormente irá resultar em uma eliminação <i>backward</i>	42
2.19	Taxonomia dos critérios de importância.	44
2.20	Exemplos de casos extremos da correlação de Pearson.	47
2.21	Abordagem Filtro.	48
2.22	Abordagem <i>Wrapper</i>	49
2.23	Abordagem <i>Embedding</i>	49
2.24	Exemplos de atributo redundante e irrelevante. (a) x_2 é considerado como atributo redundante já que consegue discriminar o mesmo <i>cluster</i> de x_1 . (b) x_3 é dito como atributo irrelevante já que a sua eliminação não modifica a discriminação dos <i>clusters</i>	51
2.25	Comportamento da métrica $M_{C'}$ para diferentes valores de $\overline{corr}_{x,y}$, $\overline{corr}_{x,x}$ e k	53
2.26	Exemplos de relações monotônicas e não monotônicas.	54
2.27	Exemplos do cálculo para correlação de Spearman.	55
3.1	Os 4 paradigmas em ciência dos materiais.	61
3.2	Linha tracejada apresenta a função original que governa a propriedade. Os pontos vermelhos são os resultados obtidos por simulações quânticas (DFT, por exemplo). Linha azul mostra a função aproximada encontrada pelo algoritmo de AM. Note que este exemplo é um caso de <i>underfitting</i> , já que as amostras não descrevem adequadamente a função original do problema.	61
3.3	AM em ciência dos materiais para a predição de propriedade e geometria molecular.	62
3.4	No processo de codificação do grafo da molécula para SMILES foram omitidos os átomos de hidrogênio e as ligações simples.	63
3.5	Representação das informações sobre a distribuição espacial da molécula de água no formato XYZ. Primeira e segunda linha apresentam a quantidade de átomos e nome do composto, respectivamente.	64
3.6	Atributo <i>fingerprint</i> com fragmentos definidos manualmente.	67
3.7	Matriz de Coulomb.	68

4.1	Fluxo de processos para extração de conhecimento com AM.	74
4.2	Módulo sobre entendimento do problema.	75
4.3	Módulo sobre entendimento do problema.	76
4.4	Histograma do número de moléculas em função dos tipo de átomos presentes. Note que existe cerca de 67 mil moléculas contento os quatro diferentes tipos de átomos C, H, N e O.	77
4.5	Histograma do número de moléculas por quantidade de átomos.	77
4.6	Dispersão dos valores por quantidade de átomos para a propriedade constante rotacional A. Pontos em preto representa a média em para cada tamanho de moléculas.	78
4.7	Dispersão dos valores por quantidade de átomos para a propriedade constante rotacional B. Pontos em preto representa a média em para cada tamanho de moléculas.	79
4.8	Dispersão dos valores por quantidade de átomos para a propriedade constante rotacional C. Pontos em preto representa a média em para cada tamanho de moléculas.	79
4.9	Dispersão dos valores por quantidade de átomos para a propriedade μ . Pontos em preto representa a média em para cada tamanho de moléculas.	80
4.10	Dispersão dos valores por quantidade de átomos para a propriedade α . Pontos em preto representa a média em para cada tamanho de moléculas.	80
4.11	Dispersão dos valores por quantidade de átomos para a propriedade ϵ_{HOMO} . Pontos em preto representa a média em para cada tamanho de moléculas.	81
4.12	Dispersão dos valores por quantidade de átomos para a propriedade ϵ_{LUMO} . Pontos em preto representa a média em para cada tamanho de moléculas.	81
4.13	Dispersão dos valores por quantidade de átomos para a propriedade Δ_e . Pontos em preto representa a média em para cada tamanho de moléculas.	82
4.14	Dispersão dos valores por quantidade de átomos para a propriedade $\langle R^2 \rangle$. Pontos em preto representa a média em para cada tamanho de moléculas.	82
4.15	Dispersão dos valores por quantidade de átomos para a propriedade zpve. Pontos em preto representa a média em para cada tamanho de moléculas.	83
4.16	Dispersão dos valores por quantidade de átomos para a propriedade U_0 . Pontos em preto representa a média em para cada tamanho de moléculas.	83
4.17	Dispersão dos valores por quantidade de átomos para a propriedade U . Pontos em preto representa a média em para cada tamanho de moléculas.	84
4.18	Dispersão dos valores por quantidade de átomos para a propriedade H . Pontos em preto representa a média em para cada tamanho de moléculas.	84

4.19	Dispersão dos valores por quantidade de átomos para a propriedade G . Pontos em preto representa a média em para cada tamanho de moléculas.	85
4.20	Dispersão dos valores por quantidade de átomos para a propriedade C_v . Pontos em preto representa a média em para cada tamanho de moléculas.	85
4.21	Moléculas com valores discrepantes para constante rotacional A	86
4.22	Exemplos de moléculas com múltiplas estruturas na representação SMILES.	87
4.23	Verificação da consistência dos SMILES.	88
4.24	Dispersão dos valores por quantidade de átomos para a propriedade atômização U_0 . Pontos em preto representa a média em para cada tamanho de moléculas.	90
4.25	Dispersão dos valores por quantidade de átomos para a propriedade atômização U . Pontos em preto representa a média em para cada tamanho de moléculas.	90
4.26	Dispersão dos valores por quantidade de átomos para a propriedade atômização H . Pontos em preto representa a média em para cada tamanho de moléculas.	91
4.27	Dispersão dos valores por quantidade de átomos para a propriedade atômização G . Pontos em preto representa a média em para cada tamanho de moléculas.	91
4.28	Módulo sobre entendimento do problema.	92
4.29	Dispersão dos valores por quantidade de átomos para a propriedade μ . Pontos em cinza representa as moléculas removidas.	93
4.30	Dispersão dos valores por quantidade de átomos para a propriedade α . Pontos em cinza representa as moléculas removidas.	93
4.31	Dispersão dos valores por quantidade de átomos para a propriedade ϵ_{HOMO} . Pontos em cinza representa as moléculas removidas.	94
5.1	Correlação de Pearson entre os pares de descritores apresentados na Tabela 5.1.	105
5.2	Cálculo do ES para a molécula de Butanol considerando fragmentos formados por vizinhos de ordem 1.	109
5.3	<i>Scatter plot</i> do valor real contra o valor predito para o conjunto de teste, no qual r é o coeficiente da correlação de Pearson entre ambas variáveis.	114

LISTA DE TABELAS

	<u>Pág.</u>
2.1 Tabela formato atributo-valor.	8
2.2 Combinações das técnicas de organização de busca com direção de busca.	43
4.1 Propriedades e suas respectivas unidades do conjunto QM9.	73
4.2 Estatística descritiva básica sobre o QM9, no qual σ representa o desvio padrão e Min e Max denomina os valores mínimo e máximo, respectivamente.	87
4.3 Quantidade média de descritores selecionados por conjunto de dados. A média foi calculada sobre as seleções por propriedade. Conjunto 17, 18 e 19 representa os subconjuntos composto por moléculas com 17, 18 e 19 átomos, respectivamente e # denomina quantidade.	101
4.4 População de valores definidos para ajustar os hiperparâmetros das RNAs.	102
5.1 Descritores mais selecionados entre todos os métodos de SA e propriedades preditas do conjunto QM9. O número entre parêntese representa a quantidade de ocorrência para cada atributo.	103
5.2 Descritores mais selecionados entre os subconjuntos com moléculas de 17, 18 e 19 átomos.	104
5.3 Restrições no somatório do atributo PEOE_VSA _k para cada condição k	106
5.4 Restrições no somatório do atributo SlogP_VSA _k para cada condição k	106
5.5 Restrições no somatório do atributo EState_VSA _k para cada condição k	107
5.6 Restrições no somatório do atributo VSA_EState _k para cada condição k	107
5.7 Restrições no somatório do atributo SMR_VSA _k para cada condição k	107
5.8 Erro de generalização, em termos de EAM, sobre o conjunto QM9. Nenhum significa que todos os descritores foram usados como entrada.	110
5.9 Erro de generalização, em termos de EAM, sobre o subconjunto com moléculas de 17 átomos. Nenhum significa que todos os descritores foram usados como entrada.	111
5.10 Erro de generalização, em termos de EAM, sobre o subconjunto com moléculas de 18 átomos. Nenhum significa que todos os descritores foram usados como entrada.	112
5.11 Erro de generalização, em termos de EAM, sobre o subconjunto com moléculas de 19 átomos. Nenhum significa que todos os descritores foram usados como entrada.	113

5.12 Erro de generalização no conjunto QM9 usando como entrada da rede os descriptores baseados no SMILES e a matriz de Coulomb.	113
---	-----

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
1.1 Objetivos	4
1.2 Organização do texto	4
2 APRENDIZADO DE MÁQUINA	7
2.1 Conceitos iniciais	7
2.1.1 Aprendizado supervisionado	8
2.1.2 Aprendizado não-supervisionado	9
2.2 Fluxo de processos em aprendizado de máquina	9
2.3 Árvore de decisão	12
2.4 Métodos <i>ensemble</i>	16
2.4.1 <i>Boosting</i>	17
2.4.2 <i>Bagging</i>	18
2.4.3 <i>Random forest</i>	19
2.5 <i>Least absolute shrinkage and selection operator</i>	20
2.6 Redes neurais artificiais	23
2.6.1 Neurônio	26
2.6.2 Termo viés	26
2.6.3 Função de ativação	27
2.6.4 Perceptron de múltiplas camadas	29
2.6.4.1 Propagação <i>feedforward</i>	30
2.6.4.2 Aprendizagem por correção de erro	31
2.7 Otimização	34
2.7.1 Parâmetros e hiperparâmetros	36
2.7.1.1 Validação cruzada	37
2.8 Extração e seleção de atributos	38
2.9 Seleção de atributos	39
2.9.1 Direção de busca	41
2.9.2 Organização de busca	42
2.9.3 Critérios de avaliação de importância	44
2.9.4 Critério de parada	45
2.9.5 Abordagens para seleção de atributos	46

2.9.5.1	Abordagem filtro	46
2.9.5.2	Abordagem <i>wrapper</i>	47
2.9.5.3	Abordagem <i>embedding</i>	48
2.10	Paradigmas em seleção de atributos	50
2.11	<i>Correlation-based feature selection</i>	52
2.12	Filtro por correlação	53
2.13	<i>Random forest</i> para seleção de atributos	56
2.14	<i>Least absolute shrinkage and selection operator</i> para seleção de atributos	56
3	APRENDIZADO DE MÁQUINA E CIÊNCIA DOS MATERIAIS	59
3.1	Conceitos iniciais	59
3.2	Estrutura molecular	62
3.3	Descritores	65
3.3.1	Descritores constitucionais	65
3.3.2	Descritores topológicos	66
3.3.3	Descritores geométricos	67
3.4	Estado da arte	67
4	MATERIAIS E MÉTODOS	73
4.1	Entendimento do problema	74
4.1.1	Definição do problema	74
4.1.2	Estratégias de aprendizado de máquina	75
4.1.3	Reconhecimento das fontes de dados	75
4.2	Coleta dos dados	76
4.2.1	Determinação do conjunto de dados	76
4.2.2	Análise exploratória dos dados	78
4.2.2.1	Seleção das propriedades para predição	89
4.3	Preparação dos dados	92
4.3.1	Pré-processamento dos dados	92
4.3.2	Extração de atributos	95
4.3.3	Seleção de atributos	99
4.3.4	Normalização dos dados	100
4.4	Modelagem	101
4.5	Avaliação	102
5	RESULTADOS E DISCUSSÕES	103
5.1	Seleção de atributos	103
5.1.1	Análise dos descritores	105

5.2	Predição das propriedades	110
6	CONCLUSÃO	115
	REFERÊNCIAS BIBLIOGRÁFICAS	117
	APÊNDICE A - DESCRITORES SELECIONADOS.	131
A.1	Subconjunto com moléculas de 17 átomos	131
A.1.1	Energia interna de atomização para a temperatura de 0 K	131
A.1.2	Energia interna de atomização para a temperatura de 298.15 K	132
A.1.3	Entalpia de atomização para a temperatura de 298.15 K	133
A.1.4	Energia livre de Gibbs de atomização para a temperatura de 298.15 K	134
A.1.5	Capacidade térmica para a temperatura de 298.15 K	135
A.1.6	Momento dipolar	136
A.1.7	Polarizabilidade isotrópica	137
A.1.8	Energia HOMO	139
A.1.9	Diferença de energias do HOMO e LUMO	140
A.2	Subconjunto com moléculas de 18 átomos	141
A.2.1	Energia interna de atomização para a temperatura de 0 K	141
A.2.2	Energia interna de atomização para a temperatura de 298.15 K	142
A.2.3	Entalpia de atomização para a temperatura de 298.15 K	143
A.2.4	Energia livre de Gibbs de atomização para a temperatura de 298.15 K	144
A.2.5	Capacidade térmica para a temperatura de 298.15 K	145
A.2.6	Momento dipolar	146
A.2.7	Polarizabilidade isotrópica	147
A.2.8	Energia HOMO	148
A.2.9	Diferença de energias do HOMO e LUMO	149
A.3	Subconjunto com moléculas de 19 átomos	150
A.3.1	Energia interna de atomização para a temperatura de 0 K	150
A.3.2	Energia interna de atomização para a temperatura de 298.15 K	151
A.3.3	Entalpia de atomização para a temperatura de 298.15 K	152
A.3.4	Energia livre de Gibbs de atomização para a temperatura de 298.15 K	153
A.3.5	Capacidade térmica para a temperatura de 298.15 K	154
A.3.6	Momento dipolar	155
A.3.7	Polarizabilidade isotrópica	156
A.3.8	Energia HOMO	157
A.3.9	Diferença de energias do HOMO e LUMO	158
A.4	QM9	159

A.4.1	Energia interna de atomização para a temperatura de 0 K	159
A.4.2	Energia interna de atomização para a temperatura de 298.15 K	160
A.4.3	Entalpia de atomização para a temperatura de 298.15 K	161
A.4.4	Energia livre de Gibbs de atomização para a temperatura de 298.15 K .	162
A.4.5	Capacidade térmica para a temperatura de 298.15 K	163
A.4.6	Momento dipolar	164
A.4.7	Polarizabilidade isotrópica	165
A.4.8	Energia HOMO	166
A.4.9	Diferença de energias do HOMO e LUMO	167

1 INTRODUÇÃO

O constante crescimento da capacidade computacional alinhado com a massiva quantidade de dados e o avanço dos métodos da inteligência artificial contribuíram para a aplicação do aprendizado de máquina (AM) nas mais diversas áreas da ciência. Pode-se tomar como exemplo, a aplicação de AM na medicina (ESTEVA et al., 2019), sistemas de recomendação de propaganda (OORD et al., 2013), descoberta de medicamentos (EKINS et al., 2019; VAMATHEVAN et al., 2019), visão computacional (KRIZHEVSKY et al., 2012), ciência dos materiais (LO et al., 2018; JORDAN; MITCHELL, 2015), entre outros. Na ciência dos materiais, em particular, um novo paradigma em fazer ciência (AGRAWAL; CHOUDHARY, 2016), utiliza tais abordagens computacionais capazes de aprender, baseando-se em experiência, relações lineares ou não lineares entre os dados entrada e saída (SANCHEZ-LENGELING; ASPURUGUZZIK, 2018), para impulsionar a extração de conhecimento sobre bases de dados da química com o objetivo de acelerar a descoberta de novos materiais (WU et al., 2018; BISBO; HAMMER, 2019; KEARNES et al., 2016).

Sobretudo, algoritmos de AM têm apresentado erros dentro do limite estabelecido como acurácia química para as predições de propriedades moleculares e baixo custo computacional quando comparado com abordagens padrões da ciência dos materiais, como a teoria do funcional da densidade (do inglês *density functional theory* — DFT) e métodos baseados em funções de onda (RAMAKRISHNAN et al., 2015; FABER et al., 2017).

Em aplicações espaciais, o Instituto Nacional de Pesquisa Espaciais tem investido no desenvolvimento de nanotecnologias. Alguns dos estudos de interesse envolvem o aprimoramento das propriedades de supercapacitores. Por exemplo, Toledo (2013) realizou o estudo da aplicação de nanotubos de carbono em eletrodos com o objetivo de aumentar a densidade de energia e a vida útil de supercapacitores. De maneira similar, Moraes (2013) propôs a aplicação de nanoestruturas de compósitos ternários (dióxido de titânio / óxido de grafeno reduzido / fibra de carbono) devido às propriedades de condutividade e estabilidade química do material. De fato, o interesse pelo desenvolvimento de nanotecnologias pode ser justificado pela redução da massa dos satélites e consequentemente, o custo do lançamento. Nestes estudos, a aplicação de AM para a realização de tais descobertas apresenta ser uma área promissora devido à sua capacidade em acelerar os processos.

Neste sentido, um fator importante para que o sucesso dos modelos de AM seja atingido são os atributos, ou descritores, usados na descrição da molécula, já que é por

meio deles que o algoritmo deverá aprender uma função-conceito até a propriedade molecular. Por sua vez, o processo de extração de atributos é dependente da informação *a priori* a respeito da molécula (por exemplo, estrutura molecular, elementos químicos) e da maneira pela qual esta informação está organizada (por exemplo, textual ou numérica) (LO et al., 2018). Além do mais, pretende-se que os descritores sejam sempre relevantes para as propriedades físico-químicas de interesse, visto que dependendo do conjunto de atributos definido, até mesmo modelos lineares podem apresentar excelentes resultados. Contudo, definir os descritores para uma determinada propriedade não é uma tarefa trivial e por isso, tal processo é considerado como uma área de pesquisa em ciência dos materiais e AM (BUTLER et al., 2018).

Com efeito, Duvenaud et al. (2015) propôs uma rede neural convolucional sobre o grafo da molécula para realizar a extração de descritores *fingerprints* com dimensão fixa e de maneira automática para a predição de propriedades. Posteriormente, (SCHÜTT et al., 2018; SCHÜTT et al., 2017b) apresentou um modelo de rede neural convolucional de grafo considerando as posições dos átomos para extração de descritores automática e atingiu o estado da arte na predição de diversas propriedades. Neste mesmo segmento, diversos outros trabalhos propuseram a construção de descritores derivados das coordenadas moleculares, como a matriz de Coulomb (MONTAVON et al., 2013), o *bag of bonds* (HANSEN et al., 2015), a representação de Faber–Christensen–Huang–Lilienfeld (FCHL) (FABER et al., 2018) e as funções simétricas para descrição das estruturas moleculares por meio de funções de dois e três corpos considerando as características dos átomos (BEHLER; PARRINELLO, 2007).

Neste mesmo contexto, a coordenada cartesiana é a representação molecular mais utilizada, já que consegue capturar melhor as variações da estrutura molecular (LO et al., 2018) e é a mesma informação usada em métodos *ab initio* (MONTAVON et al., 2013). Entretanto, tal representação não codifica adequadamente as principais características da molécula e algoritmos de AM podem enfrentar dificuldades no processo de aprendizagem, caso a extração de atributos não seja bem feita (SANCHEZ-LENGELING; ASPURU-GUZIK, 2018). Além disso, quando a molécula é conhecida, a obtenção das coordenadas é feita por um método de otimização global que faz a iteração sobre um conjunto de possíveis coordenadas candidatas até que a molécula atinja o estado de menor energia. Neste processo, a quantidade de candidatos em mínimos locais aumenta exponencialmente com a quantidade de átomos.

Uma segunda representação bastante usada como formato padrão em diversos pro-

gramas da quimioinformática e conjunto de dados da química é o *Simplified Molecular Input Line Entry System* (SMILES) (GOH et al., 2017a). Diferente da representação anterior, o SMILES codifica a molécula por meio de uma sequência de caracteres que representam os átomos e suas ligações, similar a fórmula química, porém, mais completa. A gramática usada em SMILES é simples, o que permite ser facilmente interpretada e obtida através do grafo da molécula. Em AM, exemplos de aplicações de SMILES estão presentes no processo de criação de molécula, como o *de novo design* (GÓMEZ-BOMBARELLI et al., 2018; BJERRUM; SATTAROV, 2018); técnica de *data augmentation*, já que é possível obter diversos SMILES de uma mesma molécula (BJERRUM, 2017); e desenvolvimento de *fingerprints* (GLEN et al., 2006). Nesses estudos, o SMILES apresentou desempenho similar na predição de propriedades eletrônicas em comparação aos modelos de rede neural convolucional de grafo e uma melhora dos resultados de predição usando SMILES aleatórios em *data augmentation*.

Diante do exposto e com o objetivo de entender a aplicação de algoritmos de AM em representação SMILES, este trabalho irá investigar a utilização de descritores moleculares baseados em SMILES para a realização da predição de propriedades moleculares em um conjunto de dados público chamado QM9 (RAMAKRISHNAN et al., 2014). O QM9 é composto por 133885 moléculas orgânicas formadas por átomos de C, H, O, N e F, e disponibiliza 15 propriedades moleculares, entre elas propriedades geométricas, termodinâmicas, de energia e eletrônicas. Além disso, diversos estudos anteriores utilizaram a base de dados na predição de propriedades moleculares (STUKE et al., 2019; HUANG; LILIENFELD, 2016; JØRGENSEN et al., 2018; PRONOBIS et al., 2018).

No presente trabalho foi realizada a predição de nove propriedades do QM9, estas são: (i) energia interna de atomização para a temperatura de 0 K; (ii) energia interna de atomização para a temperatura de 298.15 K; (iii) entalpia de atomização para a temperatura de 298.15 K; (iv) energia livre de Gibbs de atomização para a temperatura de 298.15 K; (v) capacidade térmica para a temperatura de 298.15 K; (vi) energia do orbital molecular ocupado mais alto (do inglês *Highest Occupied Molecular Orbital* — HOMO); (vii) diferença de energias do HOMO e do orbital molecular mais baixo desocupado (do inglês *Lowest Unoccupied Molecular Orbital* — LUMO); (viii) momento dipolar; e (ix) polarizabilidade isotrópica. Para tanto, nas simulações foi utilizada uma rede Perceptron de múltiplas camadas, um vetor de características composto por 880 descritores baseados na representação SMILES e cinco abordagens para a seleção de atributos, este são: (i) método *correlation-based*

feature selection, (ii) seleção filtro por correlação usando a correlação de Pearson e Spearman, (iii) *random forest* e (iv) *least absolute shrinkage and selection operator*. A seleção de atributos teve por objetivo melhorar o desempenho da rede neural artificial e o entendimento dos descritores mais relevantes para as propriedades de interesse.

1.1 Objetivos

Tendo em vista os aspectos levantados anteriormente, o principal objetivo dessa dissertação foi explorar a representação de moléculas e algoritmos de AM para a predição de propriedades relevantes no contexto de ciência dos materiais. De modo específico, investigou-se diferentes tipos de descritores baseados na representação SMILES da molécula de modo a codificar diversas informações a respeito da molécula.

Neste estudo, foi realizada a aplicação de técnicas para seleção de atributos visando obter ganhos na acurácia, reduzir o tempo de aprendizado e eliminar atributos redundantes e irrelevantes, que tornam o processo de aprendizado complexo devido à quantidade de atributos. Além disso, nos experimentos foi utilizado a base de dados pública chamada QM9, que é comumente usada como *benchmarking* por reportar o cálculo de 15 propriedades moleculares para mais de 133 mil moléculas, assim como também, a estrutura da molécula (WU et al., 2018). Com base no QM9, foram definidos três subconjuntos, cada qual composto por grupos de moléculas com quantidade de átomo específico, com o objetivo de verificar se a homogeneidade imposta nos dados resulta em uma acurácia melhor do modelo.

1.2 Organização do texto

Essa dissertação está organizada da seguinte forma:

Capítulo 2: APRENDIZADO DE MÁQUINA, apresenta os principais fundamentos teóricos sobre AM e seus paradigmas. Posteriormente, é dada uma breve contextualização sobre redes neurais artificiais, *random forest* e *least absolute shrinkage and selection operator*. Também é abordado a otimização de parâmetros, representação de dados no processo de extração e métodos de seleção de atributos.

Capítulo 3: APRENDIZADO DE MÁQUINA E CIÊNCIA DOS MATERIAIS, apresenta uma breve contextualização sobre ciência dos materiais na descoberta de novos materiais, limitações e aplicação de AM como método para predição de propriedades moleculares. Em seguida, é mostrado representações da

estrutura molecular, descritores moleculares e o estado da arte.

Capítulo 4: MATERIAIS E MÉTODOS, apresenta as etapas realizadas para o processo da predição de propriedades. Em especial, é mostrada a análise exploratória feito sobre o conjunto do QM9 e os processos envolvidos na extração e seleção de atributos.

Capítulo 5: RESULTADOS E DISCUSSÕES, apresenta as avaliações realizadas nas predições das propriedades moleculares, os resultados são comparados usando o erro absoluto médio sobre o QM9 e subconjuntos de moléculas de tamanho único para cada método de seleção. Além disso, foram apresentados os resultados na predição usando um atributo bastante comum na área de ciência dos materiais e AM, a matriz de Coulomb.

Capítulo 6: CONCLUSÃO, apresenta a conclusão deste trabalho, assim como trabalhos futuros.

2 APRENDIZADO DE MÁQUINA

Neste capítulo os conceitos sobre AM estão estruturados da seguinte maneira: na Seção 2.1 são apresentados os conceitos iniciais sobre o tema. Na Seção 2.2 é apresentado o fluxo de processos em uma tarefa de AM. Na Seção 2.3 é apresentado o algoritmo da árvore de decisão. Na Seção 2.4 são apresentados métodos *ensemble*. Na Seção 2.5 é apresentado o algoritmo *least absolute shrinkage and selection operator*. Na Seção 2.6 são apresentados conceitos sobre redes neurais artificiais. Na Seção 2.7 é contextualizada a otimização de parâmetros e hiperparâmetros. Na Seção 2.8 são introduzidos conceitos e métodos sobre seleção de atributos.

2.1 Conceitos iniciais

A extração manual de conhecimento em uma quantidade massiva de dados é inviável por demandar tempo e mão de obra altamente especializada. Assim, o uso do AM mostra-se como uma alternativa a tais limitações, ao aprender a extrair conhecimento de maneira automática sobre um conjunto de dados, denominado como exemplos ou experiências. AM se difere dos paradigmas da programação convencional, no qual a saída é obtida através de regras pré-definidas para cada entrada (BISHOP, 2006; MITCHELL, 1997), o que viabiliza diversas aplicações nas áreas de estudo relacionadas a biologia, medicina, finanças, química e economia (WU et al., 2014; AL-JARRAH et al., 2015). Na ciência dos materiais, o AM mostra-se como tomador de decisão para o desenvolvimento de novos materiais com propriedades específicas (ASPURU-GUZI; PERSSON, 2018).

Para que um problema seja resolvido por métodos de AM é preciso fazer uma descrição adequada dos dados. Tal descrição, por ser feita através da estruturação de uma tabela atributo-valor, ou seja, uma tabela contendo atributos e o rótulo de cada exemplo em suas linhas vide Tabela 2.1. Um atributo é toda representação que codifica alguma informação do objeto do problema. Uma lista de atributos forma um vetor de característica $C = (x_1, \dots, x_d) \in \mathbb{R}^D$ que compõe os exemplos x do conjunto de dados. Como será visto posteriormente, a estruturação apresentada na Tabela 2.1 é usada no caso do aprendizado supervisionado, porém, existem cenários onde os dados não contêm os rótulos.

Além do mais, na resolução de um problema por meio de AM são considerados os seguintes fatores: (i) definição da tarefa que deve ser executada; (ii) escolha da medida de desempenho que avalie os resultados da tarefa; (iii) amostras, exemplos do conjunto de dados usados no processo de aprendizagem do algoritmo. Dessa forma,

um algoritmo de AM estará aprendendo se os dados de treinamento melhoram o resultado da medida de desempenho. Uma definição de AM mais formal é descrita como:

Um programa de computador aprende pela experiência E, com respeito a algum tipo de tarefa T e desempenho P, se sua performance P nas tarefas em T, na forma medida por P, melhoram com a experiência E (MITCHELL, 1997, p. 2).

Tabela 2.1 - Tabela formato atributo-valor.

Exemplo	Atributos				Rótulo
	A_1	A_2	$\dots$	A_d	
1	$x_1^{[1]}$	$x_2^{[1]}$	$\dots$	$x_d^{[1]}$	y_1
2	$x_1^{[2]}$	$x_2^{[2]}$	$\dots$	$x_d^{[2]}$	y_2
$\dots$	$\vdots$	$\vdots$	$\dots$	$\ddots$	$\vdots$
n	$x_1^{[n]}$	$x_2^{[n]}$	$\dots$	$x_d^{[n]}$	y_n

Fonte: Produção do autor.

Neste contexto, em um problema de classificação se um *e-mail* é ou não é *spam*, a experiência deve provir de uma base de dados com *e-mails* previamente rotulados. A medida de desempenho pode ser a taxa da classificação correta e a tarefa é classificar o conteúdo da mensagem. Ao fim, espera-se que o algoritmo consiga fazer uma boa generalização para dados não vistos.

Grosso modo, o AM é dividido em dois paradigmas que irão aplicar técnicas computacionais, visando objetivos diferentes, de acordo com o tipo do conjunto de dado, estes são: (i) aprendizado supervisionado; e (ii) aprendizado não-supervisionado. Um terceiro paradigma, denominado aprendizado por reforço, também é comumente definido na literatura. Entretanto, considerando o escopo desse trabalho, tal paradigma não será considerado.

2.1.1 Aprendizado supervisionado

No paradigma do aprendizado supervisionado deve-se apresentar ao algoritmo de AM um conjunto de treino, no qual as amostras x contém os valores de seus atributos e do rótulo y . O rótulo é a característica de interesse que se pretende prever que

pode assumir valores discretos ou contínuos, sendo este definido no processo de rotulagem no qual, geralmente, é realizado manualmente por um especialista, o supervisor (MURPHY, 2012).

Neste paradigma, é esperado que o algoritmo de AM aprenda uma função-conceito f que mapeie as entradas à respectiva saída de cada exemplo, $f : x \rightarrow y$. Uma função-conceito cujos rótulos contenham valores contínuos é nomeado como um problema de regressão. Por outro lado, caso os rótulos sejam valores discretos usa-se o termo classificação (BISHOP, 2006).

Logo após o treinamento do algoritmo, espera-se que o modelo gerado seja capaz de realizar uma boa generalização para dados não vistos. Nesta validação, deve-se separar um conjunto de exemplos da base de dados que não tenha sido usado no processo de treinamento e os rótulos apenas devem ser considerados para a comparação das saídas preditas.

2.1.2 Aprendizado não-supervisionado

No aprendizado não-supervisionado o conjunto de treino não diz nada a respeito sobre o rótulo de cada amostra, visto que nem sempre o processo de rotulagem é viável por demandar tempo, especialista no domínio e estar sujeito a falha humana, principalmente quando se tem um grande volume de dados. Por não conter um rótulo, o algoritmo de AM deverá aprender algum padrão presente na estrutura dos dados (DUDA et al., 2012). Métodos neste paradigma inclui a formação de grupos partindo do pressuposto que dados semelhantes devem pertencer ao mesmo grupo, identificação de exemplos que apresentam um padrão diferente da normalidade, redução da dimensionalidade do vetor de característica e modelos geradores.

De maneira análoga, pode-se aplicar o aprendizado não-supervisionado para encontrar padrões em mensagens de *e-mails*, de modo a agrupar conteúdo de texto parecidos entre si e separar aqueles distintos, dado uma função de similaridade. Este agrupamento será feito sem o conhecimento *a priori* das classes, porém, com base no padrão reconhecido, pode ser possível interpretar os resultados com o objetivo de categorizar se um *e-mail* é *spam* ou não, ou se o conteúdo da mensagem está relacionado a promoção, rede social ou compras, por exemplo.

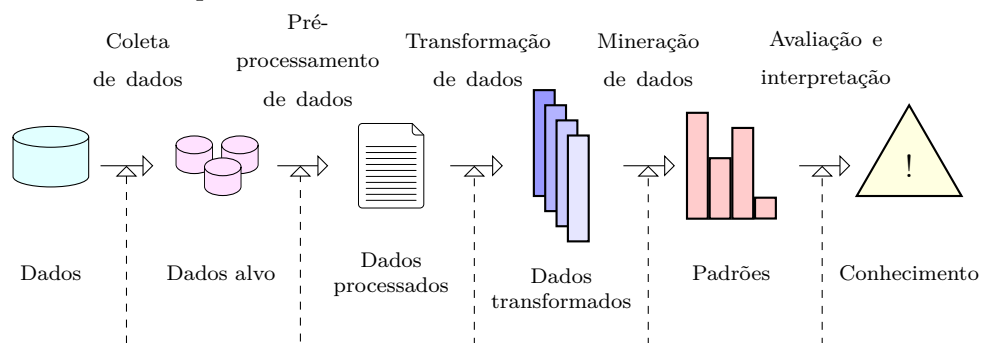
2.2 Fluxo de processos em aprendizado de máquina

Algoritmos de AM facilitam o trabalho dos especialistas ao permitir a extração de conhecimento sobre uma base de dados de forma automática. Como visto anteri-

ormente, neste cenário os dados são elementos fundamentais e é baseado neles que os algoritmos de AM irão fazer a busca do conhecimento para tomadas de decisões. Também é com base neles que os algoritmos serão capazes de melhorar a sua acurácia na generalização de dados futuros. Dessa forma, faz-se necessário estabelecer tarefas que permitam identificar, organizar e preparar os dados, atacando-se as seguintes questões: Qual problema pretende-se resolver usando um algoritmo de AM? Como os dados deste problema serão coletados? Como será feita a organização destes dados? Como será feita a extração de atributo destes dados para o algoritmo de AM?

Posteriormente, deve ser feito a busca do conhecimento por meio de um algoritmo de AM. Nesta etapa, faz-se preciso a escolha de um algoritmo de AM considerando o tipo do problema, como por exemplo, a classificação, regressão ou clusterização. Em seguida, deve ser feita a avaliação da capacidade de generalização do modelo em dados futuros. A Figura 2.1 apresenta o fluxo de processos com todas as orientações de práticas para a identificação de conhecimento embutido em dados. Cada um desses processos faz parte das fases na descoberta de conhecimento em base de dados, este são: coleta, pré-processamento, transformação, mineração de dados, avaliação e interpretação dos resultados (FAYYAD et al., 1996). A seguir serão descritas cada uma dessas fases.

Figura 2.1 - Diagrama das fases da descoberta de conhecimento em base de dados. Vale ressaltar que tais fases não necessariamente são executadas de maneira linear.



Fonte: Adaptada de Fayyad et al. (1996).

- **Coleta de dados:** Inicialmente, é necessário identificar o que pretende ser resolvido usando os métodos de mineração de dados. Neste primeiro momento é importante entender quais são os objetivos, recursos e restri-

ções existentes no contexto do problema. Com base nisso deve ser feito a aquisição dos dados pelas fontes definidas (por exemplo, em banco de dados, aplicações *web*, sensores) (HARRINGTON, 2012). Os dados coletados podem ser gravados em diferentes formatos de arquivos permitindo uma maior facilidade no tratamento das informações desejadas. Exemplos de formatos comuns são o *Comma-separated values* (CSV) e *Java Script Object Notation* (JSON). Nesta fase também é feito um estudo mais detalhado dos dados por meio de uma análise exploratória e a seleção dos dados, isso é, quais atributos e amostras devem ser submetidos para a solução do problema.

- **Pré-processamento de dados:** Nesta fase os dados passarão por inúmeros processos, como remoção de ruídos, fusão de dados providos de múltiplas fontes, eliminação de amostras duplicadas, conversão, combinação de atributos e tratamento de amostras com atributos ausentes (HAN et al., 2011).
- **Transformação de dados:** Tem por objetivo transformar os dados brutos em atributos para o problema em questão. Esta fase é uma das mais importantes, já que é por meio dos atributos que o algoritmo de AM fará o mapeamento da amostra até a saída desejada. A transformação de atributos exige a combinação de conhecimentos do domínio do problema, intuição e métodos matemáticos (SARKAR et al., 2018). Além disso, é esperado que os atributos definidos sejam sempre numéricos para serem utilizados como entrada em algoritmos de AM.

Contudo, ao finalizar o processo de extração de atributos, é possível obter um conjunto de atributos extenso, no qual nem todos os atributos sejam relevantes ao problema. Neste caso, a redução do conjunto dos atributos evita a construção de modelos complexos e é chamada de seleção de atributos. Na Seção 2.9 serão introduzidos os conceitos e alguns métodos sobre a seleção de atributos.

Além disso, alguns algoritmos de AM podem ter sua acurácia degradada por conta da magnitude dos valores dos atributos. Por exemplo, o coeficiente de um atributo em um método de regressão pode ser simplesmente maior que de outro atributo por conta da sua magnitude. Neste caso, o modelo irá considerar apenas aquele atributo de maior magnitude como importante, fazendo com que os outros tenham pouco impacto na predição final. Para evitar esse problema, pode-se utilizar técnicas de normalização

visando colocar todos os atributos na mesma faixa de valores.

- **Mineração de dados:** Nesta fase é feita a escolha dos algoritmos de AM e o ajuste dos parâmetros para a realização da extração de conhecimento. Dentre as tarefas exercidas pela mineração de dados, encontra-se a classificação, regressão, clusterização, associação (HAN et al., 2011).
- **Avaliação e interpretação dos resultados:** Nesta última fase, o conhecimento obtido na tomada de decisão é avaliado e interpretado. A avaliação é feita verificando o desempenho do modelo com base em métricas adequadas de acordo com a tarefa (por exemplo, classificação, regressão, clusterização) (HAN et al., 2011). No caso da regressão, uma medida bastante utilizada é o erro absoluto médio (EAM). O EAM é obtido por meio da média do módulo da diferença entre o valor real e o predito vide Equação 2.1.

$$\text{EAM} = \frac{1}{n} \sum_{i=1}^n |y^{[i]} - \hat{y}^{[i]}| \quad (2.1)$$

no qual $y^{[i]}$ é a saída desejada da i -ésima amostra do conjunto de dados $\mathcal{D}$ com n amostras e $\hat{y}^{[i]}$ é a saída do modelo para a i -ésima amostra pertencente a $\mathcal{D}$.

2.3 Árvore de decisão

Árvore de decisão é um algoritmo de AM supervisionado aplicado tanto em problemas de regressão como classificação. O modelo é caracterizado como não paramétrico, isto é, o conjunto de parâmetros pertinentes ao algoritmo depende da quantidade de amostras fornecidas no processo de treinamento. Logo, possibilitando a construção de modelos complexos capazes de capturar a variabilidade dos dados, o que pode acarretar excelentes predições. No caso da árvore de decisão, a quantidade de nós é o parâmetro que varia conforme o conjunto de treinamento (HARRINGTON, 2012).

Primeiramente, para melhor entender os conceitos deste modelo, será preciso introduzir algumas terminologias sobre a estrutura árvore. Uma árvore é um conjunto de um ou mais nós conectados entre si por ramos. As árvores são desenhadas de cabeça-para-baixo a fim de criar uma estrutura hierárquica, no qual se denomina como raiz o nó que sucede qualquer outro nó pertencente a árvore; nós internos, são aqueles que no sistema hierárquico sempre têm um nó inferior a eles chamado de nó

filho; e os nós folhas, que são considerados como terminais por não conter nós filhos. Na teoria de grafos, uma árvore é representada por um grafo conexo e acíclico, no qual cada nó é um vértice e o ramo uma aresta. Além disso, uma árvore é binária se um nó interno contém no máximo dois nós filhos, denominados como filho esquerdo e filho direito (ROKACH; MAIMON, 2005).

Uma árvore de decisão é um algoritmo que usa a estrutura de árvore para induzir um conjunto de regras na construção do modelo preditivo ou da hipótese. Nesta abordagem, os nós folhas são rotulados como a saída ou resposta do modelo e os demais nós desempenham a função de particionar o espaço de atributos. Em geral, o particionamento feito em um nó interno é gerado a partir de um único atributo e os seus nós filhos podem criar outra nova partição, representando uma nova regra de decisão, ou apenas a atribuição de um valor a uma dada instância de entrada, caso estes sejam um nó folha (HARRINGTON, 2012).

Considere um conjunto de treinamento composto por pares de entrada e saída (x, y) . Em uma árvore de decisão, cada atributo x_i é um potencial candidato a compor um nó da árvore e cada valor pertencente a x_i é um possível limiar para a partição do espaço de atributos. Idealmente, espera-se encontrar um valor em x_i , tal que as amostras rotuladas como y_1 caiam em uma partição e todas as outras amostras pertencentes ao rótulo y_2 caiam na outra partição, portanto, $y = y_1 \cup y_2$ (ROKACH; MAIMON, 2005). Contudo, encontrar o conjunto de atributos ótimo para a construção da árvore e o limiar do x_i de cada nó é um problema NP-completo. Neste sentido, heurísticas são empregadas para a procura de uma solução local ótima (MURTHY, 1998).

O método CART (do inglês, *Classification and Regression Trees*) é um exemplo de algoritmo que permite a criação de uma árvore de decisão binária. O método emprega um algoritmo guloso na procura do atributo ótimo e uma abordagem *top-down* para a definição do modelo. Isto é, inicialmente, o método considera que os dados fornecidos são pertencentes a uma única partição e sucessivamente, a cada nova iteração, realiza a criação de duas novas partições no espaço de atributos até que o critério de parada seja atingido. As duas partições são associadas aos nós filhos esquerdo e direito do nó particionado (ROKACH; MAIMON, 2005; SEGAL, 1988).

A criação de uma nova partição é feita usando um critério de impureza que considera tanto o atributo quanto o limiar de divisão. Tal critério pode variar dependendo do problema em estudo. No caso da regressão, o critério visa a redução da variância das predições de y por meio da soma dos quadrados (SQ) da diferença entre a saída

deseja e predita. A redução da variância faz com que as partições criadas em cada nó agrupem amostras com valores de y similares. Na equação 2.2 é apresentado o cálculo de SQ para uma única partição de um nó t (SENI; ELDER, 2010).

$$\text{SQ}(t) = \frac{1}{n_t} \sum_{i=1, i \in D_t(\mathbf{x}_i, v)}^n \left(y_i^{[i]} - h(\mathbf{x}^{[i]})_t \right)^2, \quad (2.2)$$

no qual $D_t(\mathbf{x}_i, v)$ são as amostras que caíram na partição do nó t dado o atributo $\mathbf{x}_i$ e o limiar de partição v , n_t é a quantidade de amostras em $D_t(\mathbf{x}_i, v)$, $x^{[i]}$ e $y^{[i]}$ são os valores de entrada e saída da i -ésima amostra de $D_t(\mathbf{x}_i, v)$, respectivamente, e $h(\mathbf{x}^{[i]})_t$ é o cálculo da média dos valores de y em $D_t(\mathbf{x}_i, v)$ como definido na equação 2.3.

$$h(\mathbf{x}^{[i]})_t = \frac{1}{n_t} \sum_{i \in D_t(\mathbf{x}^{[i]}, v)}^n y^{[i]} \quad (2.3)$$

No entanto, a qualidade de uma divisão s feita em um nó t deve considerar as duas partições criadas c_1 e c_2 , que são os filhos da esquerda t_e e direita t_d de t (SEGAL, 1988). Neste sentido, equação 2.4 é usada no processo de otimização para a busca da solução ótima. A relação entre a equação 2.4 e equação 2.2 é inversamente proporcional e portanto, a solução ótima é encontrada pela maximização da equação 2.4.

$$\phi(s, t) = \text{SQ}(t) - \text{SQ}(t_e) - \text{SQ}(t_d) \quad (2.4)$$

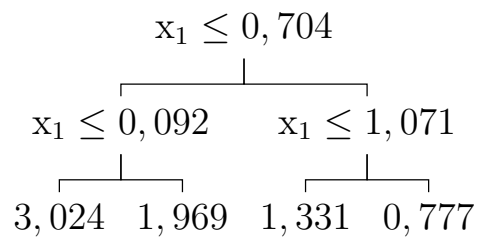
Geralmente, para a definição do limiar de um atributo $\mathbf{x}_i$, percorre-se todos os possíveis valores v , sendo $v \in \mathbf{x}_i$, com o objetivo de encontrar o v que minimize a impureza da partição de um nó t . Outra abordagem visa ordenar os elementos em $\mathbf{x}_i$ e usar a média de v adjacentes. Um terceiro método propõe o uso de percentil para reduzir o custo computacional.

No pior caso, o algoritmo da árvore de decisão pode gerar uma árvore que contenha praticamente uma partição para cada valor de y , causando um *overfitting* sobre os dados de treinamento. Para evitar tal problema é comum utilizar técnicas de poda durante a construção da árvore. Algumas abordagens de poda incluem, por exemplo, a utilização de um limiar de impureza, quantidade mínima de amostra

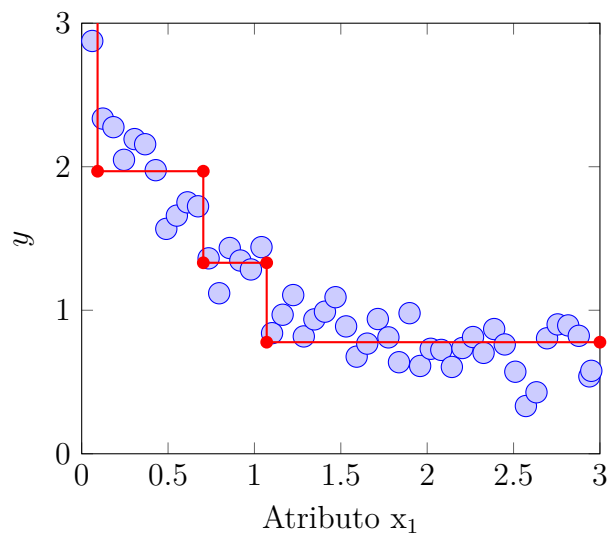
para a inserção de um nó na árvore, ou um teste de significância estatística para a realização da partição (SENI; ELDER, 2010).

Um dos pontos positivos da árvore de decisão é a facilidade em relação à interpretação. A Figura 2.2(a) apresenta a árvore gerada para um problema de regressão. A leitura deve começar do nó raiz e se $x_1 \leq 0,704$, o próximo nó a ser verificado é o filho esquerdo do nó raiz e caso contrário, o filho direito deverá ser o verificado. A leitura deve ser continuada até que um nó folha seja atingido. Já Figura 2.2(b) mostra as partições geradas (linha vertical) e a média dos valores em cada uma das partições (linha horizontal).

Figura 2.2 - Interpretação da árvore de decisão.



(a) Árvore gerada



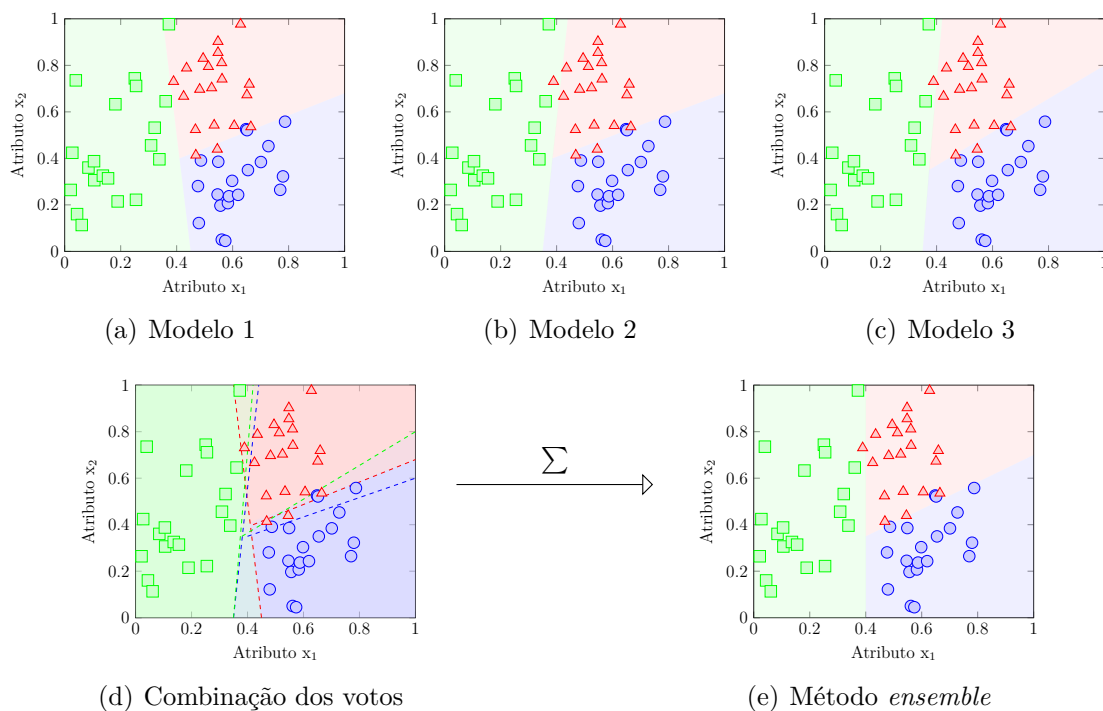
(b) Partições no espaço de atributo

Fonte: Produção do autor.

2.4 Métodos *ensemble*

Métodos *ensemble* são algoritmos que treinam múltiplos classificadores ou regressores, denominados como modelos base, e combinam a saída de cada um dos modelos treinados por meio de um esquema de voto para a geração do resultado. O esquema de votos nos métodos *ensemble* tem por vantagem evitar que pequenas modificações nos parâmetros dos modelos treinados afetem a predição final, em outras palavras, a redução da variância do modelo preditivo (ZHANG; MA, 2012). Figura 2.3 ilustra a redução da variância em um problema de classificação aplicando um método *ensemble*.

Figura 2.3 - Redução da variância usando métodos *ensemble*. (a), (b) e (c) ilustram a fronteira de decisão de 3 modelos bases. (d) A combinação dos resultados de cada modelo base. As linhas tracejadas representam as fronteiras de decisão dos modelos (a), (b) e (c), sendo que a linha vermelha representa o modelo (a), azul o modelo (b) e verde o modelo (c). (e) Fronteira de decisão do método *ensemble*.



Fonte: Adaptada de Zhang e Ma (2012).

Os modelos base devem ser gerados usando o conjunto de treino sobre um ou mais algoritmos de AM. Normalmente, os métodos *ensemble* utilizam um único algoritmo (por exemplo, árvore de decisão ou uma rede neural artificial) e o treinamento de

tais modelos têm como base uma das seguintes modalidades:

- **Método paralelo**, nesta modalidade os modelos base são treinados ao mesmo tempo, e de maneira independente para a geração de hipóteses fracamente correlacionadas (LIU et al., 2016).
- **Método sequencial**, nesta abordagem os modelos base são treinados um por vez, sendo que o modelo gerado na iteração t tem influência na geração do modelo na iteração $t + 1$ (ROKACH, 2005).

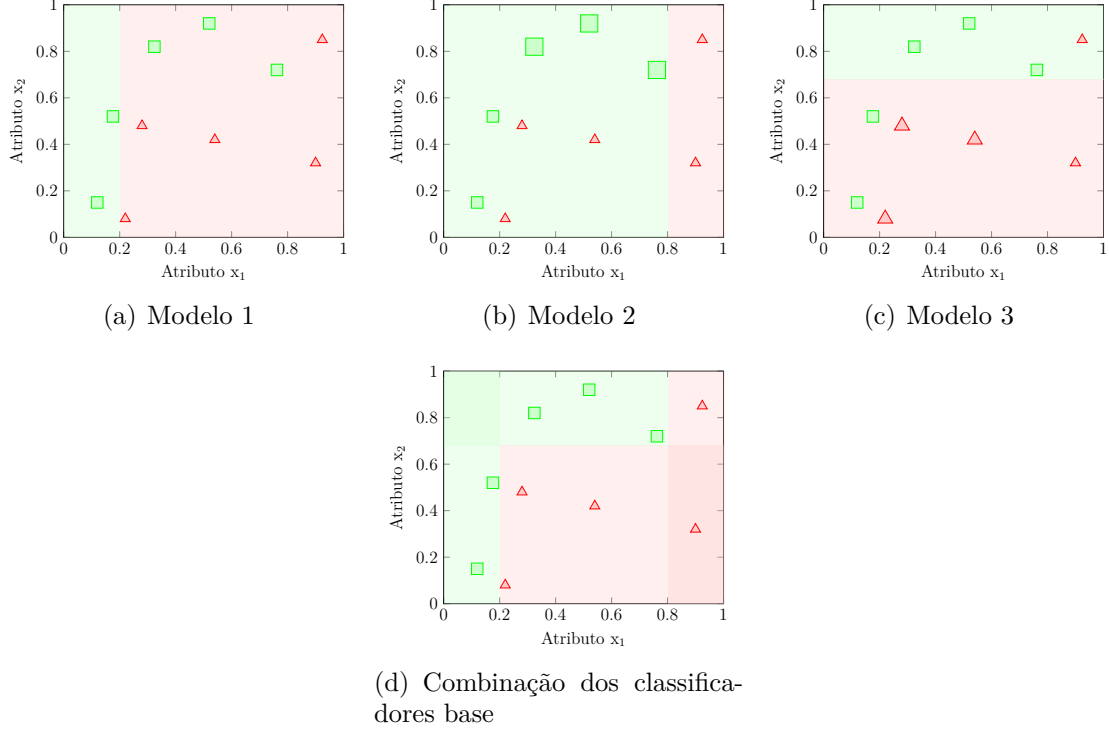
O processo de combinação dos votos é dependente do problema em estudo. Se o problema for de classificação, então o voto majoritário é o escolhido. Caso o problema seja uma regressão, então é feita uma média dos votos. Para que um método *ensemble* apresente resultados superiores comparado a qualquer um de seus modelos base é desejado que seus membros individuais sejam acurados e os mais diversos possíveis. A diversidade neste contexto pode ser inserida ao fazer uma amostragem no conjunto de treino, adicionar uma aleatoriedade no processo de inicialização dos parâmetros ou fazer uma seleção aleatória de atributos, por exemplo. Os métodos *ensemble* mais conhecidos são *boosting*, *bagging* e RF.

2.4.1 *Boosting*

O método *boosting* baseia-se em um processo sequencial, no qual cada modelo gerado na iteração $t + 1$ tenta melhorar seu desempenho sobre as amostras preditas incorretamente pelo modelo da iteração t . Inicialmente, todas as amostras são atribuídas com pesos iguais, ou seja, nenhuma amostra é enfatizada nesta etapa de treinamento. A cada nova iteração os pesos são ponderados de acordo com os erros e acertos do modelo base da iteração anterior. Se uma amostra for classificada erroneamente, então seu peso deverá ser incrementado para que o modelo base sucessor tente classificá-la corretamente e caso contrário, seu peso deverá ser reduzido. Esse processo de iteração deve ser seguido até que todos os modelos base estejam treinados. Por fim, no esquema de votação cada classificador é ponderado de acordo com a sua acurácia para a predição de uma nova instância (FREUND et al., 1996).

Em suma, a ideia principal dos métodos *boosting* é definir conjuntos de treino que permitam que cada modelo base infira algo novo sobre os dados. A Figura 2.4 ilustra este processo, nela é possível perceber que cada classificador capturou características específicas dos dados e o resultado final é a composição de todos os modelos bases. O algoritmo *boosting* mais conhecido é o AdaBoost (SENI; ELDER, 2010).

Figura 2.4 - Exemplo do método *boosting* para um problema de classificação binária. (a), (b) e (c) ilustram a fronteira de decisão de 3 modelos bases. Os pesos das amostras foram ilustrados usando pontos de diferentes tamanhos. (d) Fronteira de decisão do método *boosting*.



Fonte: Adaptada de [Chen e Millar \(2017\)](#).

2.4.2 *Bagging*

O método *bagging* baseia-se em um processo paralelo, no qual cada modelo base é treinado com uma amostragem *bootstrap* do conjunto de treino. O *bootstrap* é um método que realiza amostragem com reposição dos dados pertencentes ao conjunto de treino. A reposição permite que sejam escolhidos uma mesma amostra mais de uma vez. Dessa forma, dado um conjunto de treino $\mathcal{D} = \{(x^{[1]}, y^{[1]}), \dots, (x^{[n]}, y^{[n]})\}$ composto por n amostras, $|\mathcal{D}| = n$, o método *bagging* irá treinar cada modelo base amostrando subconjuntos de $\mathcal{D}$ com reposição de tamanho n ([BREIMAN, 1996a](#); [SENI; ELDER, 2010](#)).

Na amostragem *bootstrap*, a probabilidade da amostra com o índice $m \in \{1 : n\}$ ser amostrada é de $\mathbf{P}((x^{[i]}, y^{[i]}) = (x^{[m]}, y^{[m]})) = \frac{1}{n}$ e a probabilidade de m não ser amostrada é de $\mathbf{P}((x^{[i]}, y^{[i]}) \neq (x^{[m]}, y^{[m]})) = 1 - \frac{1}{n}$ para todo $(x^{[i]}, y^{[i]}) \in \mathcal{D}$. Portanto, a probabilidade de m ser selecionado pelo menos uma vez em um subconjunto $\mathcal{D}'$ é

de $1 - (1 - \frac{1}{n})^n$ vide equação 2.5. Para um valor grande de n , $\mathcal{D}'$ contém, em média, $1 - (1 - \frac{1}{n})^n = 63,2\%$ amostras do conjunto $\mathcal{D}$ (DIETTERICH, 2000).

$$\begin{aligned}
\mathbf{P}((\mathbf{x}^{[m]}, y^{[m]}) \in \mathcal{D}') &= 1 - \mathbf{P}((\mathbf{x}^{[m]}, y^{[m]}) \notin \mathcal{D}') \\
&= 1 - \mathbf{P}(\cap_{i=1}^n (\mathbf{x}^{[i]}, y^{[i]}) \neq (\mathbf{x}^{[m]}, y^{[m]})) = 1 - \prod_{i=1}^n \mathbf{p}((\mathbf{x}^{[i]}, y^{[i]}) \neq (\mathbf{x}^{[m]}, y^{[m]})) \quad (2.5) \\
&= 1 - (1 - \frac{1}{n})(1 - \frac{1}{n}) \cdots (1 - \frac{1}{n}) = 1 - (1 - \frac{1}{n})^n
\end{aligned}$$

As amostras que não compõe o subconjunto $\mathcal{D}'$ formam o conjunto *out-of-bag* (OOB). Neste contexto, é possível utilizar o OOB para estimar a capacidade de generalização do método *bagging*. Essa estimação pode ser feita calculando-se o erro de predição sobre a combinação de votos somente dos modelos base que contêm a amostra $(\mathbf{x}^{[i]}, y^{[i]})$ no conjunto OOB (BREIMAN, 1996b).

2.4.3 Random forest

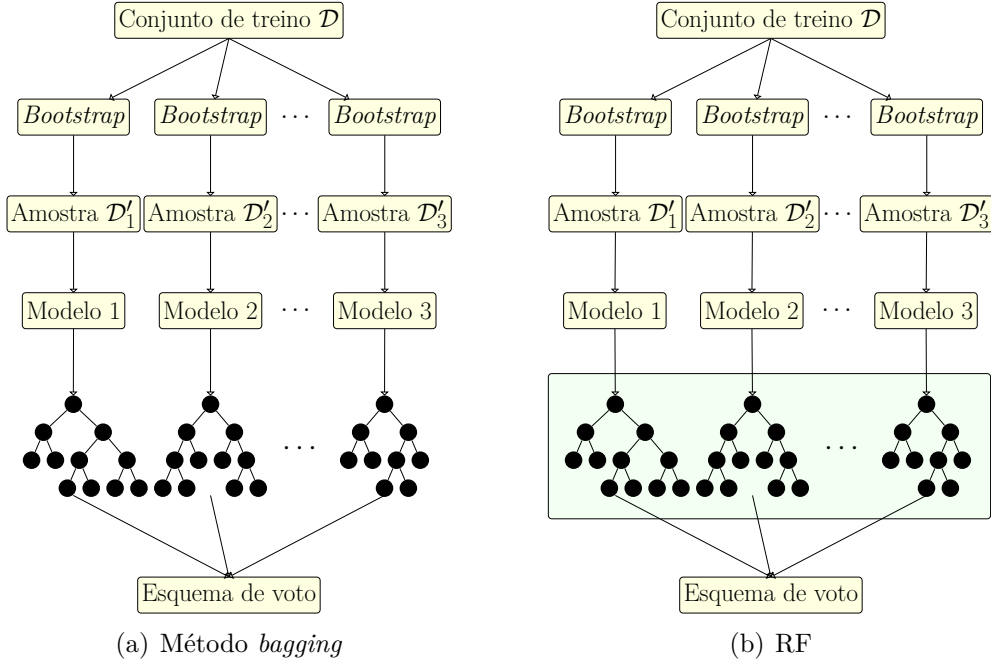
Random forest (RF) é um algoritmo semelhante ao método *bagging*, pois considera a prática da amostragem *bootstrap* para a composição do conjunto de treino dos modelos base. A diferença entre as duas abordagens está na seleção aleatória dos atributos para a construção dos modelos base. Dado um conjunto de treino $\mathcal{D}$ com n amostras e d atributos no vetor de característica C , realiza-se uma amostragem *bootstrap* $\mathcal{D}'$ do conjunto $\mathcal{D}$, no qual $|\mathcal{D}'| = n$ e escolhe-se m atributos de forma aleatória de C , tal que $m \leq d$ (BREIMAN, 2001; LOUPPE, 2014).

No caso do RF, os modelos base são árvores de decisão. A quantidade de árvores k é um hiperparâmetro do algoritmo, ou seja, k deve ser definido pelo especialista antes do processo de treinamento. Na medida que k cresce, o RF demonstra obter resultados com uma maior precisão. Além disso, a escolha de m atributos aleatórios é feita para cada nó da árvore. Isso é, a cada nova partição um subconjunto de m atributos é selecionado aleatoriamente de C , portanto, o valor de m é um hiperparâmetro do algoritmo e deve ser igual para todos os nós da árvore (SENI; ELDER, 2010).

Figura 2.5 apresenta a diferença entre o método *bagging* e o RF. Note que o processo de construção dos subconjuntos é igual nas duas abordagens, contudo, no RF as árvores são construídas selecionando-se m atributos para cada um dos nós. Além disso, é possível usar diferentes algoritmos de AM em *bagging* enquanto no RF

somente árvores de decisão.

Figura 2.5 - Método *bagging* versus RF. (b) A área destaca em verde representa a etapa em que seleção aleatória dos m atributos ocorre.



Fonte: Produção do autor.

2.5 *Least absolute shrinkage and selection operator*

Least Absolute Shrinkage and Selection Operator (LASSO) é um algoritmo que agrega em uma regressão linear um termo de regularização, também chamado de penalização. A regressão linear utiliza de um modelo matemático para representar a relação linear entre uma variável denominada como dependente e outras variáveis chamadas como independentes. No caso mais básico, o objetivo da regressão linear é encontrar a relação entre apenas duas variáveis (FRIEDMAN et al., 2001). Seja Y e X a variável dependente e independente, respectivamente, o modelo matemático assume que

$$Y \approx \beta_0 + \beta_1 X, \quad (2.6)$$

no qual β_0 e β_1 são os coeficientes linear e angular da equação da reta, respectivamente. Esses coeficientes devem ser ajustados de acordo com o conjunto de treino e

por conta disso, são considerados como parâmetros do modelo. Considere um conjunto de treino composto por n amostras, a regressão linear irá estimar os valores de $\hat{\beta}_0$ e $\hat{\beta}_1$ que satisfaçam a seguinte condição $y^{[i]} \approx \hat{\beta}_0 + \hat{\beta}_1 x^{[i]} \forall i = 1, \dots, n$. Perceba que o símbolo $\hat{}$ é empregado para denotar algum parâmetro ou valor de saída estimado. Existem diversas abordagens para estimar esses parâmetros, a mais conhecida envolve a minimização da soma dos quadrados dos erros. Seja $\hat{y}^{[i]} = \hat{\beta}_0 + \hat{\beta}_1 x^{[i]}$ a predição da i -ésima amostra, o erro da predição para $x^{[i]}$ é calculado como $e_i = y^{[i]} - \hat{y}^{[i]}$ (FRIEDMAN et al., 2001). Neste sentido, a soma dos quadrados (SQ) dos erros nada mais é do que a quantidade total do erro produzido pelo modelo vide Equação 2.7.

$$SQ = e_1^2 + e_2^2 + \dots + e_n^2 = \sum_{i=1}^n e_i^2 \quad (2.7)$$

A minimização da SQ dos erros é dada, portanto, quando a derivada de SQ em relação a cada coeficiente é igual a 0, conforme a equação 2.8.

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} , \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n (x^{[i]} - \bar{x})(y^{[i]} - \bar{y})}{\sum_{i=1}^n (x^{[i]} - \bar{x})^2} , \end{aligned} \quad (2.8)$$

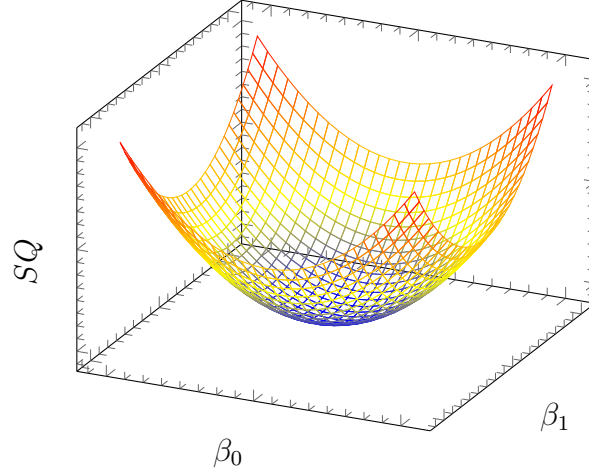
onde $\bar{y}$ e $\bar{x}$ equivale à média de todos os $y^{[i]}$ e $x^{[i]}$. Figura 2.6 ilustra o comportamento para uma função SQ dos erros. É válido salientar que no caso da regressão linear essa função do erro sempre será convexa e, portanto, é possível obter a solução ótima global.

Contudo, quando o problema de regressão envolve muitas variáveis independentes, a solução apresentada na equação 2.8 torna-se um pouco mais complexa, pois envolve o cálculo de matriz inversa e a multiplicação entre matrizes. Na prática, quando se têm um conjunto de dados muito grande a realização de tais operações acabam por demandar bastantes recursos computacionais, ou até mesmo podem ser um problema intratável, já que pode não ser possível obter a matriz inversa (JAMES et al., 2013). Uma segunda abordagem é a utilização de um método iterativo para a correção do erro, chamada como gradiente descendente. Este método será abordado posteriormente na seção sobre as redes neurais artificiais.

Como mencionado anteriormente, o método LASSO inclui na equação da regressão linear um termo de penalização. O termo de penalização faz com que os coeficientes do modelo encolham em direção a zero, permitindo uma fácil interpretação

das variáveis importantes para o problema e a redução da variância do modelo (HARRINGTON, 2012). Considere uma matriz das variáveis independentes de dimensão $n \times d$, onde n é a quantidade de amostras e d o número total de atributos, $\beta = (\beta_0, \beta_1, \dots, \beta_d)$ o vetor dos coeficientes da regressão no qual β_0 é o termo viés, e λ um fator de penalização, o LASSO minimiza a seguinte função de custo (TIBSHIRANI, 1996).

Figura 2.6 - SQ dos erros em dado os parâmetros β_0 e β_1 .



Fonte: Produção do autor.

$$\hat{\beta}_\lambda = \arg \min_{\beta} \sum_{i=1}^n \left(y^{[i]} - \beta_0 - \sum_{j=1}^d \beta_j x_j^{[i]} \right)^2 + \lambda \sum_{j=1}^d |\beta_j| \quad (2.9)$$

O termo de regularização, também conhecido como norma ℓ_1 , é dado pela soma absoluta de todos os coeficientes β_j , exceto o β_0 . O λ é um hiperparâmetro que controla a força da penalização imposta no modelo, no qual para $\lambda = 0$, LASSO equivale a uma regressão linear e para $\lambda = \infty$, a restrição acaba sendo alta e todos os coeficientes são zerados (TIBSHIRANI, 1996).

Neste sentido, ao treinar um modelo LASSO tem-se por objetivo encontrar os valores de $\hat{\beta}$ que minimizem o erro de predição, cuja soma do valor absoluto de tais coeficientes não deve ultrapassar um determinado limiar t , $\sum_{j=1}^d |\beta_j| < t$ (TIBSHIRANI, 1996). Este limiar t tem uma relação inversamente proporcional ao λ , ou seja, quando $t = \infty$, λ é igual a 0 e LASSO é reduzido para uma regressão linear (FONTI; BELITSER, 2017). Dessa mesma forma, para $t = 0$, todos os coeficientes são zerados

já que $\lambda = \infty$. De fato, a equação 2.9 pode ser formalizada usando t como o seguinte problema de otimização.

$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^n \left(y^{[i]} - \beta_0 - \sum_{j=1}^d \beta_j x_j^{[i]} \right)^2 \right\} \quad \text{sujeito a} \quad \sum_{j=1}^d |\beta_j| \leq t \quad (2.10)$$

A regularização implica que certos coeficientes devam ser reduzidos para que a restrição $\sum_{j=1}^d |\beta_j| < t$ seja válida. Simultaneamente, LASSO procura encontrar os valores de $\hat{\beta}_j$ que minimizem o erro de predição. Dessa forma, para que ambas condições sejam satisfeitas, o algoritmo LASSO reduz os coeficientes dos atributos considerados pelo modelo como menos relevante para a predição da saída desejada (TIBSHIRANI, 1996). Em casos específicos, os coeficientes atingem o valor 0, permitindo que uma seleção de atributos seja feita (FONTI; BELITSER, 2017).

2.6 Redes neurais artificiais

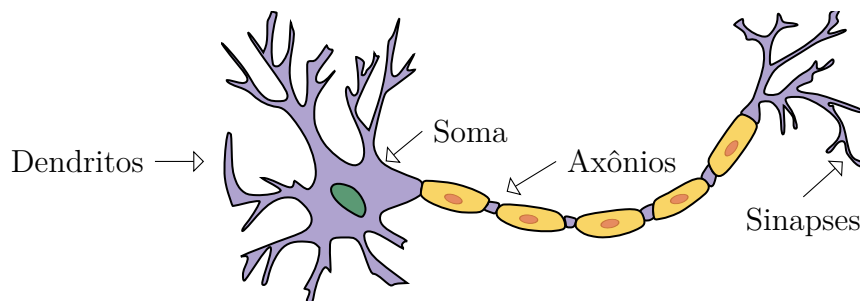
A capacidade humana em aprender e memorizar tipos distintos de informações, de tomar decisões ou pensar é dada pela maior rede complexa do corpo humano, o cérebro. O cérebro faz parte do sistema neural biológico, forma pelo qual informações internas e externas do corpo são capturadas por meio da audição, visão, tato, paladar ou olfato e compartilhadas via impulsos em uma série de neurônios interconectados. Em termos de processo de aprendizagem, o cérebro é capaz de reconhecer padrões ou classificar usando informações previamente conhecidas (LENT, 2004).

O neurônio é a unidade mais básica dessa rede complexa e é aquela que atua no processo de condução da informação através da emissão de impulsos elétricos, quando está em estado de excitação. Nos neurônios sensoriais, por exemplo, tais impulsos podem ser considerados como uma informação a respeito da interação do corpo com o ambiente ou referente a algum processo interno, porém, existem outros tipos de neurônios que desempenham funções distintas, como é o caso dos neurônios de motores e os interneurônios (KANDEL et al., 2014).

Figura 2.7 apresenta as principais partes da estrutura de um neurônio, sendo essas os dendritos — responsáveis por receber os impulsos elétricos —, o corpo celular ou soma — que desempenha a tarefa de processar e definir se um impulso será do tipo excitatório e portanto, compartilhado adiante ou apenas inibitório —, os axônios — quem realizam a transmissão dos impulsos — e as sinapses, que exercem

o processo da condução do impulso por meio de uma conexão entre o axônio do neurônio pré-sináptico e dendrito do neurônio pós-sináptico.

Figura 2.7 - Anatomia de um neurônio.



Fonte: Produção do autor.

O processamento de informação das sinapses altera o estado de um neurônio pós-sináptico, o que eventualmente pode gerar novos pulsos elétricos, denominado potencial de ação. A rede neural artificial (RNA) é um modelo computacional, inspirado no sistema neural biológico, que visa aprender uma determinada tarefa com base em experiências. Por experiência, entende-se toda informação já conhecida e contida em uma base de dados e por tarefa, alguma atividade de reconhecimento de padrões (GOODFELLOW et al., 2016).

Neste modelo, os neurônios artificiais v exercem o processamento da informação por meio de uma operação de soma sobre os dados de entrada $(x_0, x_1, \dots, x_d)$ vezes seus respectivos pesos sinápticos $(w_0, w_1, \dots, w_d)$ e produzem uma saída $\hat{y}$ usando uma função de ativação $\varphi(\cdot)$ sobre o resultado dessa soma (HAYKIN, 2007). Equação 2.11 apresenta a operação descrita anteriormente, no qual x_0 sempre é igual a 1. É possível associar o neurônio com o vértice e os seus pesos sinápticos com o ramo de um grafo, no qual vértices são distribuídos em diferentes níveis, comumente referenciado como camada, e para aqueles pertencentes ao mesmo nível não é possível existir uma conexão entre eles.

$$\hat{y} = \varphi\left(\sum_{i=0}^d x_i w_i\right), \quad (2.11)$$

Uma RNA é constituída de uma camada de entrada que irá receber os estímulos do ambiente, uma ou mais camadas ocultas, que são responsáveis pela transforma-

ção dos sinais (dados) para uma melhor discriminação no processo de aprendizado e uma camada de saída na geração do resultado final. Cada camada contém um grupo de neurônios, a definição da quantidade de neurônios e camadas varia bastante com a tarefa em questão e é considerada como um problema de otimização de hiperparâmetro, como será visto posteriormente (GOODFELLOW et al., 2016).

A primeira proposta de um neurônio artificial foi feita por Warren McCulloch e Walter Pitts, em 1943, e é conhecida como o modelo MCP. Neste modelo, os w_i eram variáveis cujo ajuste acontecia empiricamente, já que não existia um algoritmo de correção de peso e a $\varphi(v)$ resultava em saídas binárias de 0 e 1 representando, respectivamente, a inibição ou transmissão dos sinais de entrada. A saída binária era dada de acordo com um limiar definido manualmente em $\varphi(v)$, no qual para valores maiores ou iguais ao do limiar a saída era 1 e 0, caso contrário (HAYKIN, 2007).

Em 1949, Donald Hebb apresentou o primeiro conceito de aprendizagem, denominada como aprendizagem Hebbiana, que se baseia na correlação entre neurônios artificiais conectados para modificar o peso sináptico existente entre eles. Dessa forma, (i) caso o neurônio pré-sináptico esteja correlacionada positivamente com o neurônio pós-sináptico, o peso é reforçado; (ii) caso o neurônio pré-sináptico esteja correlacionada negativamente com o neurônio pós-sináptico, então o peso é inibido; (iii) caso o neurônio pré-sináptico não esteja correlacionada com o neurônio pós-sináptico, então o peso pode ser tanto reforçado quanto inibido (HEBB, 2005).

Outro marco importante na história das RNAs foi o desenvolvimento da rede Perceptron, em 1958 (ROSENBLATT, 1958). O Perceptron é um modelo de RNA similar ao MCP, porém, acrescidos de um algoritmo de aprendizagem baseado no erro para o ajuste dos pesos, no qual os pesos são acrescidos, caso a saída predita pela rede seja relativamente baixa comparado ao valor de saída desejada. Caso o valor predito seja relativamente alto em relação ao desejado, os pesos devem ser reduzidos. A princípio, a rede foi proposta com apenas uma camada de entrada e saída e conseguia resolver problemas ditos como binários e lineares.

Desde então, as RNAs foram sendo aperfeiçoadas e diversos modelos foram propostos. A rede Perceptron de múltiplas camadas, por exemplo, é composta por mais de um Perceptron e é capaz de aprender relações não lineares entre os dados de entrada e saída. Arquiteturas diferentes foram desenvolvidos com o objetivo de atacar aplicações específicas, como as redes convolucionais para classificação de imagens e as redes recorrentes para os dados sequenciais.

2.6.1 Neurônio

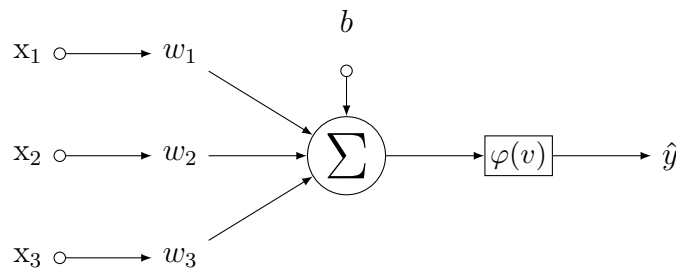
O neurônio é a unidade elementar para o funcionamento de qualquer RNA. Basicamente, um neurônio é composto por um sinal de entrada x_i , que é multiplicado pelo seu respectivo peso w_i . Posteriormente, é feito a somatória dos sinais de entrada ponderados pelos pesos, acrescido do viés b , no campo local induzido v , conforme Equação 2.12, no qual o sinal resultante v é aplicada em uma função de ativação $\varphi(\cdot)$ para geração da saída $\hat{y}$ não linear (GOODFELLOW et al., 2016).

$$v = \sum_{i=1}^d w_i x_i + b , \quad (2.12)$$

$$\hat{y} = \varphi(v)$$

Figura 2.8 ilustra o neurônio artificial e o fluxo percorrido pelo sinal de entrada. A contribuição de um peso w_i específico é dado pelo produto $x_i w_i$, sendo que w_i pode assumir qualquer valor positivo e negativo. Os pesos são ajustados durante o processo de treinamento da rede com o objetivo de que o erro entre a saída desejada y e a predita $\hat{y}$ seja reduzido.

Figura 2.8 - Neurônio artificial.



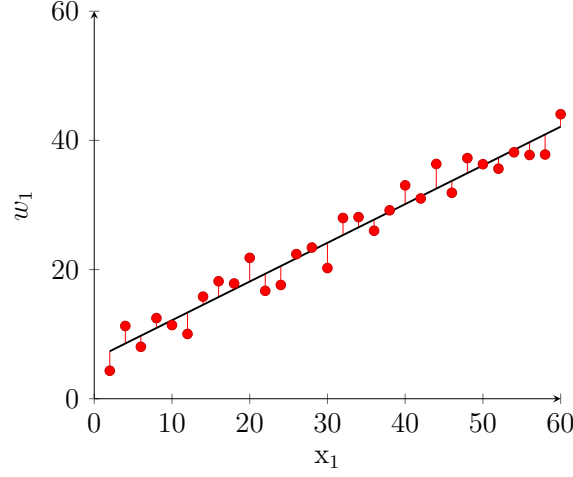
Fonte: Adaptada de Haykin (2007).

2.6.2 Termo viés

Os pesos de um neurônio representam a contribuição de um atributo para a predição da saída desejada. Suponha um problema de regressão composto por um único atributo x_1 . O neurônio irá conter um único peso w_1 e neste caso, uma função de ativação linear, que é a própria função identidade. A representação gráfica desse peso é a reta que minimiza o critério de avaliação do erro de saída vide Figura 2.9. O critério de avaliação pode ser a soma dos quadrados dos erros, como visto na Seção

2.5.

Figura 2.9 - $\hat{y} = \varphi(x_1 w_1 + b)$ para $w_1 = 0.6$, $b = 6.2$ e $\varphi(v) = v$



Fonte: Produção do autor

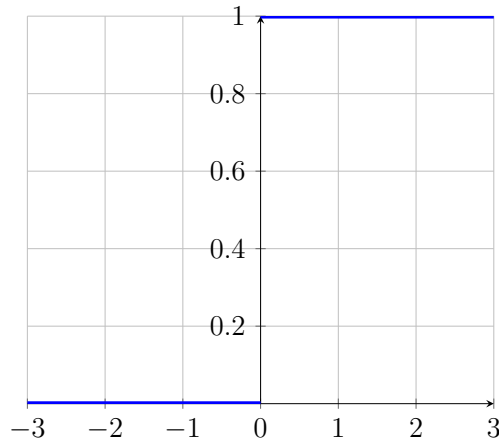
Perceba que na Figura 2.9 a reta gerada não passa pelo ponto de origem $(0, 0)$. Esse deslocamento feito no eixo das ordenadas é dado pelo termo viés b , que permite que mais funções sejam encontradas. Note que as equações 2.11 e 2.12 expressam a mesma operação, porém, em (2.12) o termo viés é denominado como b e é destacada do somatório. Já em (2.11), o termo viés w_0 é considerado dentro do somatório e sempre recebe como entrada 1.

2.6.3 Função de ativação

Elos de conexão, ou conjunto de sinapses, podem ser aplicados para a criação de uma relação linear ou não linear entre a entrada e saída de um neurônio. No primeiro caso, tais elos são denominados como elos sinápticos, enquanto no outro como elos de ativação, no qual a relação não linear é definida por uma função de ativação $\varphi(v)$ (HAYKIN, 2007). A seguir serão apresentadas as principais funções de ativação, estas são: função de limiar, sigmoide e *Rectified Linear Unit* (ReLU).

A Figura 2.10 apresenta a função de ativação de limiar, conhecida na literatura pelo modelo de neurônio MCP, que define o valor de 1 para todo v maior ou igual a 0, e 0 caso contrário vide Equação 2.13. Vale ressaltar que tal função pode assumir diferentes valores de limiar.

Figura 2.10 - Função de limiar.

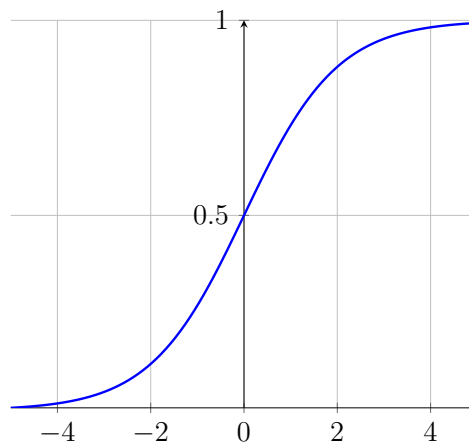


Fonte: Produção do autor.

$$\varphi(v) = \begin{cases} 0 & \text{for } v < 0 \\ 1 & \text{for } v \geq 0 \end{cases} \quad (2.13)$$

A função de ativação sigmoide é uma das mais utilizadas e é definida dentro do intervalo $[0, 1]$ conforme a Equação 2.14, no qual quanto menor for o valor de v , mais próximo do 0 a função irá tender, da mesma forma, que para valores maiores de v , mais próximo do 1 o resultado será, gerando assim um gráfico no formato de um “S” como ilustrado na Figura 2.11 (GLOROT; BENGIO, 2010).

Figura 2.11 - Função sigmoide.



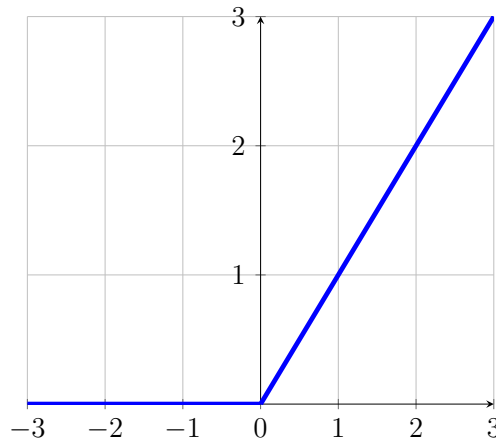
Fonte: Produção do autor.

$$\varphi(v) = \frac{1}{1 + e^v} \quad (2.14)$$

A função de ativação ReLU, vide Equação 2.15, também é bastante utilizada e é definida como 0 para valores de x menores que 0 e função identidade, caso contrário. Geralmente, tal função apresenta um desempenho melhor em relação à função sigmoide por não sofrer do problema da dissipação do gradiente (RAMACHANDRAN et al., 2017). A Figura 2.12 mostra o comportamento dessa função de ativação.

$$\varphi(v) = \begin{cases} 0 & \text{for } v < 0 \\ v & \text{for } v \geq 0 \end{cases} \quad (2.15)$$

Figura 2.12 - Função ReLU.



Fonte: Produção do autor.

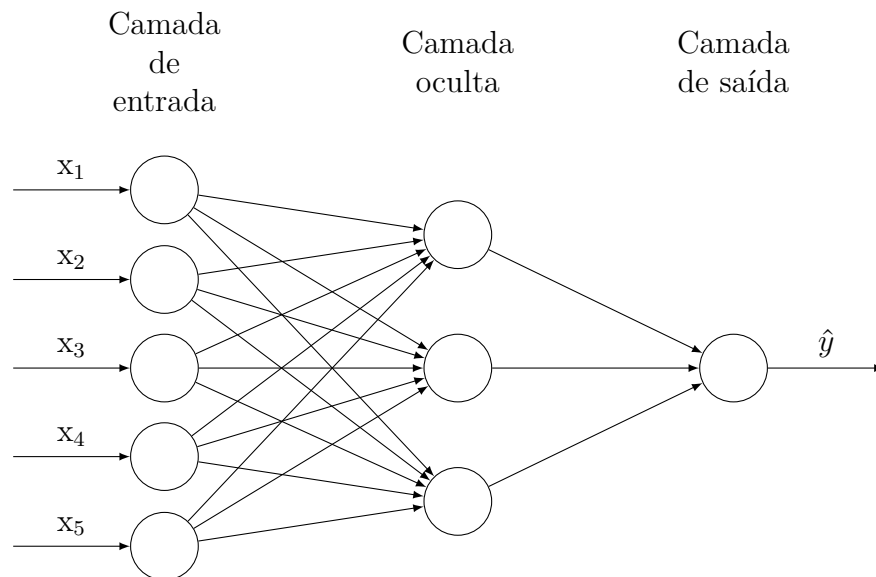
2.6.4 Perceptron de múltiplas camadas

A rede Perceptron de múltiplas camadas é composta por unidades de neurônios distribuídos em diversas camadas, contendo no mínimo uma camada de entrada, oculta e saída. No caso da camada oculta, a quantidade pode ser maior que um, sendo que a inserção dessas novas camadas permite que a rede aprenda relações mais complexas. Além disso, cada uma das camadas pode possuir j neurônios, exceto a quantidade de j na camada de entrada e saída, que fica dependente do total de atributos usados para a descrição do objeto e da quantidade de classes ou saídas do problema, respectivamente (HAYKIN, 2007).

Figura 2.13 representa a estrutura básica de uma rede Perceptron de múltiplas camadas por meio de um grafo de fluxo de sinal. Note que o sinal flui ao longo de um elo somente no sentido definido pela seta e que não existem elos entre neurônios pertencentes a mesma camada ou com uma conexão fazendo retroalimentação.

Neste modelo, a aprendizagem pode ser descrita como um processo iterativo pelo qual os parâmetros livres (intra-neurônios), constituídos pelos pesos e o viés, são adaptados usando os estímulos do ambiente que a rede está inserida. Espera-se que o estímulo recebido faça com que a rede gere uma resposta no tempo $t + 1$ que seja melhor que no tempo t . O tipo de aprendizagem é determinado pela maneira na qual a modificação dos parâmetros ocorre, um exemplo é o mecanismo da aprendizagem por correção de erro (HAYKIN, 2007).

Figura 2.13 - Rede Perceptron de múltiplas camadas.



Fonte: Produção do autor.

2.6.4.1 Propagação *feedforward*

Em uma RNA, a maneira como as unidades de processamento estão conectadas entre si define como é estabelecido a propagação dos sinais. No caso do Perceptron de múltiplas camadas, o método de propagação é denominado como *feedforward*, já que as saídas de uma camada anterior $l - 1$ são usadas como entrada na camada l ,

até que a última camada da rede seja alcançada, onde o sinal de saída é obtido. Vale ressaltar que neste tipo de propagação os pesos da rede se mantêm fixos. A Equação 2.16 apresenta como o sinal é propagado entre os neurônios. Note que (2.16) realiza o mesmo processo que (2.12), porém, (2.16) é destinada para redes com mais de um neurônio (BISHOP et al., 1995).

$$\begin{aligned} v_j^{(l)}(t) &= \sum_{i=1}^n w_{ij}^{(l)}(t) \hat{h}_i^{(l-1)}(t) + b_j^{(l)}(t) \\ h_j^{(l)}(t) &= \varphi(v_j^{(l)}(t)) , \end{aligned} \quad (2.16)$$

no qual, $v_j^{(l)}(t)$ representa a saída produzida pelo j -ésimo neurônio na camada atual l na iteração t , $w_{ij}^{(l)}(t)$ são os pesos sinápticos existentes entre o j -ésimo neurônio na camada atual l e os i -ésimos neurônios da camada anterior $l - 1$ na iteração t , $b_j^{(l)}(t)$ é o termo viés do j -ésimo neurônio na camada atual l na iteração t e $h_j^{(l)}(t)$ é saída de $v_j^{(l)}(t)$ sobre uma função de ativação.

Na Figura 2.13 é possível observar os tipos de informações processadas em um neurônio, o fluxo *feedforward* dentro da rede e o reconhecimento dos sinais interneurônio e intra-neurônios. Os sinais interneurônio, ou neurônios de associação, dizem respeito as conexões existentes entre cada neurônio, já as conexões intra-neurônios, conhecida também como pesos sinápticos, são os valores atribuídos em cada conexão para armazenar o conhecimento adquirido.

2.6.4.2 Aprendizagem por correção de erro

Aprendizagem por correção de erro é o mecanismo base da retropropagação do Perceptron de múltiplas camadas. O método propõe a reconfiguração dos pesos através da definição da energia do erro, levando-se em consideração a saída obtida $\hat{y}$ e o resultado esperado y . Neste contexto, só deverá ser feito um reajuste dos pesos quando y for diferente de $\hat{y}$, ou seja, quando o erro for diferente de zero. Como inicialmente o conjunto de pesos é definido com valores aleatórios — simbolizando que não há um conhecimento na rede — é esperado que o processo de correção seja mais frequente (BISHOP et al., 1995).

Inicialmente, é feita a propagação *feedforward* do sinal de entrada até a camada de saída da rede para a obtenção da energia do erro. Uma forma de calcular a energia do erro é através do erro quadrático médio (EQM) conforme a equação 2.17.

$$EQM = \frac{1}{n} \sum_{i=1}^n (y^{[i]} - \hat{y}^{[i]})^2, \quad (2.17)$$

onde n é a quantidade de amostras. Posteriormente, é feita a correção dos pesos com base na minimização da energia do erro. Essa minimização pode ser dada pela regra de aprendizagem normalmente referida como regra delta ou de Widrow-Hoff, no qual se utiliza a derivada da energia do erro para obtenção do quanto um peso deverá ser ajustado. Algoritmo 1 detalha melhor as etapas da retropropagação.

Algoritmo 1 Algoritmo de treinamento por retropropagação

```

1: procedure RETROPROPAGACAO( $\mathcal{D}, \eta$ )
2:    $w \leftarrow$  Inicialização dos pesos com valores aleatórios;
3:    $b \leftarrow$  Inicialização do termo viés com valor aleatório;
4:   repeat
5:     Apresentação do sinal  $(\mathbf{x}^{[i]}, y^{[i]}) \in \mathcal{D}$ ;
6:     Propagação feedforward dos sinais de entrada para cada neurônio;
7:      $h_j^{(l)}(t) = \varphi\left(\sum_{i=1}^n w_{ij}^{(l)}(t) \hat{h}_i^{(l-1)}(t) + b_j^{(l)}(t)\right)$ 
8:     Cálculo da energia do erro de cada neurônio da camada de saída;
9:     Correção dos parâmetros pela retropropagação dos pesos
10:  until Satisfazer a condição de parada;
11: end procedure

```

Algoritmo 1 recebe como entrada um conjunto de treino $\mathcal{D} = \{(\mathbf{x}^{[i]}, y^{[i]})\}_{i=1}^n$ com n amostras e o valor do hiperparâmetro de taxa de aprendizagem η . O objetivo da retropropagação é propagar o erro gerado na predição, começando pela camada de saída e indo até à camada de entrada, considerando a contribuição feita por cada neurônio no erro gerado (HAYKIN, 2007). Para que este objetivo seja alcançado, a retropropagação ajusta os pesos utilizando o gradiente instantâneo $\frac{\partial \xi}{\partial w_{ij}^{(l)}}(t)$ da função que calcula a energia do erro ξ em relação a cada peso sináptico w_{ij} da camada l na iteração t , como mostrado a seguir.

$$w_{ij}^{(l)}(t+1) = w_{ij}^{(l)}(t) - \eta \frac{\partial \xi}{\partial w_{ij}^{(l)}}(t) \quad (2.18)$$

no qual o termo ξ pode ser qualquer função para o cálculo da energia do erro, por exemplo, o EQM definido em (2.17). Usando a regra da cadeia sobre o $\frac{\partial \xi}{\partial w_{ij}^{(l)}}(t)$, tem-se que:

$$\frac{\partial \xi}{\partial w_{ij}^{(l)}}(t) = \frac{\partial \xi}{\partial \hat{h}_j^{(l)}}(t) \frac{\partial \hat{h}_j^{(l)}}{\partial \hat{y}_j^{(l)}}(t) \frac{\partial \hat{y}_j^{(l)}}{\partial w_{ij}^{(l)}}(t) \quad (2.19)$$

Contudo, ξ depende da saída desejada y_j e como y_j só é conhecida na camada de saída, então, inicialmente, os pesos só são ajustados para a camada de saída. As equações a seguir apresenta a derivação da equação 2.19.

$$\begin{aligned} \frac{\partial \xi_j}{\partial \hat{h}_j}(t) &= \frac{\partial(\frac{1}{2}(y_j(t) - \hat{h}_j(t))^2)}{\partial \hat{h}_j(t)} = -(y_j(t) - \hat{h}_j(t)) = (\hat{h}_j(t) - y_j(t)) \\ \frac{\partial \hat{h}_j}{\partial \hat{y}_j}(t) &= \varphi'(\hat{y}_j(t)) \\ \frac{\partial \hat{y}_j}{\partial w_{ij}}(t) &= \frac{w_{ij}(t)\hat{h}_i(t) + b_j(t)}{\partial w_{ij}(t)} = \hat{h}_i(t) \end{aligned} \quad (2.20)$$

Substituindo as derivações acima na equação 2.19, tem-se que:

$$\frac{\partial \xi_j}{\partial w_{ij}}(t) = (\hat{h}_j(t) - y_j(t))\varphi'(\hat{y}_j(t))\hat{h}_i(t) \quad (2.21)$$

Note que ξ_j e $\hat{h}_j$ representam o erro quadrático de uma amostra e o valor de saída do j -ésimo neurônio da camada de saída, respectivamente. Por conviência usa-se $\frac{1}{2}$ no ξ_j para remover o termo exponencial durante a derivação. Já $\varphi'(\cdot)$ é a derivada da função de ativação. Além disso, a notação l foi omitida já que as equações acima são aplicadas apenas na última camada. Baseando-se em (2.21) é possível definir o gradiente local $\delta_j(t)$ como

$$\begin{aligned} \delta_j(t) &= (\hat{h}_j(t) - y_j(t))\varphi'(\hat{y}_j(t)) \\ \therefore \frac{\partial \xi}{\partial w_{ij}^{(l)}}(t) &= \delta_j(t)\hat{h}_i(t) \end{aligned} \quad (2.22)$$

No caso dos pesos da camada oculta, é preciso considerar os erros de todos os neurônios k da camada subsequente $l + 1$ que estão conectados ao neurônio j na camada l e retropropagar o erro ponderado pelos pesos w_{jk} . A equação para o ajuste dos pesos neste caso só é modificada no cálculo do gradiente local, que é apresentado a seguir.

$$\delta_j(t) = \varphi'(\hat{y}_j(t)) \sum_k \delta_k(t) w_{jk}(t) \quad (2.23)$$

Já derivada da função da energia do erro ξ em relação ao termo viés $b_j^{(l)}$ da camada l na iteração t é apresentada na equação abaixo, no qual $\frac{\partial \hat{y}_j^{(l)}}{\partial b_j^{(l)}}(t) = 1$.

$$\frac{\partial \xi}{\partial b_j^{(l)}}(t) = \frac{\partial \xi}{\partial \hat{h}_j^{(l)}}(t) \frac{\partial \hat{h}_j^{(l)}}{\partial \hat{y}_j^{(l)}}(t) \frac{\partial \hat{y}_j^{(l)}}{\partial b_j^{(l)}}(t) \quad (2.24)$$

2.7 Otimização

Na Seção 2.6.4.2, foi apresentado a aprendizagem por correção de erro para a adaptação dos parâmetros de uma RNA. Esse tipo de ajuste é um problema de otimização, no qual é desejado que o modelo consiga obter a energia de erro mínimo, dada por uma função de custo.

Neste contexto, denomina-se por parâmetro qualquer variável que é ajustada no processo de aprendizado através das amostras e que o seu valor não é definido antes da execução do modelo (DUDA et al., 2012). O cálculo do erro para avaliar a qualidade de generalização de um modelo se baseia na saída obtida e esperada do conjunto de teste, ou seja, sobre os dados não utilizados durante a fase de treinamento. A divisão da base de dados em treino e teste, é conhecida como *hold-out* e é um processo de geração de dados (GOODFELLOW et al., 2016).

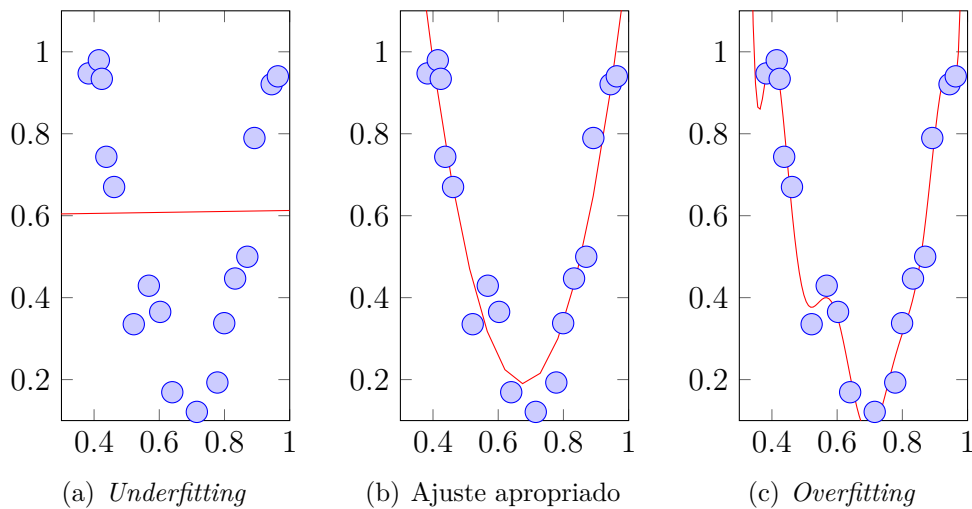
É esperado que parâmetros bem ajustados ao domínio do problema consigam fazer uma boa generalização para dados não vistos e, com isso, apresentem um erro de teste que seja um pouco maior ou igual ao erro de treino. Em determinados casos o modelo obtido pode apresentar um erro ruim sobre o conjunto de treino e teste, ou seja, um *underfitting* dos parâmetros em relação aos dados. Por outro lado, tal modelo pode apresentar um erro aceitável sobre o conjunto de treino, mas não no de teste, neste caso um *overfitting* sobre as amostras de treino.

Problemas relacionados ao *underfitting* geralmente são causados quando o modelo não consegue fazer o mapeamento adequado entre a entrada e saída, podendo ser o motivo dessa causa tanto a quantidade de amostras usadas para o treinamento quanto a escolha de atributos não representativos. Uma terceira alternativa, poderia estar relacionado a escolha de algoritmos muito simples em relação ao comportamento dos dados, por exemplo, a aplicação de um algoritmo capaz de aprender

uma relação linear, sendo que o apropriado seria obter uma superfície de decisão quadrática. Já no caso do *overfitting*, o problema pode estar associado a complexidade superfície de decisão, escolhida através do hiperparâmetros, o que torna o modelo capaz de memorizar as características do conjunto de treino, porém, falha ao generalizar em exemplos não vistos (GOODFELLOW et al., 2016).

Figura 2.14 apresenta três regressões distintas feitas por modelos de diferentes complexidades para a predição de dados regidos por uma função de segundo grau. No primeiro caso, é possível observar que o ajuste da função linear é simples demais sobre a curvatura do problema, caracterizando assim um *underfitting*. Já a regressão apresentada no segundo caso é o resultado esperado, pois é a função que melhor representa a variação dos dados. No terceiro caso, a regressão ajustou-se muito bem aos dados utilizados, no entanto, as oscilações presentes não representam a função original, dessa forma resultando em um problema de *overfitting*.

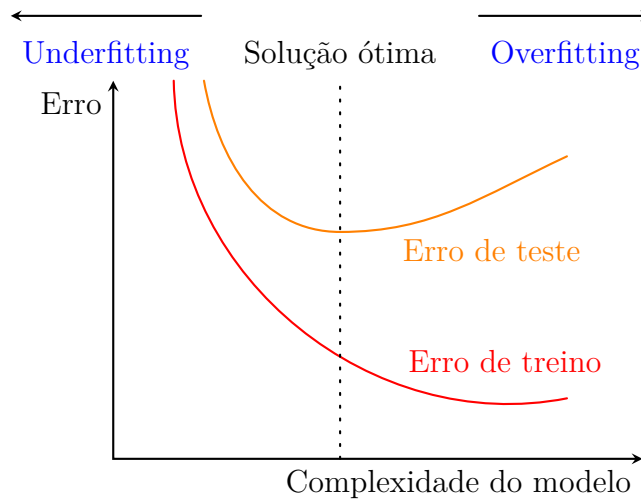
Figura 2.14 - Representação gráfica do ajuste de curva do polinômio para diferentes ordens. (a) ajuste de uma reta. (b) ajuste de uma função quadrática. (c) ajuste de uma função polinomial do nono grau.



Fonte: Adaptada de Goodfellow et al. (2016).

A verificação do *underfitting* e *overfitting* em um algoritmo de AM geralmente é feita comparando o erro de treinamento com o de teste ao modificar os hiperparâmetros do modelo. Dessa forma, se a diferença entre os dois erros crescer conforme a complexidade do modelo aumenta, tem-se um *overfitting* e *underfitting*, caso ambos erros forem altos e o modelo tenha baixa complexidade vide Figura 2.15.

Figura 2.15 - Representação gráfica do *underfitting* e *overfitting* por meio do erro de treino e teste.



Fonte: Produção do autor.

2.7.1 Parâmetros e hiperparâmetros

Denominam-se como parâmetros todas as variáveis de um algoritmo que são ajustadas através de dados, e sem interferência de um especialista, no processo de treinamento (GOODFELLOW et al., 2016). Pode se citar como exemplo, os pesos de uma RNA que são ajustados usando o *backpropagation* para a minimização da energia do erro.

Por outro lado, diz como hiperparâmetros de um modelo todo o conjunto de variáveis que são configuradas por um especialista, sem que as mesmas sejam ajustadas por meio dos dados (GOODFELLOW et al., 2016). Os hiperparâmetros permitem definir a complexidade do modelo e portanto, uma má configuração deles pode resultar em um problema de *overfitting* ou *underfitting*.

Cada algoritmo de AM possui seu próprio conjunto de hiperparâmetros. Por exemplo, no RF tem-se o ajuste da quantidade de árvores no modelo e a quantidade m de atributos do vetor de características que serão escolhidos aleatoriamente para cada nó da árvore. No LASSO, tem-se o λ para o ajuste do termo de penalização. Já nas RNAs, encontra-se a quantidade de camadas da rede, a quantidade de neurônios por camada e a taxa de aprendizagem η .

2.7.1.1 Validação cruzada

Enquanto os parâmetros de um modelo são ajustados pelo próprio algoritmo de AM, os hiperparâmetros devem ser configurados pelos especialistas ou através de técnicas de otimização. No caso dos hiperparâmetros, o ajuste precisa seguir uma abordagem que possibilite avaliar o desempenho do algoritmo para dados fora do conjunto de treinamento e fazer a comparação com diversos modelos candidatos.

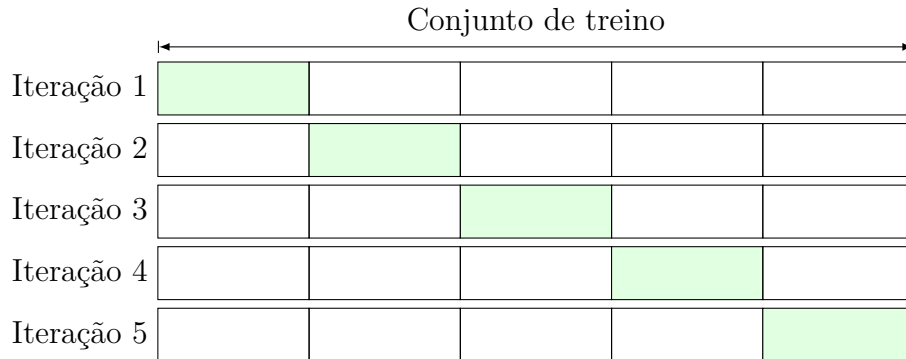
Uma primeira abordagem é a utilização do método *hold-out*, que fora citado anteriormente. O objetivo desta técnica é realizar uma divisão do conjunto de dados entre treino e teste, e treinar os diversos modelos com configurações de hiperparâmetros distintas para que em seguida, esses modelos sejam avaliados pelo conjunto de teste. Dessa forma, o modelo que apresentar o menor erro de teste é então escolhido. Contudo, essa abordagem apresenta um resultado tendencioso, já que o conjunto usado para configuração dos hiperparâmetros é o mesmo que o usado para a avaliação de capacidade de generalização (KOHAVI et al., 1995).

Para evitar a geração de resultados tendenciosos, pode-se utilizar uma outra abordagem que realiza a divisão do conjunto de dados em três novos conjuntos, este são: treino, validação e teste. Neste caso, o conjunto de validação é o responsável por escolher qual configuração de hiperparâmetros gera o melhor resultado. Após isso, o candidato escolhido é avaliado na sua capacidade de generalização pelo conjunto de teste.

No entanto, quando se tem um conjunto de dados pequeno, o conjunto de validação pode não abranger toda a diversidade existente nos dados. Visando solucionar tal problema, uma terceira abordagem, conhecida como validação cruzada, propõe a divisão do conjunto de dados em treino e teste. Em seguida, o conjunto de treino é dividido em k partes de tamanhos iguais e o modelo é treinado em k iterações, no qual a cada iteração, $k - 1$ conjuntos são usados para o treinamento e um conjunto é considerado como validação (KOHAVI et al., 1995).

Figura 2.16 ilustra o esquema da validação cruzada. Vale ressaltar que a cada nova iteração o conjunto de treino e validação é necessariamente diferente. Após percorrer as k iterações, o resultado final para um dado candidato é obtido pela média do erro nas k vezes. Apesar de todas as amostras do conjunto de dados serem usados tanto para treino quanto para teste em algum momento, a validação cruzada tem um custo computacional alto, o que a torna inviável quando se tem uma grande quantidade de hiperparâmetros ou um alto valor de k .

Figura 2.16 - Divisão do conjunto de treino em k partições, no qual a partição com a cor verde ilustra o conjunto de validação na k -ésima iteração e as partições em branco, o conjunto de treino.



Fonte: Produção do autor.

2.8 Extração e seleção de atributos

No aprendizado supervisionado espera-se que o algoritmo de AM aprenda com base no conjunto de treino uma relação linear, ou não linear, entre os dados de entrada e a saída alvo e que tal modelo, uma vez treinado, seja capaz de fazer uma boa generalização para dados não vistos. Neste contexto, a escolha apropriada dos dados de entrada do algoritmo é fundamental e é definida no processo de extração de atributos.

Os dados de entrada devem ser descritos por uma lista de atributos, conhecida como vetor de característica, com tamanho fixo e formada por valores numéricos que podem ser binários, categóricos ou contínuos. Formalmente, considere x um exemplo de dimensão d de um conjunto de dados, logo $x = (x_1, \dots, x_d)$, o processo de extração de atributos tem por objetivo transformar o dado bruto x em um vetor de característica x' de dimensão d' (GUYON et al., 2008).

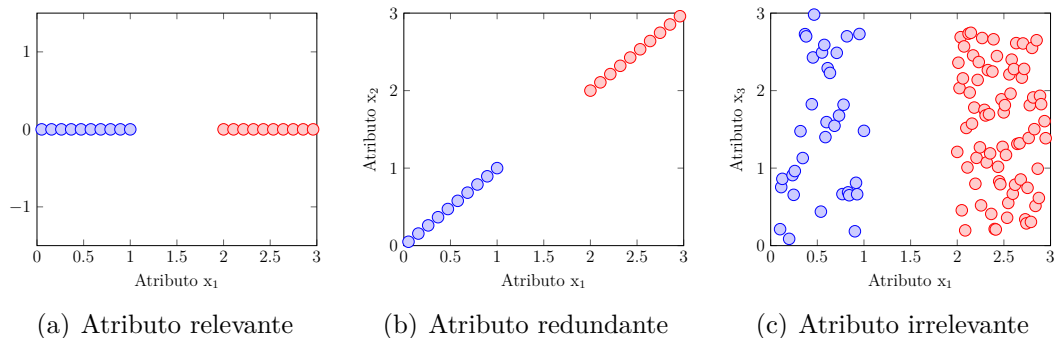
O processo de extração de atributos pode ser feito através de um especialista, que escolhe manualmente representações comuns no domínio da aplicação, ou através de um algoritmo de AM, sendo que em ambos casos visam-se obter o mínimo de perda de informação dos dados originais (NETZER et al., 2011). Contudo, nem todos os atributos do vetor de característica podem ser considerados representativos ou informativos. Enquanto a representatividade está relacionada ao processo de extração, a verificação do ganho de informação é considerada como seleção de atributos (SA), que pode ser categorizada como relevante, irrelevante ou redundante vide Figura

2.17 (KHALID et al., 2014).

De modo geral, a SA é classificada como uma técnica de pré-processamento que obedece a um determinado critério para eliminação de atributos irrelevantes e redundantes. Neste contexto, critérios comumente empregados utilizam a teoria da informação e estatística, por exemplo, a informação mútua e o coeficiente de correlação de Pearson (GUYON; ELISSEEFF, 2006).

A SA ditos como representativos e informativos contribui no desempenho do algoritmo de AM: (i) ao reduzir o consumo de memória para o armazenamento dos dados, (ii) tempo de treinamento, (iii) melhoria na interpretação e acurácia do modelo e (iv) visualização dos dados (MOTODA; LIU, 2002). Geralmente, ambas técnicas resultam na redução da dimensionalidade dos dados e por conta disso, podem causar a perda de informação caso o processo de extração produza atributos redundantes ou irrelevantes ao problema em predição, ou se a seleção ocorrer sobre um conjunto de atributos informativos que estejam pouco correlacionados entre si (CAI et al., 2018).

Figura 2.17 - Exemplo de atributo relevante, redundante e irrelevante. (a) x_1 é considerado como atributo relevante já que consegue discriminar bem as duas classes do problema. (b) x_2 é dito como atributo redundante devido a correlação alta com x_1 . (c) x_3 é categorizado como atributo irrelevante já que a separação das classes é feita completamente pelo x_1 .



Fonte: Adaptada de Li et al. (2017).

2.9 Seleção de atributos

A SA, bem como a extração, é uma área de estudo de pré-processamento de dados para a descoberta de conhecimento em banco de dados. O campo de estudo está em constante desenvolvimento desde da década 70 e tem por objetivo a ordenação, ou seleção de um subconjunto de atributos do vetor de características, com base

em um critério de importância que visa mensurar a redundância, irrelevância, ou ambos aspectos de um atributo. Como resultado de uma SA tem-se, por exemplo, a facilitação da interpretabilidade dos dados, eliminação de atributos custosos de serem calculados, remoção de dados ruidosos, redução do tempo de treinamento do algoritmo de aprendizagem e aumento da acurácia do modelo (LIU; MOTODA, 2012; CHANDRASHEKAR; SAHIN, 2014).

A formalização de um problema de SA é dado no seguinte modo: seja o subconjunto de atributos C' pertencente ao vetor de característica C , e $x_{C'}$ e x_C os atributos pertencentes a C' e C , respectivamente. O objetivo da SA é encontrar o subconjunto ótimo de atributos C' de dimensão d mínimo, tal que a distribuição de probabilidade da classe A dado $x_{C'}$, $P(A|y = f_{C'})$, seja aproximadamente igual a distribuição de probabilidade dos valores dos atributos x_C , $P(A|y = f_C)$. No geral, a qualidade de uma SA é medida com base nos resultados do algoritmo de AM e pode ser aplicada nos mais diversos cenários de dados, como no contexto supervisionado, não-supervisionado, ou semi-supervisionado, que serão abordados mais adiante (YU; LIU, 2004; KOLLER; SAHAMI, 1996).

Por exemplo, uma SA designada a realizar a ordenação dos atributos de C utiliza apenas como base um critério de importância, que define o mérito de cada atributo individualmente. Em seguida, os atributos são escolhidos por meio de um valor k , que se pode referir a uma quantidade fixa ou porcentagem de atributos que serão selecionados da lista ordenada, ou um limiar sobre o critério de importância. Uma outra alternativa de seleção do C' requer, além da medida de importância, uma estratégia de busca. Neste contexto, a estratégia de busca é o método responsável por encontrar o C' ótimo sobre todas as possíveis combinações. Por exemplo, a pesquisa exaustiva faz com que o método de SA examine todas as combinações possíveis C' para posteriormente selecionar a melhor. Contudo, na busca exaustiva, o número de combinações cresce exponencialmente de acordo com o tamanho do vetor de característica, d , o que na prática torna essa abordagem computacionalmente intratável para valores de $d \geq 20$ (KUMAR; MINZ, 2014).

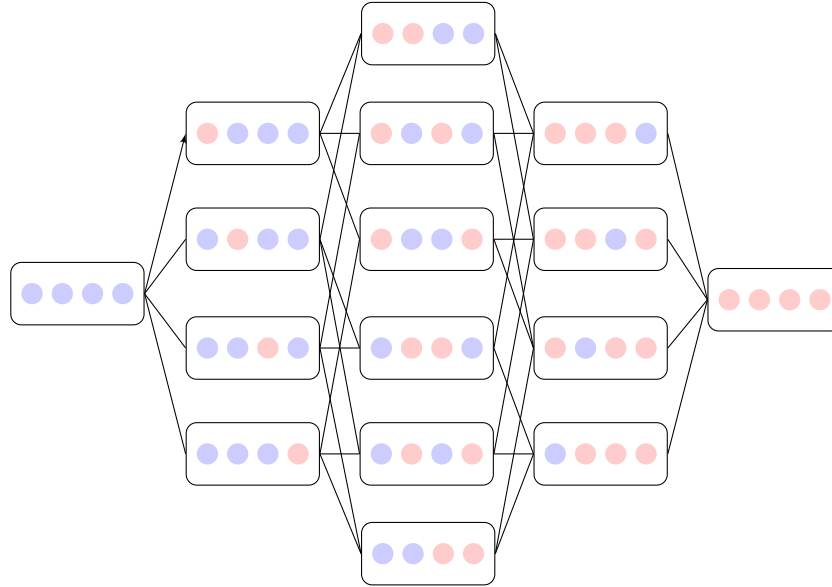
A estratégia de busca utiliza de uma heurística e abordagem gulosa para definição do C' ótimo, que leva em consideração os seguintes aspectos para a sua formulação: (i) direção de busca, (ii) organização de busca, (iii) critério de avaliação e (iv) critério de parada.

2.9.1 Direção de busca

A primeira questão a ser desenvolvida é o ponto de partida, que é responsável pela seleção dos subconjuntos do espaço de busca. A escolha do ponto de partida influencia diretamente na direção e emprega as operações de adição ou remoção que virá a ser utilizada na geração do C' sucessor (MOLINA et al., IEEE. 2002). Por exemplo, uma opção é começar C' como um conjunto vazio, calcular o critério de avaliação ao adicionar um atributo em C' e escolher o subconjunto que apresentar a solução ótima naquela iteração. Neste contexto, a direção da busca define a estratégia de geração do C' sucessor baseado no estado atual de C' , essas estratégias são:

- **Forward:** denominada também como seleção *forward*, a estratégia começa com um C' vazio de atributos e sucessivamente vai adicionando o atributo que produz a solução *ótima* em cada iteração, até atingir o critério de parada (MOLINA et al., IEEE. 2002). A Figura 2.18 ilustra o processo de seleção *forward* para um conjunto formado por quatro atributos, no qual o subconjunto mais à esquerda representa o estado inicial da técnica.
- **Backward:** ou eliminação *backward*, opera a busca do C' ótimo usando apenas a operação de remoção. A técnica começa inicialmente com o conjunto de atributos completo e a cada iteração um atributo é eliminado (MOLINA et al., IEEE. 2002). Na Figura 2.18, a eliminação *backward* pode ser retratada ao iniciar o processo no subconjunto mais à direita, com uma busca ocorrendo da direita para esquerda.
- **Bidirecional:** diferente dos métodos citados anteriormente, a abordagem bidirecional não é limitada apenas na utilização de uma direção no processo de busca. Neste caso, a busca é feita de maneira simultânea, uma em cada direção, até que a melhor solução seja encontrada ou ambas execuções atinjam o centro do espaço de busca. O ponto inicial da busca pode ser dado aleatoriamente (LIU; MOTODA, 2012).
- **Aleatória:** abordagem que não agrega o conceito de direção na busca e que realiza a geração do C' de forma aleatória. O ponto de partida, assim como o caso anterior, é designado em qualquer ponto do espaço de soluções. O método visa, sem nenhuma garantia, que a busca não fique presa em mínimos locais, devido à aleatoriedade inserida no processo (KUMAR; MINZ, 2014).

Figura 2.18 - Exemplo de espaço de soluções de atributos considerando $d = 4$. Uma seleção *forward* é ilustrada percorrendo o diagrama da esquerda para a direita, um percurso ao contrário do citado anteriormente irá resultar em uma eliminação *backward*.



Fonte: Adaptada de Langley et al. (1994).

2.9.2 Organização de busca

A organização de busca determina a estratégia usada para percorrer o espaço de todas as possíveis combinações de atributos, que pode vir a explorar cada C' existente por meio de uma busca exaustiva, ou empregando uma abordagem para a geração de candidatos factíveis através de heurísticas, por exemplo. Nem sempre uma organização de busca garante encontrar a solução global, dependendo do valor de d é desejável aplicar um método que requer um tempo de processamento menor e que produza uma solução local ótima.

De acordo com Liu e Motoda (2012), métodos de organização de busca podem ser agrupadas em três categorias distintas, estas são:

- **Busca completa:** essa abordagem garante que a solução global ótima sempre seja encontrada. Neste contexto, uma possível opção é a utilização de métodos exaustivos, como por exemplo, a busca em profundidade ou largura. Uma segunda alternativa, é percorrer parcialmente o espaço de soluções aplicando restrições que assegurem que a solução ótima seja alcançada. Um exemplo é o algoritmo *Branch-and-Bound*, que aplica a

propriedade de monotonicidade no critério de avaliação, ou seja, dado dois subconjunto C'_1 e C'_2 , sendo $C'_2 \subseteq C'_1$, a monotonicidade garante que o critério de avaliação de C'_1 seja sempre maior ou igual ao do C'_2 . Logo, se C'_1 não for um possível candidato a solução global ótima, C'_2 também não será (LIU; MOTODA, 2012).

- **Busca heurística:** Visa reduzir o tempo de execução, em relação à busca completa, ao utilizar uma heurística que não garante a solução ótima global para percorrer o espaço de soluções. Algoritmos comumente utilizados nesta abordagem são: *Best-First Search* (BFS), *beam* e *approximate branch-and-bound search* (LANGLEY et al., 1994).
- **Busca não-determinística:** difere dos métodos acima ao introduzir o conceito de não-determinismo na busca do C' , ou seja, dado as mesmas condições iniciais o algoritmo pode vir a produzir resultados diferentes. Por utilizar uma abordagem aleatória para a geração do C' , é necessário definir uma quantidade máxima de iterações. Além disso, essa abordagem não garante que o C' é ótimo, porém, assegura que o C' sucessor seja melhor que o C' atual. Exemplo de busca não-determinística são os algoritmos genéticos (LIU; MOTODA, 2012).

A Tabela 2.2 apresenta como a organização de busca está relacionada com a direção de busca. Por exemplo, a estratégia aleatória requer que C' sucessor seja gerado de maneira aleatória, assim como também, na busca exponencial não é possível assegurar a solução global ótima de forma aleatória. Já a busca heurística permite utilizar qualquer tipo de direção de busca.

Tabela 2.2 - Combinações das técnicas de organização de busca com direção de busca.

Organização de busca			
Direção de busca	Completa	Heurística	Aleatória
Seleção <i>forward</i>	Sim	Sim	Não
Eliminação <i>backward</i>	Sim	Sim	Não
Bidirecional	Sim	Sim	Não
Aleatória	Não	Sim	Sim

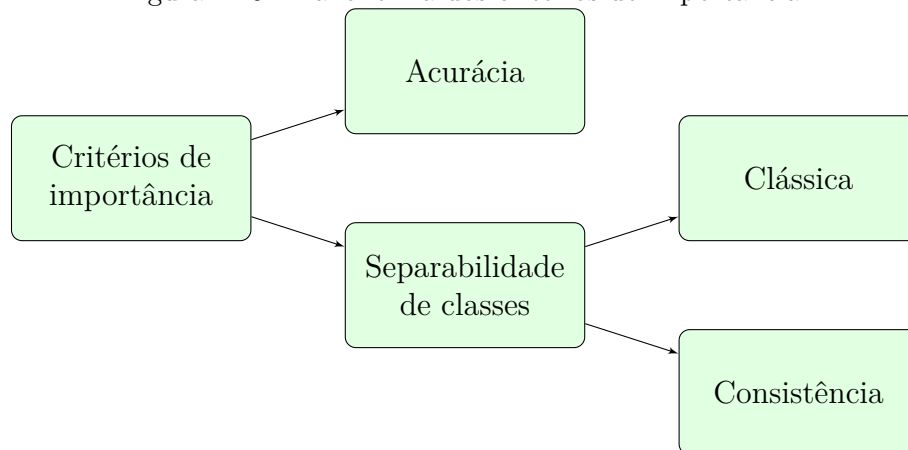
Fonte: Adaptada de Liu e Motoda (2012).

2.9.3 Critérios de avaliação de importância

Critérios de avaliação de importância são métricas utilizadas para quantificar a capacidade de um atributo, ou conjunto de atributos, em discriminar as classes de um problema. Nesse contexto, a SA é feita através da minimização ou maximização dos critérios de avaliação para a definição da solução ótima. O conceito de importância está relacionado com a relevância, que pode ser definida no seguinte modo: seja $C = (x_1, \dots, x_d)$, diz que $x_i \in C$ é um atributo relevante, se para duas amostras $x^{[1]}$ e $x^{[2]}$ com classe diferentes, x_i contribui para que $f(x^{[1]}) \neq f(x^{[2]})$ pela função-conceito f (KUMAR; MINZ, 2014).

No caso da seleção do subconjunto C' , tais critérios podem levar também em consideração a redundância. Normalmente, a redundância é referida como a correlação entre dois atributos. Dessa forma, dois atributos fortemente correlacionados são considerados como redundantes e um deles deve ser removido do subconjunto. A seguir serão descritos os critérios de avaliação de importância apresentados na taxonomia da Figura 2.19.

Figura 2.19 - Taxonomia dos critérios de importância.



Fonte: Adaptada de Liu e Motoda (2012).

- **Clássica:** Os métodos clássicos são englobados como critérios de avaliação de separabilidade de classe, ou seja, são métricas que se configuram como independentes de um algoritmo de AM. Esta abordagem é caracterizada por utilizar medidas estatísticas durante a SA, tais como:

a) **Medida de informação:** emprega o ganho de informação para men-

surar o quanto se aprende sobre uma determinada classe ao incluir um novo atributo no conjunto, podendo este conjunto ser vazio ou não. A quantidade de informação é calculada pela entropia de Shannon, que se baseia na probabilidade de ocorrência dos valores do atributo (DASH; LIU, 1997).

b) **Medida de distância:** quantifica a importância de um atributo considerando a sua capacidade de agrupamento entre exemplos pertencentes a mesma classe e separação de amostras de classes distintas, ou seja, calcula a distância interclasses e intraclasses de cada atributo. Geralmente, utiliza-se a distância Euclidiana para o cálculo dessa medida (KUMAR; MINZ, 2014).

c) **Medida de dependência:** baseia-se na dependência estatística entre atributo e classes para mensurar a importância de um atributo. Nesta abordagem, o coeficiente de correlação pode ser empregado para realizar a ordenação dos atributos mais relevantes e a eliminação de atributos redundantes (MOLINA et al., IEEE. 2002).

– **Medida de consistência:** é considerada também como um método de separabilidade de classes, porém, que visa reduzir a ocorrência de um ou mais pares de amostras, pertencentes a rótulos diferentes, com valor de atributo similar. As medidas de consistência são caracterizadas como monotônicas, isso permite, por exemplo, a realização de uma busca completa usando o método de *Branch-and-Bound*. Contudo, uma desvantagem desse tipo de abordagem é que atributos redundantes não são reconhecidos, já que a medida de consistência não consegue fazer uma distinção entre dois atributos bons (KUMAR; MINZ, 2014).

– **Medida de acurácia:** O processo de SA nesta abordagem é feita com um viés no algoritmo de AM. A medida é a acurácia obtida pelo modelo, no qual, comumente, o algoritmo utilizado é o mesmo aplicado para a classificação ou regressão do problema. A escolha do algoritmo é um fator crucial, já que o tempo necessário para o treinamento do mesmo pode se tornar uma limitação, como por exemplo, RNAs (LIU; MOTODA, 2012).

2.9.4 Critério de parada

O critério de parada é responsável por definir a condição de finalização na busca do subconjunto C' (KUMAR; MINZ, 2014). Normalmente, a parada é realizada quando uma das seguintes estratégias são atingidas:

- Número máximo de iterações
- Limiar baseado na acurácia do modelo
- Quantidade máximo de atributos em C'
- A adição ou remoção de atributos não traz melhoras significativas

2.9.5 Abordagens para seleção de atributos

Métodos que realizam a SA podem ser classificados em três tipos de abordagens: filtro, *wrappers* e *embedding*. As seções a seguir revisam as três abordagens citadas, sendo que a principal diferença entre tais abordagens é como o critério de avaliação é empregado para realizar a seleção.

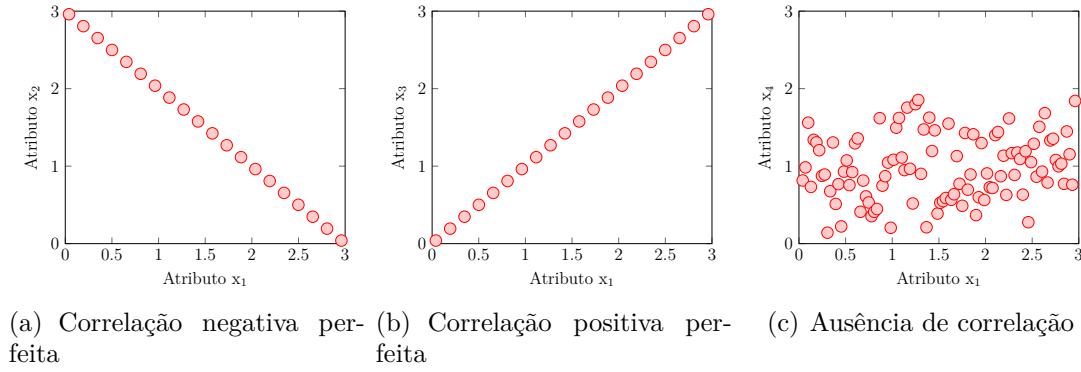
2.9.5.1 Abordagem filtro

A abordagem filtro contempla métodos baseados em medidas que não consideram a acurácia do modelo de AM. Essa abordagem é categorizada em dois tipos de análise, estes são: (i) filtros univariados e (ii) filtros multivariados. No primeiro caso, o método de filtro faz a ordenação das variáveis por meio de uma métrica que avalia um atributo por vez, podendo o cálculo considerar ou não a relação desse atributo com a saída alvo (LIU; MOTODA, 2012). Por exemplo, pode-se utilizar o coeficiente da correlação de Pearson, vide Equação 2.25, para fazer o ranqueamento dos atributos mais relevantes para a saída alvo ou utilizar da variância para remover atributos irrelevantes (GUYON; ELISSEEFF, 2003).

$$r = \frac{\text{cov}(x, y)}{\text{var}(x) \text{var}(y)}, \quad (2.25)$$

onde cov e var representam a covariância e variância, respectivamente. A correlação de Pearson mede a dependência linear entre duas variáveis com valores que variam entre -1 e 1, no qual valores próximos a -1 representam uma correlação negativa (Figura 2.20(a)), isso é, as duas variáveis movem-se em sentido oposto, valores próximos a 1 indicam uma correlação positiva (Figura 2.20(b)), ou seja, a variação entre as duas variáveis ocorre no mesmo sentido e valores próximos a 0 diz que não existe uma dependência linear entre as duas variáveis (Figura 2.20(c)) (GUYON; ELISSEEFF, 2003).

Figura 2.20 - Exemplos de casos extremos da correlação de Pearson.



Fonte: Produção do autor.

Como dito anteriormente, tanto a correlação de Pearson quanto a variância podem ser usados para fazer o ranqueamento de atributos. Desse modo, uma vez feito o ranqueamento, utiliza-se como critério de seleção os n -ésimos primeiros atributos do ranque ou um limiar baseado no critério de avaliação. Contudo, uma desvantagem dessa abordagem é que a remoção de atributos redundantes não é realizada, já que se avalia apenas uma variável por vez.

Neste caso, a eliminação de variáveis redundantes é dada pelos filtros multivariados. Métodos multivariados utilizam métricas que consideram tanto a correlação entre pares de atributos quanto entre atributos e saída alvo (LIU; MOTODA, 2012). Pode-se citar como exemplo o método *Correlation-based Feature Selection* (CFS), que será apresentado posteriormente. Geralmente, alguns filtros multivariados demandam uma estratégia de busca para fazer a seleção do subconjunto, o que faz o processo ser mais demorado quando comparado ao filtro univariado (HALL, 1999).

Figura 2.21 apresenta de maneira simplista o esquema utilizado na abordagem filtro no processo de SA, no qual o filtro pode ser um qualquer critério de avaliação, como a correlação de Pearson, por exemplo.

2.9.5.2 Abordagem *wrapper*

A abordagem *wrapper* destaca-se por empregar um algoritmo de AM na avaliação do subconjunto de atributos (MOLINA et al., IEEE. 2002). Em específico, este método utiliza de uma estratégia de busca, no qual a cada iteração retorna-se um subconjunto de atributos que é avaliado pelo desempenho do algoritmo de AM. Dentre os critérios de parada, utiliza-se geralmente a acurácia do modelo como um limiar ou

na verificação de uma melhora significativa do modelo em até t iterações, com um número máximo de iterações pré-definido pela busca no espaço de todos os subconjuntos de atributos (TANG et al., 2014).

Figura 2.21 - Abordagem Filtro.



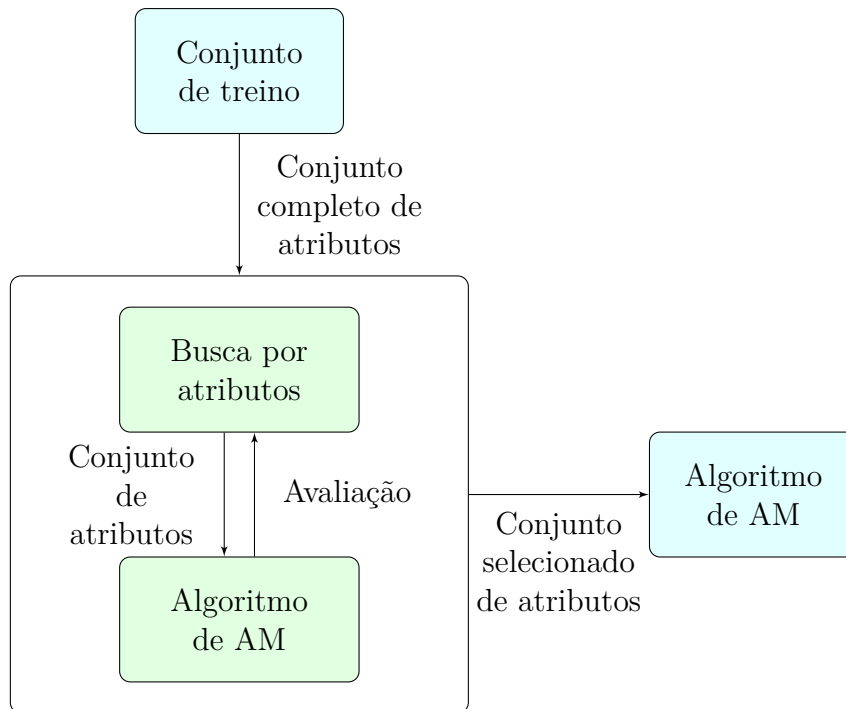
Fonte: Adaptada de Liu e Motoda (2012).

Espera-se que métodos *wrapper* apresentem resultados melhores em relação à abordagem filtro devido ao viés ao classificador ou regressor inserido no processo de seleção. Entretanto, uma desvantagem da abordagem é o alto custo computacional, já que tal método depende do treinamento do algoritmo de AM para a avaliação dos subconjuntos de atributos (KOHAVI et al., 1997; TANG et al., 2014). Figura 2.22 ilustra o processo de seleção utilizado na abordagem *wrapper*.

2.9.5.3 Abordagem *embedding*

Na abordagem *embedding* a SA ocorre naturalmente como parte do algoritmo de AM, ou seja, a escolha do subconjunto é feita durante a construção do modelo na fase de treinamento (TANG et al., 2014; ANG et al., 2015). Exemplos de algoritmos que contêm esse tipo de característica são: (i) árvore de decisão, que realiza a seleção pelos nós que compõe a árvore; (ii) LASSO, removendo os atributos que tiveram os coeficientes iguais a 0; (iii) RNAs, removendo atributos que tiveram o módulo do peso próximo a 0. Figura 2.23 apresenta o esquema de seleção utilizado na abordagem *embedding*.

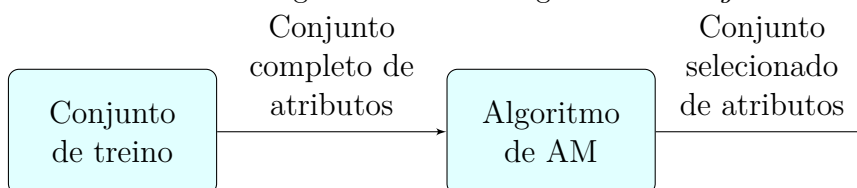
Figura 2.22 - Abordagem *Wrapper*.



Fonte: Adaptada de Tang et al. (2014).

Por exemplo, no caso das redes neurais artificiais, pode-se usar os pesos de uma rede *feedforward* de camada única como medida para a SA. No processo de treinamento, espera-se que a rede aprenda os pesos que maximizam a acurácia da predição. Dessa forma, os pesos podem ser interpretados como o quanto um determinado atributo contribui na predição da saída alvo, logo, pesos próximos a 0 indicam uma contribuição fraca no cálculo da saída desejada e por conta disso, podem ser removidos. Além disso, como os pesos podem assumir valores negativos, utiliza-se o módulo para fazer o ranqueamento dos atributos mais relevantes.

Figura 2.23 - Abordagem *Embedding*.



Fonte: Adaptada de Ang et al. (2015).

2.10 Paradigmas em seleção de atributos

Métodos de SA, assim como os algoritmos de AM, podem ser categorizados em três paradigmas de aprendizado, este são: (i) supervisionado, (ii) não-supervisionado e (iii) semi-supervisionado. Neste trabalho, a seleção supervisionada foi empregada para encontrar o subconjunto ótimo de atributo já que a saída alvo do problema é conhecida.

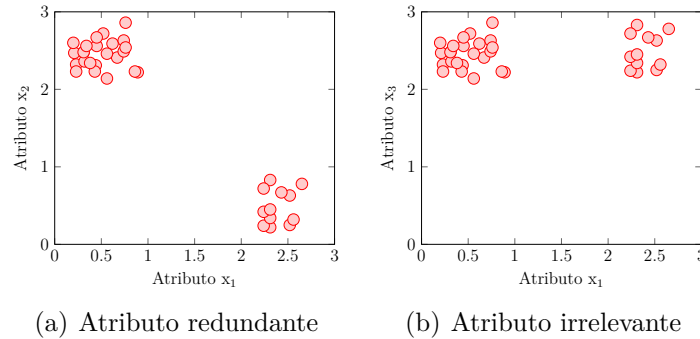
- **Seleção Supervisionada:** a seleção supervisionada é dada quando o conjunto de dados tem *a priori* a informação a respeito dos rótulos do problema, o que traz como vantagem a procura do subconjunto de atributos que contém uma maior relação com a saída desejada. Métodos neste paradigma visam maximizar a acurácia do algoritmo de AM ou a correlação entre atributos e as classes do problema, como por exemplo, o método CFS que busca encontrar o subconjunto de atributos que minimize a correlação entre atributo e atributo e maximize a correção entre atributo e saída desejada (CAI et al., 2018).
- **Seleção Não Supervisionada:** neste paradigma a informação sobre o rótulo dos dados não é conhecida. Uma possível abordagem é a utilização de métodos *wrapper* para a clusterização dos subconjuntos de atributos gerados pela estratégia de busca e consequentemente, a escolha daquele que apresentar os resultados mais interessantes com base no critério de avaliação (MITRA et al., 2002; DY; BRODLEY, 2004). A detecção de atributos redundantes e irrelevantes pode ser feita baseado no resultado de clusterização, como ilustrado na Figura 2.24.

Uma segunda opção é a utilização de abordagens filtro, como a seleção pela variância, que fora explicado anteriormente, ou o desvio absoluto (DA) (BHARTI; SINGH, 2014; FERREIRA; FIGUEIREDO, 2012). O DA é um método univariado que calcula a média da diferença dos valores pelo valor médio da variável vide Equação 2.26. Neste método, atributos com valores de DA altos são ditos como mais discriminativos.

$$DA_i = \frac{1}{n} \sum_{j=1}^n |x_i^{[j]} - \bar{x}_i| , \quad (2.26)$$

no qual DA_i refere-se ao i -ésimo atributo, $\bar{x}_i$ representa a média da variável x_i , e n é o número total de valores em x_i .

Figura 2.24 - Exemplos de atributo redundante e irrelevante. (a) x_2 é considerado como atributo redundante já que consegue discriminar o mesmo *cluster* de x_1 . (b) x_3 é dito como atributo irrelevante já que a sua eliminação não modifica a discriminação dos *clusters*.



Fonte: Adaptada de Dy e Brodley (2004).

- **Seleção Semi-supervisionada:** o termo semi-supervisionada refere-se a um conjunto de dados composto por amostras rotuladas e não rotuladas, no qual, geralmente, a quantidade de dados não rotulados é maior que os rotulados. A SA neste cenário é, em sua maioria, dada pela utilização de técnicas que buscam construir um preditor baseado nos dados rotulados para a predição das classes dos dados não rotulados (CHAPELLE et al., 2009; PISE; KULKARNI, 2008). Pode-se citar como exemplo, técnicas como *self-training* e *co-training* empregadas na abordagem *wrapper* durante o treinamento do algoritmo (SHEIKHPOUR et al., 2017).

As abordagens *self-training* e *co-training* são métodos de aprendizado classificados como indutivo. No caso do *self-training*, utiliza-se um classificador para prever a saída das amostras não rotuladas. Em seguida, as predições mais confiantes são adicionadas ao conjunto de dados rotulados e o modelo é retreinado. Este processo é repetido até que todo o conjunto não rotulado esteja vazio (PISE; KULKARNI, 2008).

O método *co-training* utiliza dois classificadores para prever a classe dos dados não rotulados, sendo que o algoritmo de AM não necessariamente seja o mesmo para ambos modelos. Além disso, o treinamento dos algoritmos é realizado com conjunto de dados rotulados distintos e uma vez que os modelos estejam treinados, as predições ditas como as mais confiantes de um classificador são adicionadas no conjunto de dados do classificador oposto e vice-versa. Em seguida, os modelos são retreinados sobre o novo

conjunto e este processo é repetido até que um critério de parada seja atingido (CHAPELLE et al., 2009).

2.11 Correlation-based feature selection

Correlation-based Feature Selection (CFS) é um método filtro, supervisionado e multivariado, que avalia um subconjunto de atributos C' usando uma métrica $M_{C'}$, que considera a correlação entre atributos para a minimização da redundância e correlação entre atributo e saída desejada para a maximização de atributos relevantes vide Equação 2.27 (HALL, 1999; ZHAO et al., 2010).

$$M_{C'} = \frac{k\overline{corr_{x,y}}}{\sqrt{k + k(k-1)\overline{corr_{x,x}}}}, \quad (2.27)$$

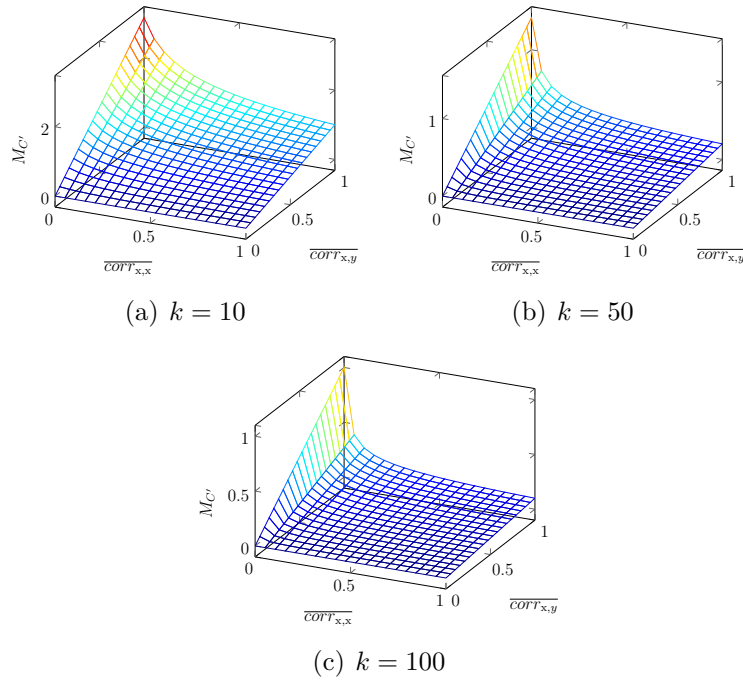
no qual k é a quantidade de atributos no subconjunto C' , $\overline{corr_{x,y}}$ é a média das correlações entre atributo e saída desejada e $\overline{corr_{x,x}}$ é a média entre todos os pares de atributos. Dessa forma, $\overline{corr_{x,y}}$ pode ser interpretado como o quão fácil é predizer a saída alvo baseando-se nos atributos em C' e $\overline{corr_{x,x}}$ como um indicador da redundância existente entre tais atributos. Figura 2.25 apresenta o comportamento de $M_{C'}$ para diferentes valores de $\overline{corr_{x,y}}$ e $\overline{corr_{x,x}}$, é possível notar que quanto maior o valor de $M_{C'}$ mais informativo e menos redundante são os atributos em C' . Contudo, ao adicionar mais atributos fortemente correlacionados com a saída alvo, espera-se que a correlação entre atributos também aumente, logo, atingir o valor máximo de $M_{C'}$ é uma tarefa difícil.

Normalmente, utiliza-se a informação mútua para dados categóricos e a correlação de Pearson em dados contínuos como medida de dependência estatística entre as variáveis. A informação mútua permite capturar qualquer tipo de relação linear ou não linear e assume qualquer valor igual ou maior que zero, no qual se diz que duas variáveis são estatisticamente independentes caso a informação mútua entre ambas seja igual a 0. Para garantir que o valor da informação mútua tenha o mesmo peso no cálculo da média, é feita uma normalização sobre o valor da correlação considerando a soma da entropia das duas variáveis (HALL, 1999).

Para percorrer o espaço de soluções geralmente utiliza-se o BFS, seleção *forward* ou a eliminação *backward*. O BFS é um método que verifica o subconjunto mais promissor a cada iteração por meio de uma heurística, neste caso a métrica $M_{C'}$. O estado inicial do BFS pode ser definido com um conjunto vazio ou completo, no

qual a cada iteração o melhor atributo é adicionado ou removido do subconjunto, respectivamente. Além disso, em alguns casos onde o resultado da busca começa a não ser promissor, o método retrocede para o próximo subconjunto mais promissor da iteração anterior. Como BFS é um método que percorre todo o espaço de soluções é necessário definir um critério de parada, no caso do CFS o método irá finalizar a busca caso $M_{C'}$ não apresente uma melhora no resultado em 5 iterações (HALL, 1999).

Figura 2.25 - Comportamento da métrica $M_{C'}$ para diferentes valores de $\overline{corr}_{x,y}$, $\overline{corr}_{x,x}$ e k .



Fonte: Adaptada de Hall (1999).

2.12 Filtro por correlação

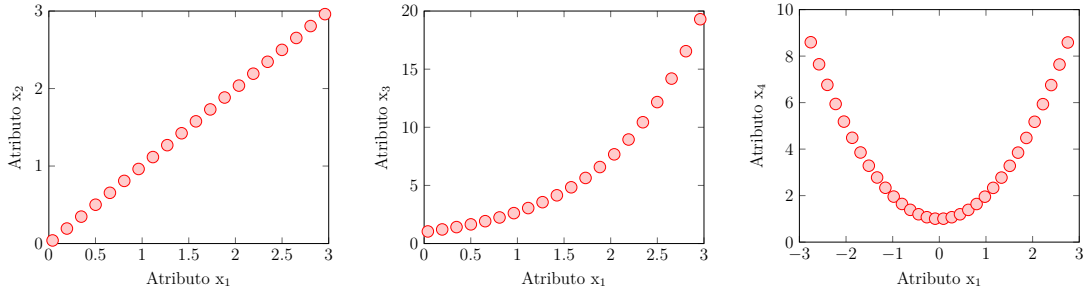
Dado um conjunto de atributos C , os métodos baseados em medidas de correlação identificam (i) a relevância de um atributo x_i considerando correlações $corr_{x_i,y}$ altas, e (ii) a redundância verificando correlações entre pares de atributos $corr_{x_i,x_j}$ altas. Além disso, diz-se que um atributo tem correlação $corr_{x_i,y}$ predominante se $|corr_{x_i,y}| \geq |corr_{x_j,y}| \forall x_i, x_j \in C$ (BIESIADA; DUCH, 2007).

Neste sentido, a SA dado pelo método filtro por correlação é iniciado criando uma lista L que contém os atributos ordenados de maneira decrescente sobre a $corr_{x_i,y}$.

Em seguida, é eliminado de L os atributos redundantes respeitando o seguinte critério $|corr_{x_i, x_j}| \geq \delta$, no qual δ é um liminar de correlação que indica quando considerar x_i e x_j como redundante. A ordenação pela relevância garante que os atributos x_i predominantes sejam mantidos no processo de eliminação da redundância, de tal maneira que ao encontrar um $|corr_{x_i, x_j}| \geq \delta$, x_j deverá ser removido se $corr_{x_i, y} \geq corr_{x_j, y}$ e caso contrário, x_i será removido.

Neste trabalho, utilizamos a correlação de Pearson e Spearman para medir relações lineares e monotônicas, respectivamente. Uma relação monotônica entre duas variáveis x_i e x_j pode ser categorizada como (i) crescente, se x_i incrementa x_j também incrementa e (ii) decrescente, se x_i incrementa x_j decrementa. Figura 2.26 ilustra relações monotônicas lineares, não lineares e não monotônicas. No caso da Figura 2.26(b) e 2.26(c), é possível perceber que apesar de existir uma correlação forte entre as duas variáveis a correlação de Pearson não refletiria um alto valor de coeficiente por capturar apenas relações lineares.

Figura 2.26 - Exemplos de relações monotônicas e não monotônicas.



(a) Relação monotônicas linear (b) Relação monotônicas não linear (c) Relação não monotônicas

Fonte: Produção do autor.

O coeficiente de correlação de Spearman ρ , ou coeficiente de correlação de postos, tem uma interpretação similar ao da correlação de Pearson, ou seja, valores próximos a 1 representa uma correlação positiva, valores próximos a -1 indica correlação negativa e valores próximos a 0 mostra que os dados não estão correlacionados (KOSKA; ZWILLINGER, 2000). Considere uma variável $x_i = (x_i^{[1]}, \dots, x_i^{[n]})$, a posição dos elementos de x_i quando x_i está organizado em ordem crescente define o valor do posto k de cada um dos seus elementos. Em casos onde existem valores de postos iguais $x_i^{[k]} = x_i^{[k+1]}$, deve-se atribuir um novo valor de posto para os dois elementos usando a média dos postos de $x_i^{[k]}$ e $x_i^{[k+1]}$, isso é, $\frac{(k+k+1)}{2}$. O coeficiente ρ pode ser

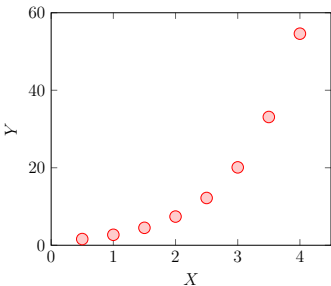
calculado através do coeficiente de correlação de Pearson sobre os dados convertidos em posto ou por meio da Equação 2.28.

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \, , \tag{2.28}$$

no qual d é a distância entre os postos de duas variáveis x_i e x_j , e n é a quantidade de amostra. Figura 2.27 apresenta o processo para o cálculo da correlação de Spearman para duas variáveis com dependência monotônica crescente, $Y = e^X$. É possível notar que $\rho = 1$ quando a distância entre todos os postos é nula, ou seja, X e Y tem os mesmos postos.

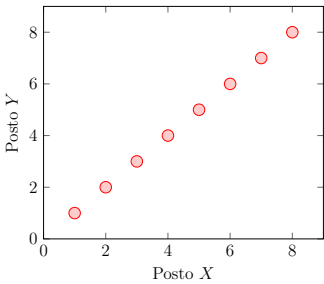
Figura 2.27 - Exemplos do cálculo para correlação de Spearman.

	X	Y
1	0.5	1.6
2	1	2.7
3	1.5	4.5
4	2	7.4
5	2.5	12.2
6	3	20.1
7	3.5	33.1
8	4	54.6



(a) Variáveis X e Y (b) Visualização das variáveis X e Y . $r = 0.9057$ e $\rho = 1$

	X	Posto X	Y	Posto Y
1	0.5	1	1.6	1
2	1	2	2.7	2
3	1.5	3	4.5	3
4	2	4	7.4	4
5	2.5	5	12.2	5
6	3	6	20.1	6
7	3.5	7	33.1	7
8	4	8	54.6	8



(c) Cálculo dos postos (d) Visualização dos postos da variáveis X e Y . Neste caso, $r = \rho = 1$

Fonte: Produção do autor.

2.13 *Random forest* para seleção de atributos

RF é um método *ensemble* aplicado tanto em problemas de classificação quanto em regressão. O algoritmo constrói k árvores de decisão baseando-se em (i) amostras *bootstrap* $\mathcal{D}'$ do conjunto de treino $\mathcal{D}$ com reposição, no qual a quantidade de amostra em $\mathcal{D}'$ deve ser igual em $\mathcal{D}$ e (ii) usando uma quantidade de atributos em $\mathcal{D}'$ menor que em $\mathcal{D}$ para cada nó da árvore. Cada árvore realiza o mapeamento da entrada para sua respectiva saída e um esquema de voto é utilizado para combinar a resposta de todos os classificadores fracos. Desse modo, RF apresenta-se como uma técnica similar ao método *bagging*, exceto pelo fato da aleatoriedade inserida na definição dos atributos utilizados em cada nó (BREIMAN, 2001; GENUER et al., 2010).

Como a árvore de decisão usa um critério de importância para avaliar os atributos na divisão de um nó da árvore, a SA em RF é dado pela média da importância de cada atributo em todas as árvores de decisão treinadas no método. A importância de uma variável é medida usando o conjunto de generalização chamado *out-of-bag*, que é composto pelas amostras $\mathcal{D} \setminus \mathcal{D}'$. Com base nesse critério, faz-se um ranqueamento para selecionar os k atributos mais importantes, no qual quanto maior o valor é, mais importante é o atributo (LOUPPE, 2014).

Por realizar a SA no treinamento das árvores, RF é considerada como uma abordagem *embedding*. A combinação do resultado de várias árvores traz como vantagem a redução na variância do resultado final e a aleatoriedade inserida no processo de criação das árvores garante uma menor correlação entre as árvores da floresta e a oportunidade de investigar atributos que não seriam usados na construção da árvore, caso fosse mantido o atributo que sempre apresenta uma maior correlação com a saída alvo no conjunto de treino (LOUPPE, 2014).

2.14 *Least absolute shrinkage and selection operator* para seleção de atributos

LASSO é um algoritmo de AM que emprega o termo de regularização ℓ_1 em um modelo de regressão linear. Isto é, o algoritmo visa minimizar a soma quadrática da diferença entre a saída esperada e predita acrescido com a penalidade da soma absoluta dos coeficientes. O termo de regularização é acompanhado pelo hiperparâmetro λ que determina o quanto a regularização vai impactar no modelo, no qual para $\lambda = 0$ a regularização é desconsiderada e, portanto, tem-se uma regressão linear. Caso λ seja muito grande, a flexibilidade do modelo é restringida (FONTI; BELITSER, 2017).

Dessa forma, a regularização desempenha o papel de penalizar ou encolher os coeficientes com o objetivo de diminuir a complexidade do modelo e nos casos onde o valor do coeficiente é igual a zero, entende-se que tais atributos são considerados como irrelevantes pelo algoritmo. Com isso, é possível utilizar LASSO para realizar a SA por meio da eliminação das variáveis que obtiveram o coeficiente exatamente igual a zero ([TANG et al., 2014](#)).

3 APRENDIZADO DE MÁQUINA E CIÊNCIA DOS MATERIAIS

3.1 Conceitos iniciais

Desde a revolução industrial, a busca pela automação de processos é bastante explorada em diversos campos de estudo, em específico, em atividades químicas voltadas para a área da medicina, biologia e ciência dos materiais ([ASPURU-GUZYK; PERSSON, 2018](#)). Neste contexto, plataformas robóticas apresentaram resultados superiores em relação a atividades manuais e por conta disso, algumas dessas aplicações são consideradas como estado da arte ([PRABHU; URBAN, 2017](#)).

Contudo, em ciência dos materiais, a automação de processos quando relacionada ao descobrimento de novos materiais ainda é constituída por métodos tradicionais compostos por experimentos laboratoriais ou simulações computacionais, que em sua maioria demandam tempo e recursos para execução dos cálculos das propriedades do material em estudo ([LIU et al., 2017](#)).

A descoberta de novos materiais está relacionada ao desenvolvimento de sistemas moleculares com propriedades físico-químicas específicas ([RAJAN, 2005](#)). O estudo dessas propriedades no que lhe concerne é dada pela Mecânica Quântica (MQ) que, entre outros fatores, difere da mecânica clássica por sua natureza probabilística ([MATTA; GILLESPIE, 2002](#)).

As propriedades físico-químicas de um material são obtidas por meio de funções de onda, devido ao movimento ondulatório que governa os sistemas moleculares. Uma interpretação estatística simplista da função da onda, pode ser feita ao assumir uma função de densidade de probabilidade de encontrar um elétron em um determinado ponto em coordenadas cartesianas. A teoria quântica faz uso da equação de Schrödinger para descrever o comportamento de partículas subatômicas, através das suas funções de onda, que são soluções da equação ([EISBERG; RESNICK, 1979](#); [TIPLER, 1994](#)).

A solução da equação de Schrödinger para um sistema molecular fornece a função de onda que contém, a princípio, toda a informação a respeito da propriedade da molécula. No entanto, o cálculo dessa equação é altamente custoso, pois seu custo computacional escala com o número de elétrons (N) no sistema ([GRIFFITHS, 2010](#)). De modo a reduzir esse custo computacional, algoritmos foram propostos usando métodos matematicamente mais sofisticados ([LEDOUX et al., 2005](#)).

Em 1964, Pierre Hohenberg e Walter Kohn demonstraram que toda informação

da função de onda do sistema está contida na densidade eletrônica e, com isso, propuseram uma abordagem para aproximar a distribuição dos elétrons nos átomos, conhecida como DFT. A DFT viabilizou o cálculo de estruturas maiores, com maior número de átomos, como também foi capaz de manter o poder descritivo e acurácia satisfatória (BURKE, 2012). Tipicamente a DFT tem uma complexidade de tempo que cresce na ordem $\mathcal{O}(N^3)$ (RYCZKO et al., 2018).

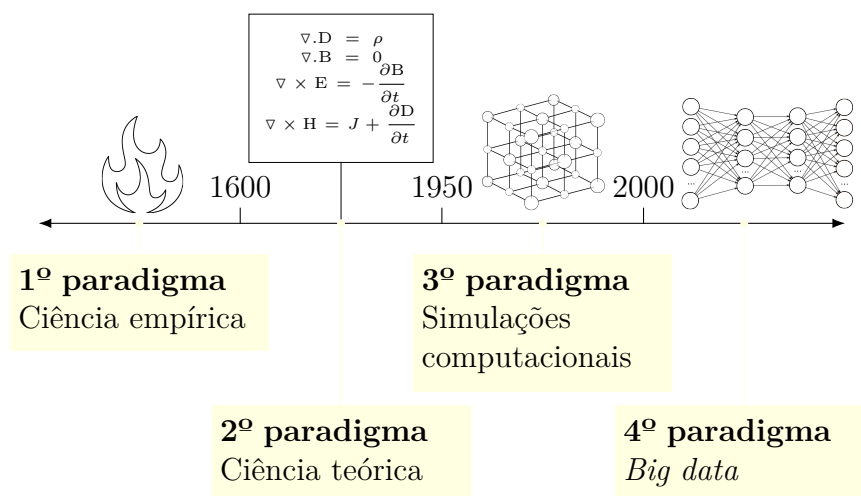
Outras abordagens, como experimentos laboratoriais são usadas para medição das propriedades desejadas, porém, neste caso, exige-se recursos físicos devido à necessidade de profissionais especializados na manipulação dos componentes e laboratórios com o ambiente adequado para a operação do experimento (NOSENGO, 2016).

Abordagens apresentadas anteriormente são consideradas como o terceiro paradigma na ciência dos materiais. O registro dos resultados dessas simulações computacionais, constituiu e motivou o conceito do quarto paradigma, o *big data*, que possibilita aplicações de métodos de AM para predição de propriedades ou clusterização de moléculas similares (AGRAWAL; CHOUDHARY, 2016). Por exemplo, em 2008, o banco de dados público PubChem¹ continha informações a respeito de 25 milhões de compostos químicos. Em 2018, essa quantia aumentou para 96 milhões (ZHU, 2020). Figura 3.1 apresenta os 4 paradigmas na ordem cronológica no tempo. No primeiro deles, as descobertas de materiais eram feitas por meio da ciência empírica, enquanto no segundo através de modelos teóricos. A necessidade de novos métodos para a aceleração da descoberta de materiais impulsionou o uso de computadores a partir da década de 50.

Além disso, as DFTs produzem resultados correlacionados quando aplicadas em grupos de moléculas similares (RUPP, 2015). Com isso, a exploração dessa correlação pode ser dada ao aplicar algoritmos de AM que aprendam a encontrar algum tipo de padrão entre a estrutura da molécula e a propriedade desejada. Figura 3.2 ilustra como o AM pode ser aplicado para a predição das propriedades moleculares. Por meio dos cálculos das simulações quânticas, pontos de exemplos são criados, dessa forma modelos de regressão podem ser usados para verificar uma relação de dependência entre os descritores escolhidos para descrever a molécula e sua propriedade gerada na simulação e, com isso, extrapolar para dados não vistos (BARKER et al., 2016; RUPP, 2015).

¹<https://pubchem.ncbi.nlm.nih.gov>

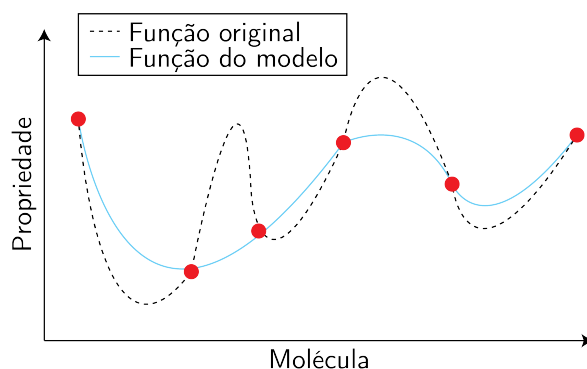
Figura 3.1 - Os 4 paradigmas em ciência dos materiais.



Fonte: Adaptada de Agrawal e Choudhary (2016).

Contudo, para que o AM seja apropriado na aplicação em ciência dos materiais, é preciso que os resultados gerados tenham uma boa precisão. Isso dependerá de alguns fatores, esses são: (i) definição da propriedade a ser predita; (ii) capacidade de aprendizagem do modelo escolhido; (iii) conjunto de descritores definidos para a representação da molécula. Dessa forma, uma vez treinado o algoritmo, o cálculo da propriedade de uma nova molécula pode ser realizado em questões de segundos (RAMPRASAD et al., 2017; WARD; WOLVERTON, 2017).

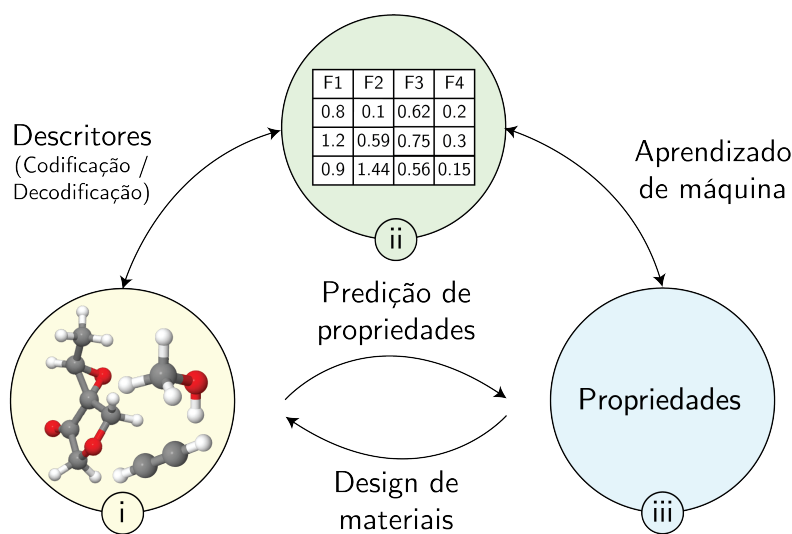
Figura 3.2 - Linha tracejada apresenta a função original que governa a propriedade. Os pontos vermelhos são os resultados obtidos por simulações quânticas (DFT, por exemplo). Linha azul mostra a função aproximada encontrada pelo algoritmo de AM. Note que este exemplo é um caso de *underfitting*, já que as amostras não descrevem adequadamente a função original do problema.



Fonte: Adaptada de Rupp (2015).

De modo geral, a aplicação da AM vem mostrando diversas oportunidades para soluções de problemas na área de ciência dos materiais. A abordagem previamente apresentada, baseia-se em moléculas predefinidas para a predição da propriedade. Contudo, novos estudos propõem métodos para encontrar moléculas tendo como entrada a propriedade almejada (SANCHEZ-LENGELING et al., 2017). Figura 3.3 apresenta ambos métodos, conhecidos como abordagem direta e *design* inverso, ou *design* de materiais, respectivamente.

Figura 3.3 - AM em ciência dos materiais para a predição de propriedade e geometria molecular.



Fonte: Produção do autor.

3.2 Estrutura molecular

Na química, é possível descrever uma molécula em termos qualitativos por uma representação da estrutura molecular. Geralmente, esse tipo de representação não contém toda informação real da molécula, pois, tem por objetivo ser simples para facilitar a interpretação humana e o seu armazenamento em arquivos químicos (MAURI et al., 2017).

A descrição da estrutura molecular visa ser única e exclusiva, isto é, para uma mesma molécula deve-se gerar sempre uma única saída e para moléculas diferentes deve-se obter saídas distintas. Uma representação bem comum é a fórmula química, que contém informações a respeito da ocorrência e dos tipos de elementos químicos, porém, é possível encontrar na literatura outras representações que variam desde

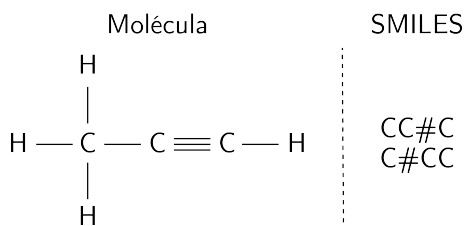
notações textuais até modelagens em grafos (POLANSKI; GASTEIGER, 2017).

No caso da fórmula química, a informação a respeito das conexões dos átomos não é relevada. Em 1988, Weininger propôs uma representação textual, de fácil cálculo, para codificação do grafo da molécula, denominada como *Simplified Molecular Inputs Line Entry System* (SMILES), sendo possível armazenar detalhes a respeito das conexões baseando-se nas seguintes regras (WEININGER, 1988):

- Átomos são representados pelos seus símbolos atômicos;
- A omissão do átomo de hidrogênio é opcional;
- Ligação simples, dupla, tripla e aromática são representadas por -, =, # e :, respectivamente. Sendo que, ligações simples e aromáticas podem ser omitidas;
- Ramificações são especificados usando o parêntese;
- Abertura e fechamento de um anel são indicados através de um dígito;
- Descreve-se os átomos de um anel aromático em letra minúscula.

SMILES é bastante aplicada devido a simplicidade das regras, porém, tal representação permite gerar saídas distintas para uma mesma entrada (POLANSKI; GASTEIGER, 2017). Figura 3.4 exemplifica este problema com a molécula de Propino, no qual é possível obter 2 saídas diferentes mudando apenas a forma de leitura da molécula. Caso a leitura seja feita da esquerda para direita, a saída será CC#C e C#CC, caso a leitura tenha sido feita da direita para esquerda.

Figura 3.4 - No processo de codificação do grafo da molécula para SMILES foram omitidos os átomos de hidrogênio e as ligações simples.



Fonte: Produção do autor.

Uma terceira opção é a representação da molécula através de um grafo não dirigido, no qual os átomos e as ligações são os vértices e arestas, respectivamente. Comumente, tais grafos costumam ser ponderados (associando pesos aos tipos de ligações) e esparsos, devido à quantidade de átomos presente no sistema molecular ser bem maior do que o número de ligações (MAURI et al., 2017).

A representação computacional de um grafo é feita por meio da matriz ou lista de adjacência. No primeiro caso, define-se uma matriz quadrada $A = [a_{ij}]$, de tamanho $n \times n$, onde n é quantidade de vértices do grafo, no qual $a_{ij} = 1$ se existir uma aresta entre um par de vértices (a_i, a_j) e $a_{ij} = 0$, caso contrário. A lista de adjacência, por outro lado, é formada por uma lista encadeada para cada vértice V , em que para cada lista encadeada de V será armazenada todos os vizinhos pertencentes àquele vértice (NETTO, 2003). O tamanho da matriz e da lista de adjacência é quadrático e linear, respectivamente, em relação à quantidade de vértices. Dessa forma, a representação da estrutura molecular por SMILES tem por vantagem reduzir o espaço de armazenamento requerido, além de permitir recuperar o grafo original da molécula.

Outra forma de representação é o formato XYZ, que permite salvar a distribuição espacial dos átomos da molécula, em um documento semelhante a uma planilha, composto por $d + 2$ linhas, sendo d a quantidade de átomos na molécula. Geralmente, o arquivo gerado apresenta o número de elementos químicos na primeira linha, seguido por uma observação a respeito da molécula na linha de baixo. Posteriormente, para cada linha d é armazenada as informações sobre o elemento químico e sua coordenada cartesiana, conforme mostrado na Figura 3.5. Vale ressaltar que neste método não é considerado a ordem dos átomos entre as linhas do arquivo.

Figura 3.5 - Representação das informações sobre a distribuição espacial da molécula de água no formato XYZ. Primeira e segunda linha apresentam a quantidade de átomos e nome do composto, respectivamente.

3			
Água			
H	0.7493682	0.0000000	0.4424329
O	0.0000000	0.0000000	-0.1653507
H	-0.7493682	0.0000000	0.4424329

Fonte: Produção do autor.

3.3 Descritores

Geralmente, as representações da estrutura molecular apresentam uma relação difícil e complicada de aprender sobre a propriedade desejada. Isso porquê, essas representações podem carregar diversas informações que não dizem nada a respeito do problema tratado (WARD et al., 2018). Neste sentido, a geração de descritores é considerada como um passo crucial (SWANN et al., 2018), pois permite transformar a estrutura da molécula em um dado que seja sempre numérico e adequado para o algoritmo de AM (GHIRINGHELLI et al., 2015). Em razão disso, é importante que um descritor seja (FABER et al., 2015; WARD; WOLVERTON, 2017):

- Completo, a informação contida no descritor deverá ser representativa ao problema;
- Único, o descritor deverá gerar sempre a mesma saída para uma mesma molécula, por exemplo, ser invariante a rotação dos eixos da molécula;
- Simples, deverá ser computacionalmente rápido de ser calculado;
- Descritivo, moléculas semelhantes deverão ter descritores parecidos.

A teoria da informação e do grafo são utilizadas na concepção de descritores moleculares, seguindo princípios da química quântica e orgânica, sobre os tipos variados de representações da estrutura da molécula (TODESCHINI; CONSONNI, 2009). Por derivar de diversas representações e com isso, serem calculados sobre informações diferentes do sistema, descritores são categorizados em constitucionais, topológicos e geométricos.

3.3.1 Descritores constitucionais

Descritores constitucionais descrevem a composição química da molécula, sem considerar as conexões e distribuição espacial de cada átomo (POLANSKI; GASTEIGER, 2017). Alguns desses descritores são:

- Massa molecular;
- Quantidade de um tipo de átomo;
- Quantidade de um tipo de ligação;
- Quantidade total de átomos;

- Quantidade total de ligações;
- Quantidade de anéis aromáticos;

3.3.2 Descritores topológicos

Descritores topológicos derivam do grafo molecular ou qualquer outra representação que contenha informações a respeito das conexões dos átomos, por exemplo, o SMILES (SHARMA et al., 1997). Neste sentido, a teoria do grafo pode ser aplicada para a extração de qualquer medida topológica, tal como ordem, tamanho, diâmetro, matriz de distância, índice Wiener, matriz laplaciana (MAURI et al., 2017).

O atributo de autocorrelação de Moreau-Broto (ATS_k), proposto por Broto et al. (1984), utiliza uma função de autocorrelação das propriedades dos átomos da molécula por meio da estrutura topológica vide Equação 3.1. Para tanto, é usado o delta de Kronecker $\delta(d_{ij}; k)$ para a autocorrelação do i -ésimo e j -ésimo átomo com distância k , no qual $\delta(d_{ij}; k) = 1$ quando $d_{ij} = k$ e 0, caso contrário.

$$ATS_k = \frac{1}{2} \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} w_i w_j \delta(d_{ij}; k), \quad (3.1)$$

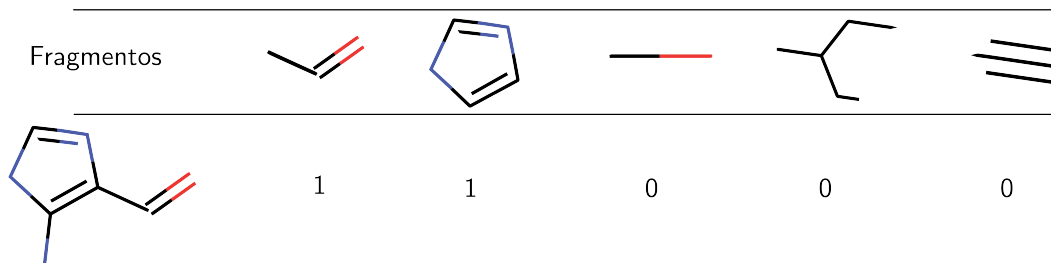
onde $|V|$ é a cardinalidade do conjunto de vértices e w é um peso, geralmente, uma propriedade atômica, como a massa atômica, eletronegatividade ou polarizabilidade. No caso onde a distância k é definida como 0, o descritor irá ser o somatório de cada peso ao quadrado.

Descritores *fingerprints* descrevem a composição da molécula, geralmente, usando uma lista binária, no qual cada elemento representa a presença ou não de uma estrutura contida na molécula. A definição da lista de estrutura ou fragmentos, deve ter tamanho fixo, e pode ser formulada manualmente ou usando métodos que visam explorar fragmentos significativos das moléculas de um conjunto de dados (XUE; BAJORATH, 2000). A Figura 3.6 ilustra o processo de codificação do *fingerprint*.

Vale ressaltar que a escolha de fragmentos representativos desempenha uma função importante no processo de aprendizado do algoritmo de AM. Por exemplo, não adiantará nada se a maioria das estruturas presente na lista definida não estiver contida nas moléculas do conjunto de dados, ou prevalecer apenas em um grupo de moléculas. Além disso, o tamanho da lista é outro fator que deverá ser considerado, não podendo ser muito pequeno para que não haja perda de informação e nem muito

grande, para não aumentar a complexidade do problema.

Figura 3.6 - Atributo *fingerprint* com fragmentos definidos manualmente.



Fonte: Produção do autor.

3.3.3 Descritores geométricos

Descritores geométricos baseiam-se em representações que contenham as coordenadas cartesianas dos átomos da molécula e suas conexões (MAURI et al., 2017). A matriz de Coulomb, proposto por Matthias Rupp, é um descritor que codifica as informações do hamiltoniano quântico ao calcular uma matriz C quadrada de dimensão d , considerando as coordenadas Z_i e número atômico R_i , por meio da Equação 3.2 (RUPP et al., 2012).

$$C_{ij} = \begin{cases} 0.5Z_i^{2.4} & \text{for } \forall i = j \\ \frac{Z_i Z_j}{|R_i - R_j|} & \text{for } \forall i \neq j \end{cases} \quad (3.2)$$

Por gerar uma matriz simétrica, pode-se converter a matriz em uma lista desconsiderando os elementos superiores ou inferiores em relação à diagonal principal vide Figura 3.7.

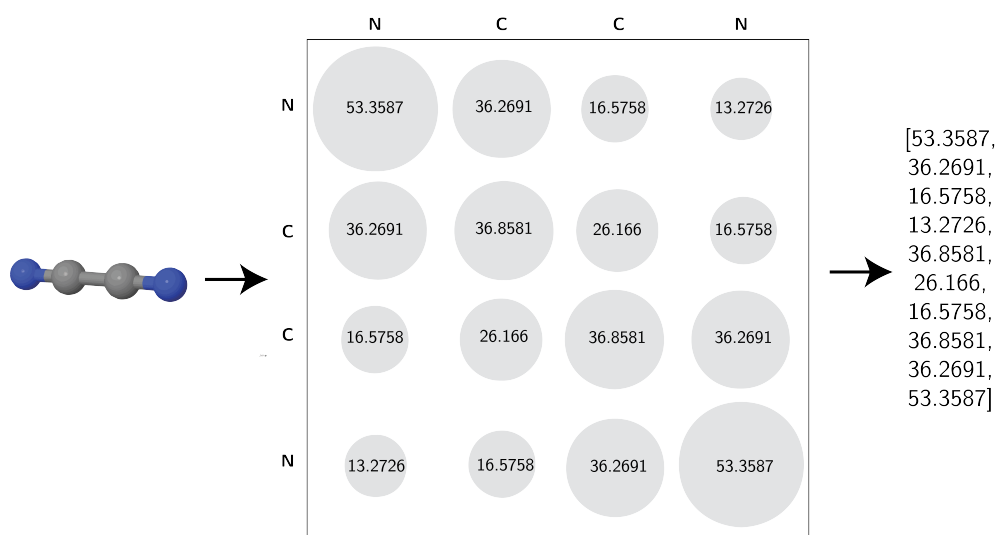
Contudo, a matriz de Coulomb não é um atributo único, ou seja, é possível obter várias saídas de uma mesma molécula, apenas trocando a ordem dos átomos na entrada do cálculo. Dessa forma, uma possível solução é a ordenação da matriz, sobre cada linha i de C , satisfazendo a seguinte condição $\|C_i\| \geq \|C_{i+1}\| \forall i$ (FABER et al., 2015; MONTAVON et al., 2012).

3.4 Estado da arte

A aplicação de AM em ciência dos materiais visa solucionar problemas que requerem a predição de uma composição, estrutura ou propriedade tendo como base de conhe-

cimento dados químicos e ou físicos da molécula. Na literatura, as aplicações de AM envolvem (i) a predição de propriedades moleculares, (RUPP et al., 2015; FABER et al., 2015; MONTAVON et al., 2013), (ii) o desenvolvimento de descritores moleculares (BUTLER et al., 2018; SCHÜTT et al., 2017b; SCHÜTT et al., 2017a; GILMER et al., 2017; GÓMEZ-BOMBARELLI et al., 2018), (iii) o *design* inverso (SANCHEZ-LENGELING; ASPURU-GUZYK, 2018), (iv) a síntese de compostos químicos (SEGLER; WALLER, 2017), (v) a triagem virtual (MELVILLE et al., 2009) e entre outros. Contudo, dado o contexto do presente trabalho, nesta seção será revisado apenas a predição de propriedades moleculares.

Figura 3.7 - Matriz de Coulomb.



Fonte: Adaptada de Hansen et al. (2013).

No âmbito da predição de propriedades moleculares, modelos de RNAs baseados em grafos apresentaram excelentes desempenhos por aprenderem a codificar o grafo da estrutura molecular, no qual átomos são representados pelos nós e as ligações como os arcos, em um vetor de característica significativo para a tarefa supervisionada. Enquanto nas redes convolucionais o filtro percorre sequencialmente uma estrutura em grade para extrair as características, na rede baseada em grafos essa extração é feita percorrendo cada nó do grafo e combinando a informação do nó com a da sua vizinhança.

Duvenaud et al. (2015) apresentou uma arquitetura de rede convolucional em grafo com o objetivo de aprender *fingerprints*. Neste modelo, as camadas escondidas combinam a informação de um átomo com a sua vizinhança. Logo, dado uma rede com

n camadas escondidas, a informação a respeito de um átomo na camada n é a sua informação combinada com os átomos vizinhos de distância n . Nos experimentos realizados pelo trabalho, tal modelo conseguiu aumentar a acurácia na predição da solubilidade e eficiência fotovoltaica em relação ao método *fingerprints* circular.

Schütt et al. (2017b) propôs um modelo chamado *Deep Tensor Neural Networks* (DTNN) que utiliza hamiltoniano de muitos corpos para aprender uma representação adequada da molécula. A DTNN é considerada como uma versão adaptativa da matriz de Coulomb e recebe como entrada o número atômico Z e a matriz de distância D entre cada par de átomos. Inicialmente, Z é mapeado para um conjunto de coeficientes c_i^0 para a representação de cada átomo i da molécula. Durante T iterações, a DTNN realiza a atualização do coeficiente c_i^t por meio de uma combinação não linear entre o estado anterior c_i^{t-1} do átomo i e os coeficientes dos átomos vizinhos ponderados pela matriz de distância e o peso da rede. Após as T iterações, a propriedade de interesse é predita usando uma camada totalmente conectada para a ponderação das contribuições atômicas c_i^t sobre uma função não linear. O trabalho conseguiu reportar um EAM de $0.84 \text{ kcal mol}^{-1}$ na predição da energia livre de Gibbs de formação para a temperatura de 298.15 K (G) do QM9.

Posteriormente, Gilmer et al. (2017) apresentou um modelo de rede baseado em grafo levando em consideração as informações a respeito dos átomos e das ligações. Neste trabalho, diferentes modelos baseados em grafo foram revisados ((LI et al., 2015), (BATTAGLIA et al., 2016), (KEARNES et al., 2016)) e tiveram seus resultados comparados para a predição de 13 propriedades moleculares do QM9, entre elas as propriedades termodinâmicas, momento dipolar (μ), polarizabilidade isotrópica (α) e outras incluídas no conjunto de dados chamado QM9. Em todos os casos, a rede proposta superou a acurácia dos outros modelos apresentados.

Faber et al. (2017) realizou um *benchmark* para o mesmo conjunto QM9. No trabalho é reportado a predição das propriedades do QM9 usando diferentes descritores moleculares e algoritmos de AM. Para as propriedades de atomização da energia interna para a temperatura de 0 K (U_0), energia de ponto zero (zpve) e capacidade térmica para a temperatura de 298.15 K (C_v), o *kernel ridge regression* foi o algoritmo que apresentou os melhores resultados, atingindo um EAM de 0.0251 eV , 0.00191 eV e $0.0441 \text{ cal mol}^{-1} \text{ K}^{-1}$, respectivamente. A rede convolucional em grafo, proposta por Kearnes et al. (2016), teve os melhores desempenhos na predição da energia do HOMO (ϵ_{HOMO}), energia do LUMO (ϵ_{LUMO}), diferença de energias do HOMO e LUMO ($\Delta\epsilon$), e μ com EAM equivalentes a 0.0549 eV , 0.0620 eV , 0.0869

eV e 0.101 Debye, respectivamente. Um segundo modelo de rede baseado em grafo apresentou o menor EAM igual a $0.161 a_0^3$ para a propriedade α .

Seguindo a mesma linha do trabalho anterior e inspirado em projetos como ImageNet e WordNet, Wu et al. (2018) propõe um dos maiores *benchmarks* sobre dados moleculares, denominado como MoleculeNet. O MoleculeNet oferece em uma única ferramenta mais de 700 mil compostos químicos distribuídos em conjuntos de dados das áreas da mecânica quântica, físico-química, biofísica e fisiologia, assim como também métodos para extração de descritores, implementação de algoritmo de AM e avaliação de desempenho de modelos de AM. Neste trabalho, a rede DTNN apresentou os melhores resultados para 7 propriedades do QM9.

Schütt et al. (2018) propôs um modelo chamado de SchNet, considerado como variante da rede DTNN, que atingiu o estado da arte para a predição de diversas propriedades moleculares. Para o conjunto QM9, por exemplo, a SchNet teve um EAM próximo de 0.014 eV em quase todas as propriedades termodinâmicas. ϵ_{HOMO} , ϵ_{LUMO} e Δ_ϵ tiveram EAM iguais a 0.041, 0.034 e 0.063 eV, respectivamente. Em 2019, CHEN et al. publicou um artigo apresentando um novo modelo de rede baseado em grafo, intitulada como MegNet, que ultrapassou o estado da arte para na predição de 11 propriedades do QM9. As propriedades termodinâmicas tiveram um EAM próximo a 0.010 eV, exceto U_0 que obteve um erro de 0.009 eV. A propriedade z_{pve} teve um erro de 1.40 meV, μ obteve um EAM igual 0.040 Debye, α gerou um erro de $0.083 a_0^3$, C_v teve um erro de $0.030 \text{ cal mol}^{-1} \text{ K}^{-1}$ e a maior frequência vibracional obteve um erro de 1.10 cm^{-1} . ϵ_{HOMO} , ϵ_{LUMO} e Δ_ϵ atingiram um EAM de 0.038, 0.031 e 0.061 eV, respectivamente.

Outra abordagem realizou a predição de propriedades usando uma RNA convolucional sobre a imagem do desenho da molécula (GOH et al., 2017b) para um outro conjunto de dados. Neste trabalho, a rede convolucional teve um desempenho melhor do que uma rede Perceptron de múltiplas camadas recebendo como entrada *fingerprints* circular na maioria das propriedades preditas. Outros trabalhos utilizaram redes neurais recorrentes e *autoencoders* para a predição de propriedades usando como entrada o SMILES da molécula (SEGLER et al., 2018; GÓMEZ-BOMBARELLI et al., 2018).

Em suma, na predição de propriedades físico-químicas, os resultados vêm mostrando que AM é capaz de produzir um cálculo rápido com níveis de acurácia química aceitáveis, enquanto em métodos tradicionais da ciência dos materiais a melhora da acurácia é influenciada pelo tempo necessário para a precisão dos cálculos (HANSEN

et al., 2013; YE et al., 2018). Segundo Faber et al. (2017), algoritmos de AM treinados sobre dados calculados por meio de DFT conseguem produzir um erro menor do que duas vezes o erro gerado pela DFT e um erro de generalização menor que o erro da DFT, caso os cálculos das propriedades providas sejam feitas usando um alto nível de teoria ou usando dados experimentais.

Já no campo da engenharia de descritores, é possível investigar algoritmos de AM para aprender representações significativas para uma determinada tarefa supervisionada, como fora apresentado anteriormente, ou propor descritores baseados em conceitos da química e ou física. Na literatura, a matriz de Coulomb é um dos descritores mais populares por ter apresentado excelentes resultados em tarefas de predições e utilizar a mesma informação do hamiltoniano quântico.

4 MATERIAIS E MÉTODOS

Este trabalho realizou a comparação do desempenho de modelos Perceptron de múltiplas camadas na tarefa de predição de propriedades moleculares para o estado fundamental das moléculas. Para tanto, foram utilizados descritores baseados na representação SMILES, já que SMILES é de fácil obtenção e de ampla utilização nos mais diversos *softwares* da quimioinformática e em bases de dados.

Para a construção dos modelos preditivos utilizou-se uma base de dados pública, chamada QM9, que oferece aproximadamente 130 mil moléculas orgânicas compostas de átomos de C, H, O, N e F, com quantidade de átomos variando entre 3 até 29. Todas as moléculas do QM9 contêm suas estruturas otimizadas e 15 propriedades moleculares, vide Tabela 4.1, obtidas usando o DFT B3LYP/6-31G(2df, p).

Tabela 4.1 - Propriedades e suas respectivas unidades do conjunto QM9.

Propriedades	Unidade
Constante rotacional A	GHz
Constante rotacional B	GHz
Constante rotacional C	GHz
Momento dipolar (μ)	Debye
Polarizabilidade isotrópica (α)	a_0^3
Energia do HOMO (ϵ_{HOMO})	Ha
Energia do LUMO (ϵ_{LUMO})	Ha
Diferença de energias do HOMO e LUMO (Δ_ϵ)	Ha
Extensão espacial eletrônica ($\langle R^2 \rangle$)	a_0^2
Energia de ponto zero (zpve)	Ha
Energia interna para a temperatura de 0 K (U_0)	Ha
Energia interna para a temperatura de 298.15 K (U)	Ha
Entalpia para a temperatura de 298.15 K (H)	Ha
Energia livre de Gibbs de formação para a temperatura de 298.15 K (G)	Ha
Capacidade térmica para a temperatura de 298.15 K (C_v)	cal mol ⁻¹ K ⁻¹

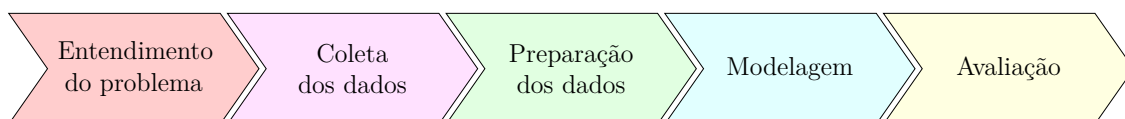
Fonte: Produção do autor.

No processo dos experimentos, também foi utilizado 3 subconjuntos do QM9, cada um composto por grupo de moléculas de tamanho específico, estes são moléculas com 17, 18 e 19 átomos. Tal restrição imposta na definição dos subconjuntos tem como objetivo verificar se a homogeneidade dos dados de cada subconjunto resulta

em uma acurácia melhor do modelo.

De modo geral, o fluxo de atividades usado neste trabalho segue uma abordagem similar aos processos convencionais empregados em atividades de mineração de dados, como por exemplo o CRISP-DM (WIRTH; HIPPEL, 2000). Figura 4.1 apresenta a sequência de cada um dos processos feitos neste trabalho e nas seções seguintes será detalhado cada uma dessas respectivas etapas.

Figura 4.1 - Fluxo de processos para extração de conhecimento com AM.



Fonte: Produção do autor.

4.1 Entendimento do problema

Nesta seção serão discutidas as etapas necessárias para a compreensão do problema, assim como também os métodos de AM e a base de dados que será empregada para solucionar o problema. Figura 4.2 ilustra cada passo pertencente a esta etapa.

4.1.1 Definição do problema

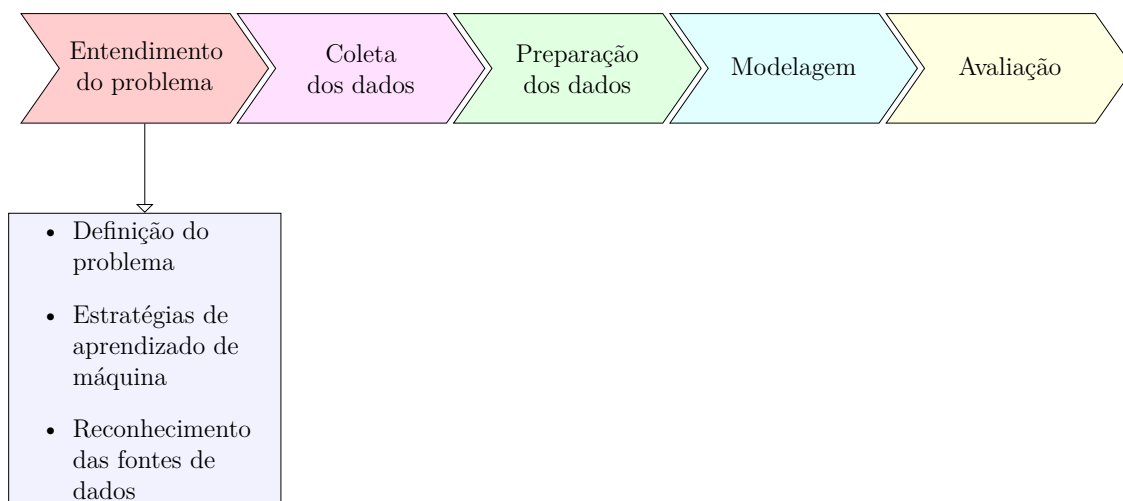
O estudo das propriedades físico-químicas permite investigar como as moléculas reagem e se comportam em condições específicas. Tipicamente, o processo do cálculo de uma propriedade demanda bastante recurso computacional, pois requer, primeiramente, a otimização da molécula para o estado de menor energia e em seguida, o cálculo da propriedade desejada por meio de simulações computacionais.

Na literatura, modelos de AM estão sendo aplicados para realizar a predição de propriedades físico-químicas como uma abordagem computacionalmente menos custosa (BUTLER et al., 2018). Neste caso, geralmente, é utilizado como entrada do algoritmo descritores derivados da posição espacial em coordenadas XYZ obtidas por meio da molécula otimizada, que por sua vez é um processo computacionalmente custoso.

Por outro lado, o SMILES é uma representação molecular que contém informações a respeito das ligações e átomos existentes na molécula e é uma representação fácil de ser obtida, pois, não depende do processo de otimização. Neste sentido, o obje-

tivo deste trabalho é elaborar um modelo preditivo usando descritores baseados na representação SMILES como entrada, capazes de prever propriedades moleculares com acurácias semelhantes ou melhores que descritores baseados na posição espacial dos átomos.

Figura 4.2 - Módulo sobre entendimento do problema.



Fonte: Produção do autor.

4.1.2 Estratégias de aprendizado de máquina

Na solução do problema em questão, foram utilizadas RNAs do tipo Perceptron de múltiplas camadas no contexto supervisionado, no qual se definiu como as entradas x , os descritores baseados na representação moleculares (SMILES e arquivos XYZ) e a saída desejada y , referente a uma propriedade físico-química de interesse. Além disso, algoritmos de AM como RF e LASSO foram usados para a realização da SA, no qual foram aplicados para cada uma das propriedades de interesse.

4.1.3 Reconhecimento das fontes de dados

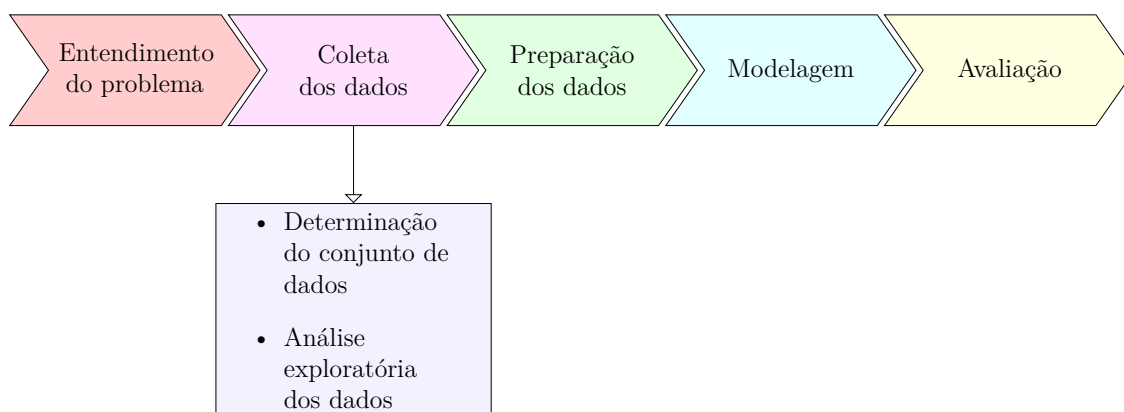
A escolha da fonte de dados considerou a utilização de um conjunto de dados que fosse público e já consolidado na área de interesse, contemplando uma grande quantidade de amostra e diversidade nos dados, isto é, incluindo moléculas de diferentes composições químicas e quantidade de átomos. Além disso, considerou-se também a variedade de representações e propriedades moleculares reportadas no conjunto, bem como a confiabilidade dos cálculos e otimizações realizadas sobre a molécula

para a obtenção das propriedades.

4.2 Coleta dos dados

Nesta seção será apresentada a base de dados utilizada neste trabalho e as análises feitas sobre este conjunto. Figura 4.3 detalha os processos envolvidos nesta etapa.

Figura 4.3 - Módulo sobre entendimento do problema.



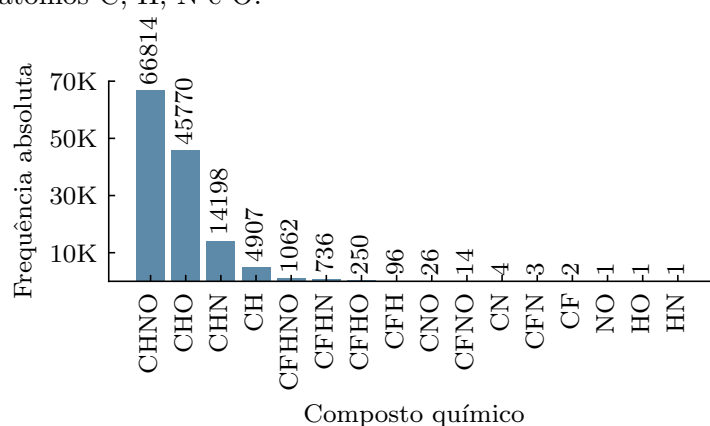
Fonte: Produção do autor.

4.2.1 Determinação do conjunto de dados

Neste trabalho, foi usado uma base pública chamada como QM9 que contém 133.885 moléculas orgânicas compostas por átomos de C, H, O, N e F e tamanho variando entre 3 a 29 átomos. Figura 4.4 e 4.5 apresentam o histograma da quantidade de moléculas por compostos químicos e quantidade de átomos, respectivamente.

O conjunto de dados QM9 é um subconjunto da base GDB-17, que por sua vez contém 166 bilhões de moléculas do espaço químico. Contudo, o QM9 disponibiliza outros dados sobre as moléculas além do SMILES oferecido pelo GDB-17, estes são: (i) o cálculo de 15 propriedades moleculares como apresentado na Tabela 4.1; (ii) as coordenadas XYZ obtidas através da conversão do SMILES original usando o programa Corina que, em seguida, foram otimizadas para o cálculo das propriedades; (iii) o identificador químico internacional (do inglês *International Chemical Identifier*, InChI) gerado pelo XYZ obtido pelo Corina e também pelo XYZ otimizado; (iv) e o SMILES derivado do XYZ otimizado.

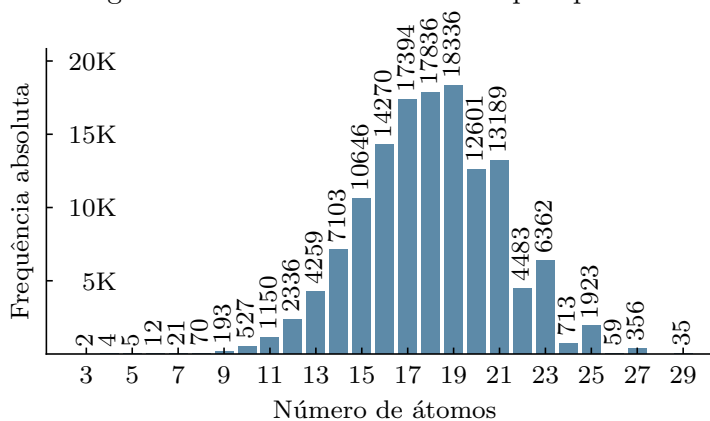
Figura 4.4 - Histograma do número de moléculas em função dos tipo de átomos presentes. Note que existe cerca de 67 mil moléculas contento os quatro diferentes tipos de átomos C, H, N e O.



Fonte: Produção do autor.

Um dos objetivos do QM9 é servir para o desenvolvimento de novos métodos envolvendo AM e para o *benchmarking* de modelos e descritores moleculares já propostos. Para tanto, todas as moléculas do QM9 foram armazenadas em um formato XYZ modificado, para que as informações referentes às propriedades, coordenadas cartesianas, SMILES e demais dados fossem organizados em um único arquivo.

Figura 4.5 - Histograma do número de moléculas por quantidade de átomos.



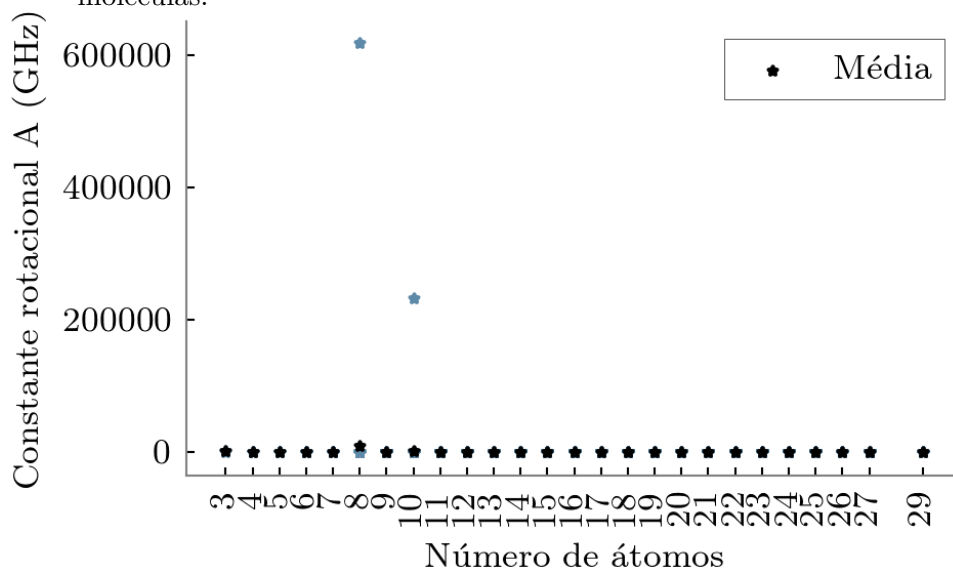
Fonte: Produção do autor.

4.2.2 Análise exploratória dos dados

Na análise exploratória dos dados foram realizadas diversas representações visuais para o entendimento básico dos dados, investigação de tendências e detecção de moléculas atípicas. Por exemplo, na Figura 4.5 fora apresentado o número de moléculas pela quantidade de átomos, no histograma é possível notar que existem poucas moléculas com quantidade de átomos exatamente igual ou menor que 8. Isso permite entender que dada as características do QM9, é mais complexo criar moléculas com uma quantidade de átomos tão restrita.

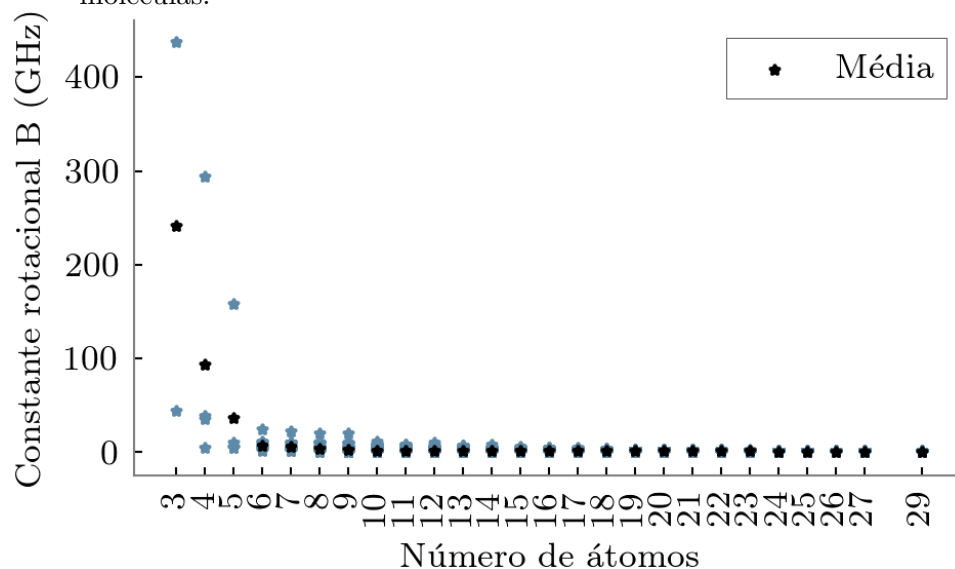
Para cada uma das propriedades relatadas na Tabela 4.1, fez-se um gráfico mostrando a dispersão dos valores pelo tamanho da molécula, ou quantidade de átomos, vide Figuras 4.6 até 4.20. Note que as propriedades termodinâmicas (U_0 , U , H e G) e eletrônicas (ϵ_{HOMO} , ϵ_{LUMO} e $\Delta\epsilon$) tiveram a unidade convertida de Hartree (Ha) para eletrovolts (eV), já que eV é usualmente usado na literatura para esses casos.

Figura 4.6 - Dispersão dos valores por quantidade de átomos para a propriedade constante rotacional A. Pontos em preto representa a média em para cada tamanho de moléculas.



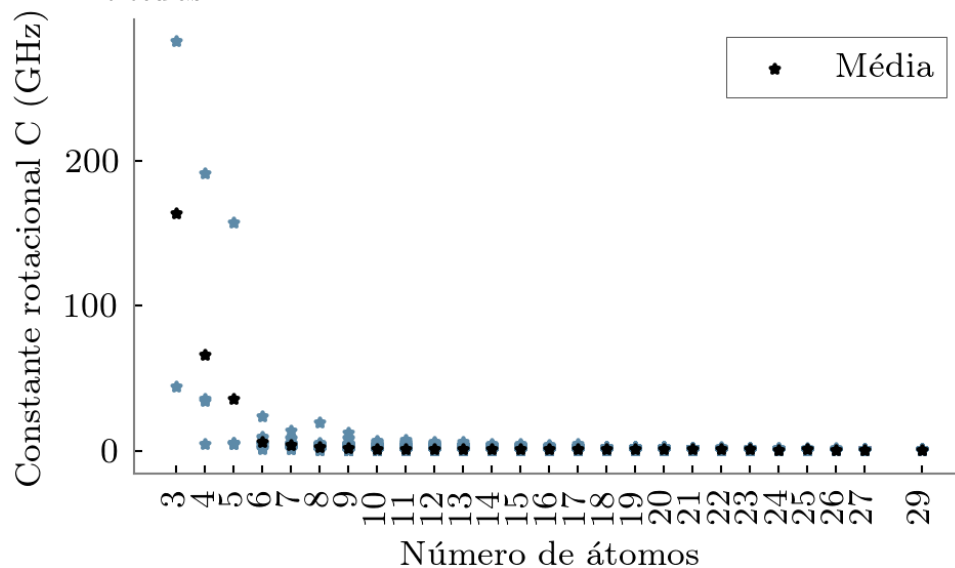
Fonte: Produção do autor.

Figura 4.7 - Dispersão dos valores por quantidade de átomos para a propriedade constante rotacional B. Pontos em preto representa a média em para cada tamanho de moléculas.



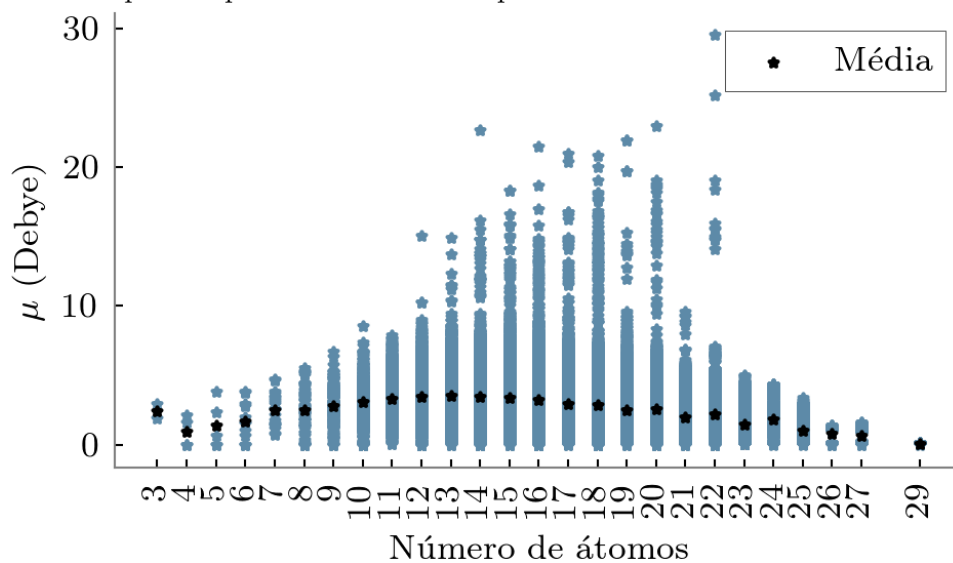
Fonte: Produção do autor.

Figura 4.8 - Dispersão dos valores por quantidade de átomos para a propriedade constante rotacional C. Pontos em preto representa a média em para cada tamanho de moléculas.



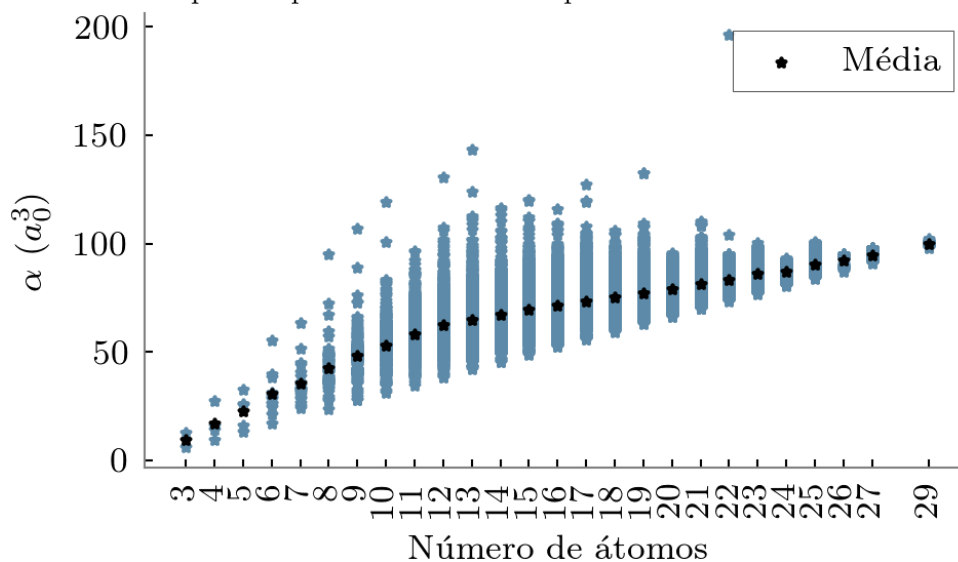
Fonte: Produção do autor.

Figura 4.9 - Dispersão dos valores por quantidade de átomos para a propriedade μ . Pontos em preto representa a média em para cada tamanho de moléculas.



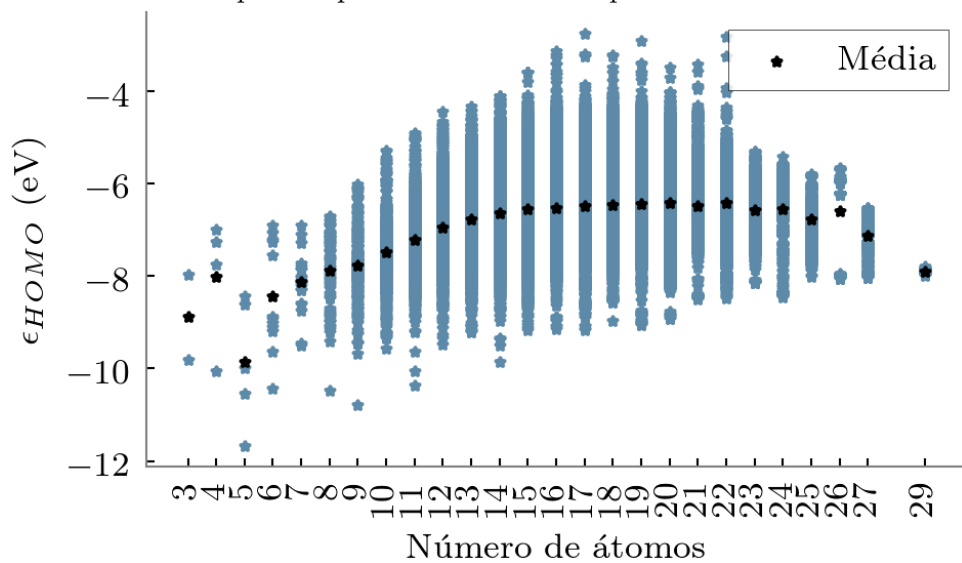
Fonte: Produção do autor.

Figura 4.10 - Dispersão dos valores por quantidade de átomos para a propriedade α . Pontos em preto representa a média em para cada tamanho de moléculas.



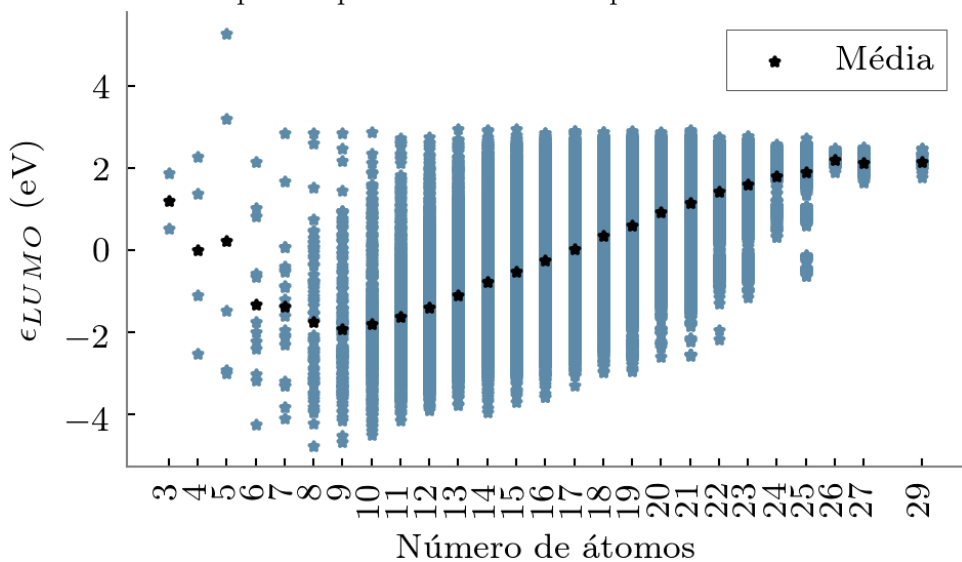
Fonte: Produção do autor.

Figura 4.11 - Dispersão dos valores por quantidade de átomos para a propriedade ϵ_{HOMO} . Pontos em preto representa a média em para cada tamanho de moléculas.



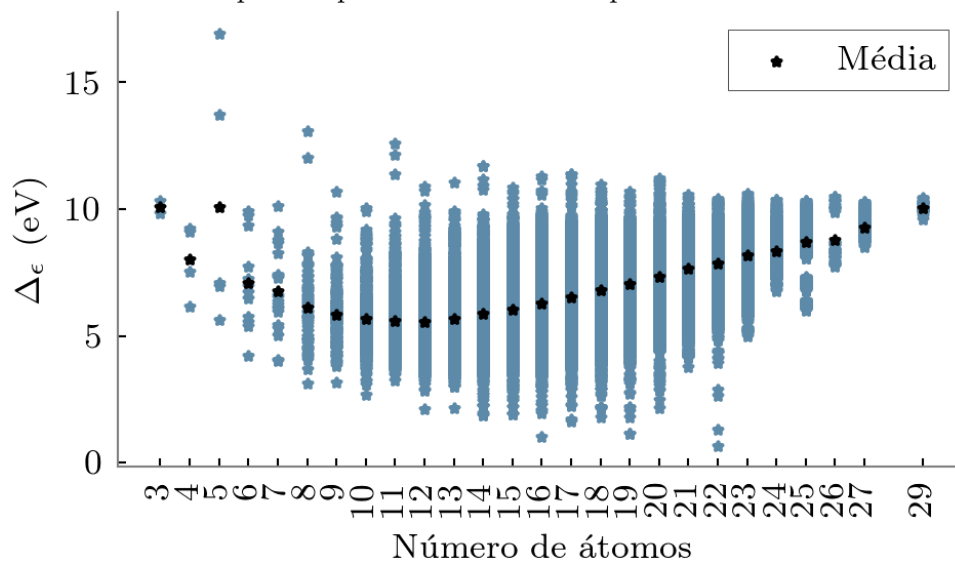
Fonte: Produção do autor.

Figura 4.12 - Dispersão dos valores por quantidade de átomos para a propriedade ϵ_{LUMO} . Pontos em preto representa a média em para cada tamanho de moléculas.



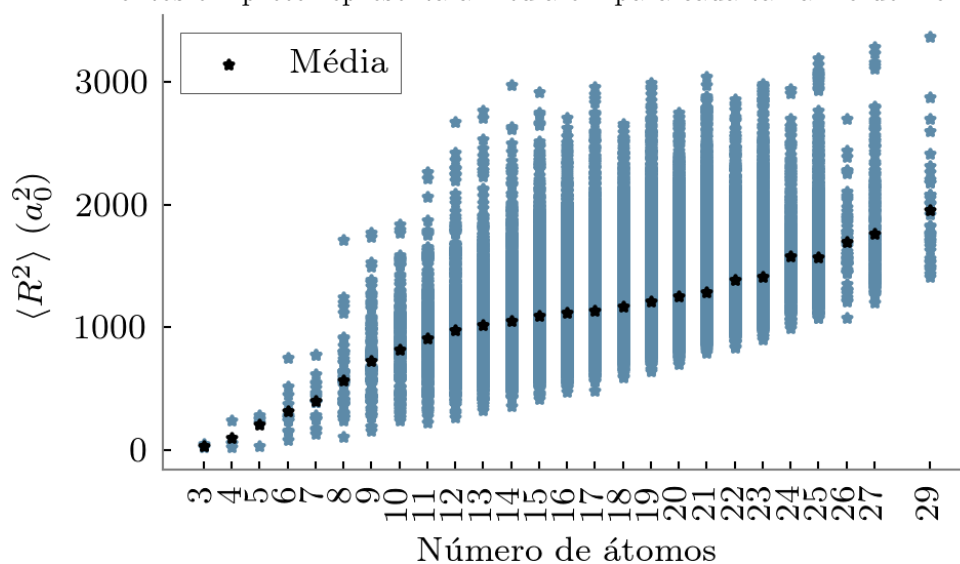
Fonte: Produção do autor.

Figura 4.13 - Dispersão dos valores por quantidade de átomos para a propriedade Δ_ϵ . Pontos em preto representa a média em para cada tamanho de moléculas.



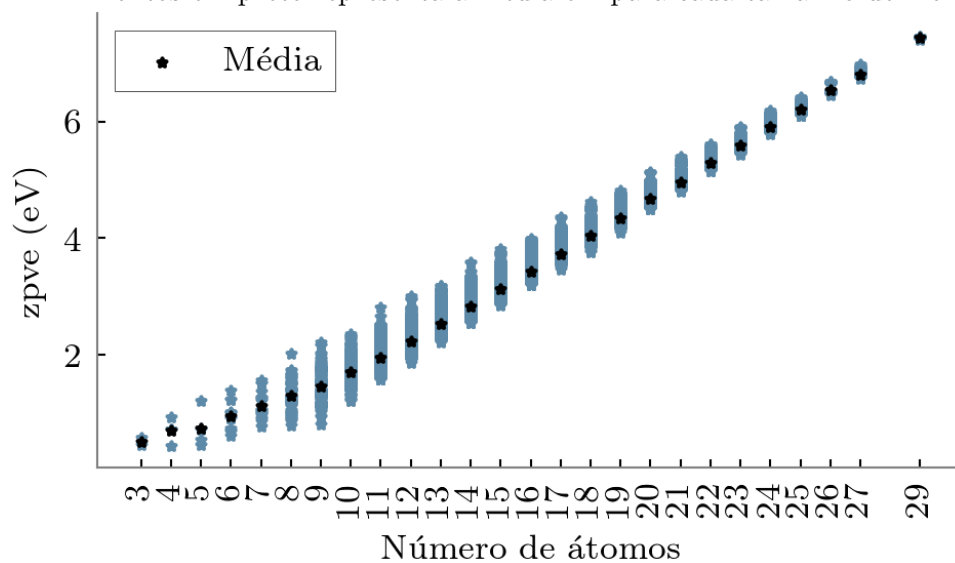
Fonte: Produção do autor.

Figura 4.14 - Dispersão dos valores por quantidade de átomos para a propriedade $\langle R^2 \rangle$. Pontos em preto representa a média em para cada tamanho de moléculas.



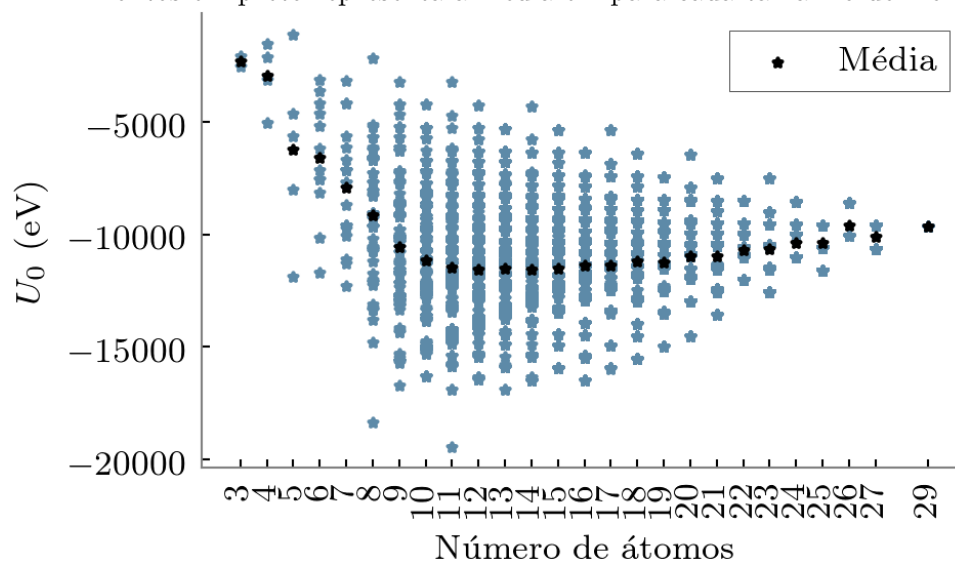
Fonte: Produção do autor.

Figura 4.15 - Dispersão dos valores por quantidade de átomos para a propriedade zpve. Pontos em preto representa a média em para cada tamanho de moléculas.



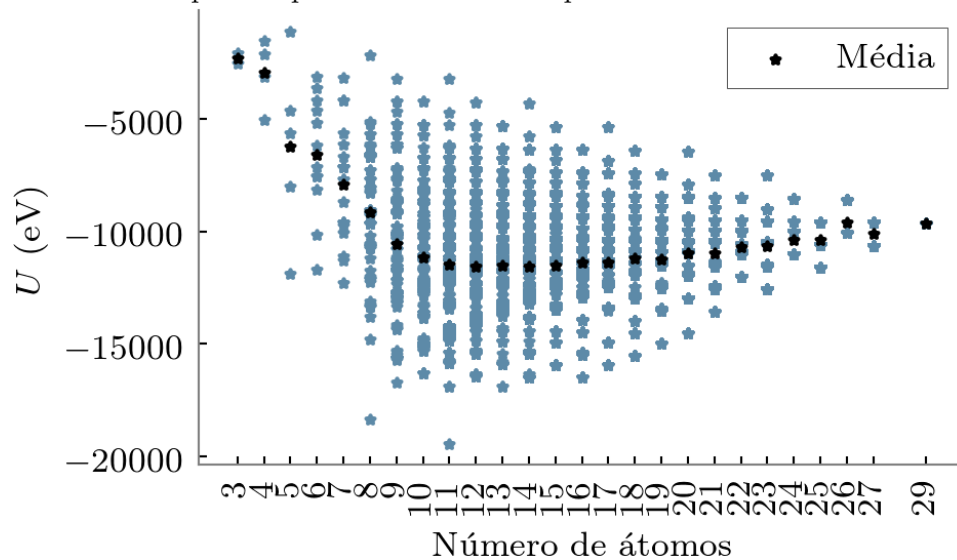
Fonte: Produção do autor.

Figura 4.16 - Dispersão dos valores por quantidade de átomos para a propriedade U_0 . Pontos em preto representa a média em para cada tamanho de moléculas.



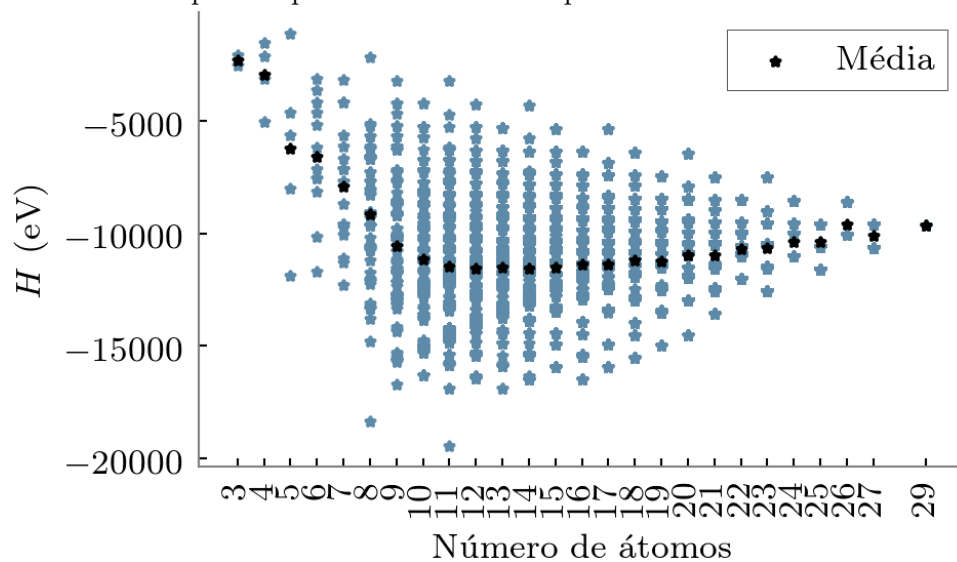
Fonte: Produção do autor.

Figura 4.17 - Dispersão dos valores por quantidade de átomos para a propriedade U . Pontos em preto representa a média em para cada tamanho de moléculas.



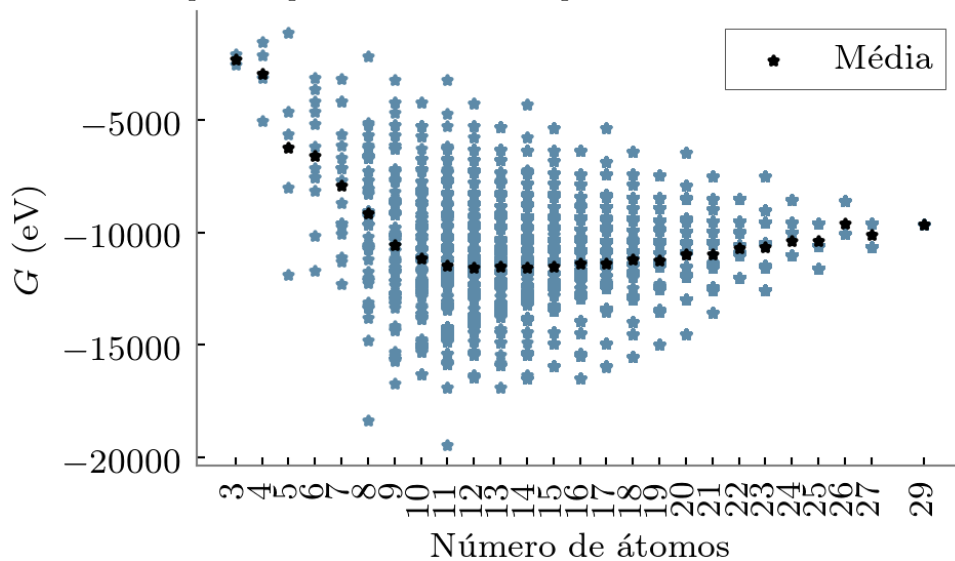
Fonte: Produção do autor.

Figura 4.18 - Dispersão dos valores por quantidade de átomos para a propriedade H . Pontos em preto representa a média em para cada tamanho de moléculas.



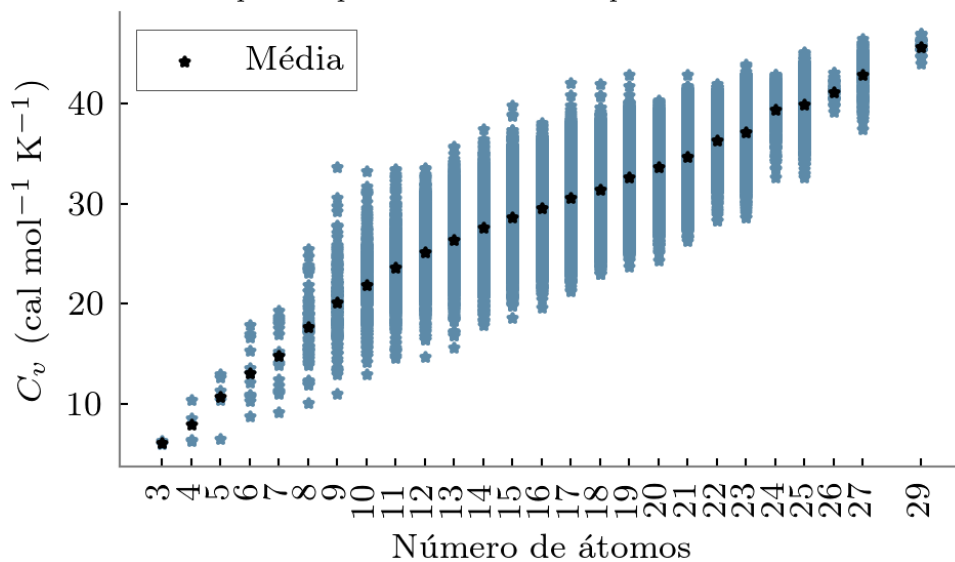
Fonte: Produção do autor.

Figura 4.19 - Dispersão dos valores por quantidade de átomos para a propriedade G . Pontos em preto representa a média em para cada tamanho de moléculas.



Fonte: Produção do autor.

Figura 4.20 - Dispersão dos valores por quantidade de átomos para a propriedade C_v . Pontos em preto representa a média em para cada tamanho de moléculas.



Fonte: Produção do autor.

Nos gráficos 4.6 até 4.20 é possível notar que a quantidade de moléculas com número de átomos até 8 refletiu no desvio da tendência das distribuições esperadas na mai-

oria dos casos. Esse mesmo desvio da tendência acontece também para moléculas com 26 e 29 átomos, cuja quantidade de amostras são de 59 e 35, respectivamente. Acredita-se que com mais amostras destes casos seja possível obter um ajuste mais fino da distribuição. Além disso, na Figura 4.6 é possível visualizar duas moléculas com o valor bem discrepante (uma molécula apresentando valor superior a 200 mil e outra acima de 600 mil). A Figura 4.21 mostra o grafo dessas duas moléculas, no qual percebe-se que ambas contêm estruturas bem semelhantes.

Figura 4.21 - Moléculas com valores discrepantes para constante rotacional A.



ID 484	ID 14562
Constante rotacional A 619 867.683 14	Constante rotacional A 232 663.788 79

Fonte: Produção do autor.

Foi realizada uma estatística descritiva básica, vide Tabela 4.2, para sintetizar os dados através de métricas como média, desvio padrão a fim de obter-se uma visão geral da tendência central e variação dos valores de cada propriedade.

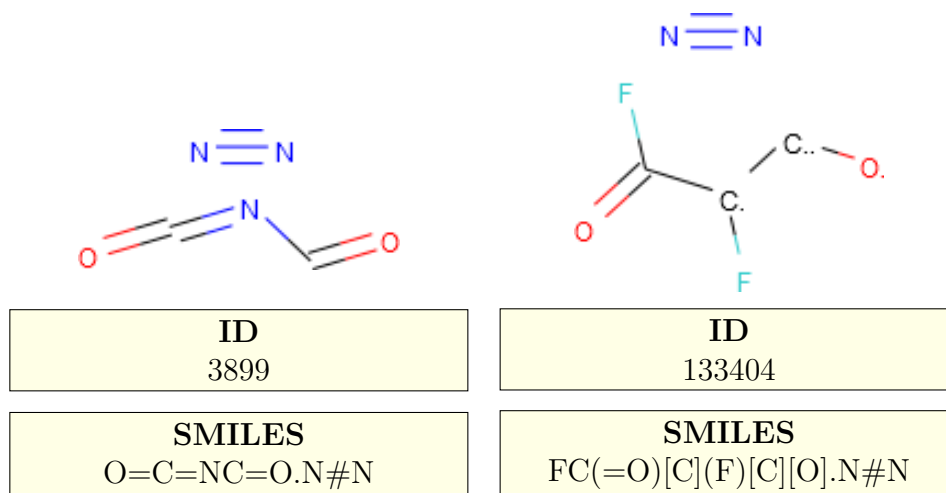
Além da análise sobre as propriedades, foi investigado a representação SMILES. O SMILES utilizado neste trabalho é o obtido pelas coordenadas cartesianas otimizadas, pois deriva do mesmo XYZ usado no cálculo das propriedades e é conhecido como SMILES relaxado. Inicialmente, nas análises foram encontradas 229 moléculas com múltiplas estruturas, em outras palavras, moléculas com estruturas disjuntas. No SMILES tal condição é representada pelo carácter ponto “.”, indicando que os símbolos da cadeia do lado esquerdo do “.” fazem parte de uma estrutura e os símbolos do lado direito de outra estrutura, como apresentado pela Figura 4.22. Acredita-se que tal condição tenha sido provocada durante o processo de otimização da molécula, que inclui utilização de métodos de campo de forças e do DFT/B3LYP.

Tabela 4.2 - Estatística descritiva básica sobre o QM9, no qual σ representa o desvio padrão e Min e Max denomina os valores mínimo e máximo, respectivamente.

Propriedade	Média	σ	Min	Max
Constante rotacional A	3.45	3.23	0	799.59
Constante rotacional B	1.41	1.60	0.34	437.90
Constante rotacional C	1.13	1.11	0.33	282.95
μ	2.65	1.41	0	19.69
α	75.28	8.16	6.31	143.53
ϵ_{HOMO}	-6.54	0.59	-11.66	-3.64
ϵ_{LUMO}	0.33	1.27	-4.76	5.27
Δ_ϵ	6.87	1.27	2.14	16.93
$\langle R^2 \rangle$	1189.48	280.71	19.00	3374.75
zpve	4.06	0.90	0.43	7.45
U_0	-76.16	10.31	-113.11	-9.24
U	-76.63	10.40	-113.89	-9.28
H	-77.07	10.47	-114.61	-9.33
G	-70.88	9.48	-104.81	-8.73
C_v	31.63	4.07	6.00	46.97

Fonte: Produção do autor.

Figura 4.22 - Exemplos de moléculas com múltiplas estruturas na representação SMILES.



Fonte: Produção do autor.

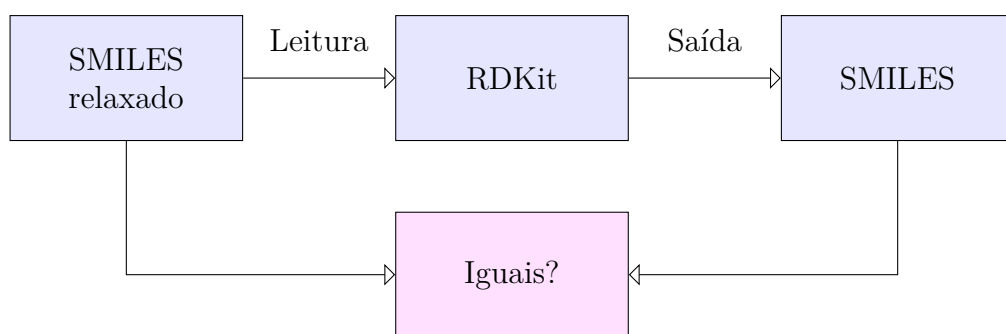
As análises restantes do SMILES relaxado foram realizadas por meio do pacote RD-

Kit¹. RDKit é uma ferramenta amplamente utilizada na área da quimioinformática, que contém métodos para leitura e tratamento de moléculas, e extração de descritores moleculares para atividades de AM. Neste trabalho, o RDKit desempenhou o papel da interpretação do SMILES relaxado para extração dos descritores, como será apresentado a seguir.

A princípio, o RDKit apresentou erros no processo de interpretação do SMILES relaxado de 746 moléculas, sendo o problema mais comum associado ao número de valência. Estes erros estão relacionados ao fato do RDKit empregar uma abordagem para o tratamento do SMILES diferente daquela oferecida pelo Open Babel, programa utilizado na base QM9 para a obtenção do SMILES relaxado.

Devido a diversas formas de interpretações, é possível que o RDKit faça o reconhecimento de uma molécula, sem manifestar nenhum erro, porém, a molécula gerada pelo programa acabar sendo totalmente diferente da original. Neste caso, o processo de consistência da interpretação da molécula consistiu em comparar os SMILES gerados pelo RDKit com aqueles fornecidos pelo QM9, como descrito no fluxo apresentado pela Figura 4.23.

Figura 4.23 - Verificação da consistência dos SMILES.



Fonte: Produção do autor.

Após as comparações, encontrou-se 24.446 moléculas que tiveram o SMILES modificado. Deste total, não foi notada nenhuma modificação que alterou de fato a estrutura da molécula, mas sim apenas apresentou uma segunda forma de representar a mesma molécula. Por exemplo, no SMILES, é possível expressar anéis aromáticos usando dígitos para indicar abertura e fechamento do anel, e omitindo ou não as

¹<https://www.rdkit.org/>

ligações presentes entre os elementos do anel aromático. No caso da omissão das ligações, as letras dos elementos pertencentes ao anel são transcritas em minúsculos. Para este último caso, em específico, em moléculas com a presença de heteroátomos (como o N, O), o RDKit acabou modificando o SMILES relaxado. No total, foram 160 casos encontrados, porém, em nenhum deles a estrutura molecular foi afetada, apenas a cadeia de caracteres do SMILES original. As demais modificações encontradas estão listadas abaixo.

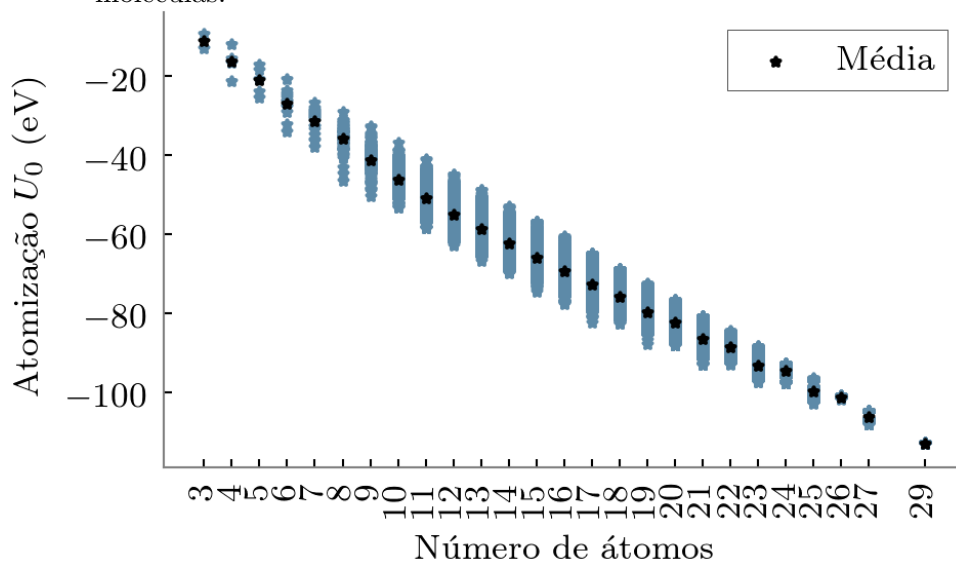
- 1820 moléculas sofreram mudanças nas especificações estereoquímicas do carbono quaternário ao substituir os caracteres “@@” pelo “@” ou vice-versa.
- 8574 moléculas tiveram os dígitos de abertura e fechamento do anel aromático trocados.
- 709 moléculas tiveram átomos de H inseridos no SMILES. Neste caso, a inserção de átomos de H não é visto como problema, já que é uma forma do RDKit tornar explícito os H omitidos.
- 8644 moléculas tiveram o SMILES modificado, porém, a espécie química não foi alterada.

4.2.2.1 Seleção das propriedades para predição

Neste trabalho, foi utilizado nove das quinze propriedades presentes no QM9, estas são: todas as propriedades termodinâmicas, C_v , ϵ_{HOMO} , $\Delta\epsilon$, μ e α . A escolha das propriedades deu-se pela sua relevância para a compreensão de vários aspectos físico-químicos, sendo que o ϵ_{LUMO} foi desconsiderado devido a sua dependência entre o ϵ_{HOMO} e o $\Delta\epsilon$; U_0 contém informação a respeito do zpve, portanto, zpve não foi incluso; $\langle R^2 \rangle$ e as constantes rotacionais não foram inclusas, devido à falta de informação a respeito do cálculo da propriedade na literatura e a forte relação com a morfologia da molécula, respectivamente.

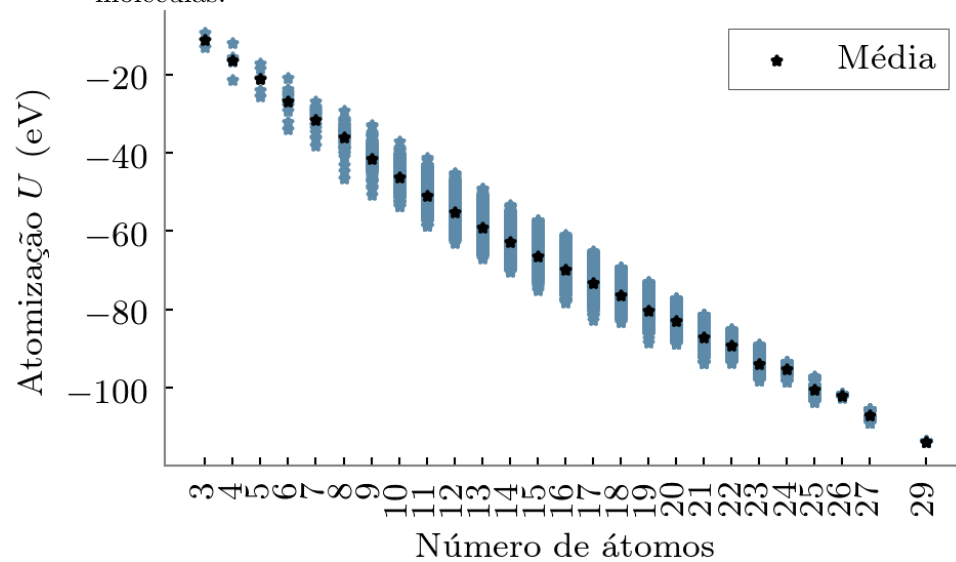
Além disso, para as propriedades termodinâmicas, foi usada a sua energia de atomização reportado em [Faber et al. \(2017\)](#). A energia de atomização é frequentemente usada em tarefas de predição na literatura e apresenta uma escala de valor menor. As Figuras 4.24 até 4.27 apresentam os gráficos de dispersão dos valores pelo tamanho da molécula para cada uma das energias de atomização.

Figura 4.24 - Dispersão dos valores por quantidade de átomos para a propriedade atomização U_0 . Pontos em preto representa a média em para cada tamanho de moléculas.



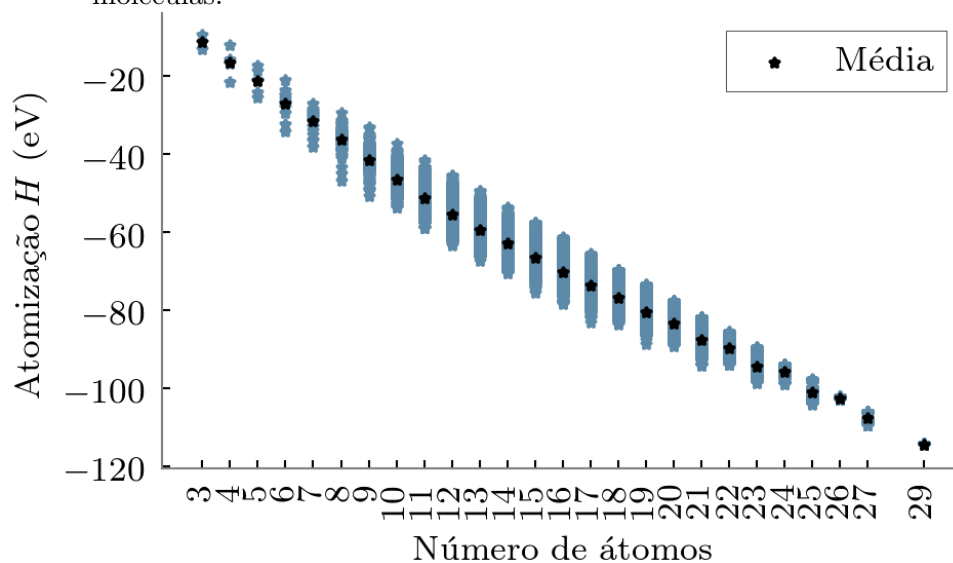
Fonte: Produção do autor.

Figura 4.25 - Dispersão dos valores por quantidade de átomos para a propriedade atomização U . Pontos em preto representa a média em para cada tamanho de moléculas.



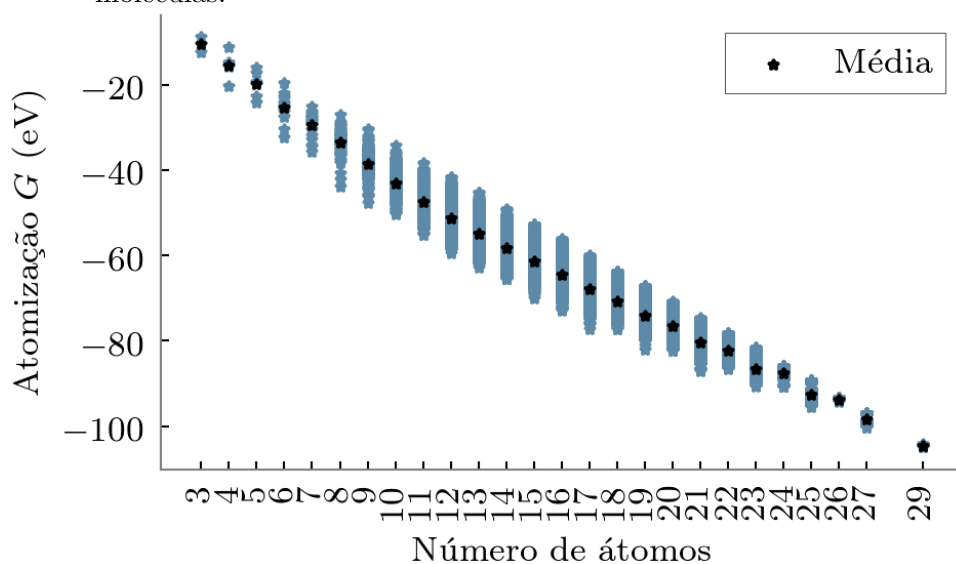
Fonte: Produção do autor.

Figura 4.26 - Dispersão dos valores por quantidade de átomos para a propriedade atomização H . Pontos em preto representa a média em para cada tamanho de moléculas.



Fonte: Produção do autor.

Figura 4.27 - Dispersão dos valores por quantidade de átomos para a propriedade atomização G . Pontos em preto representa a média em para cada tamanho de moléculas.

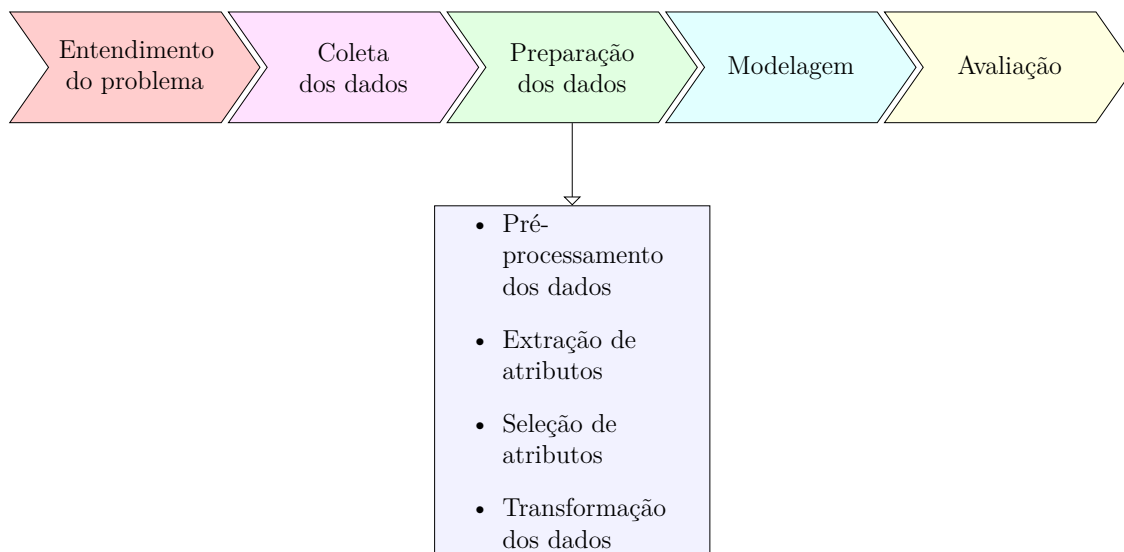


Fonte: Produção do autor.

4.3 Preparação dos dados

Nesta seção serão discutidas as etapas realizadas para a preparação dos dados para os algoritmos de AM, que inclui a limpeza e normalização dos dados, extração e SA. Figura 4.28 apresenta cada passo deste processo.

Figura 4.28 - Módulo sobre entendimento do problema.



Fonte: Produção do autor.

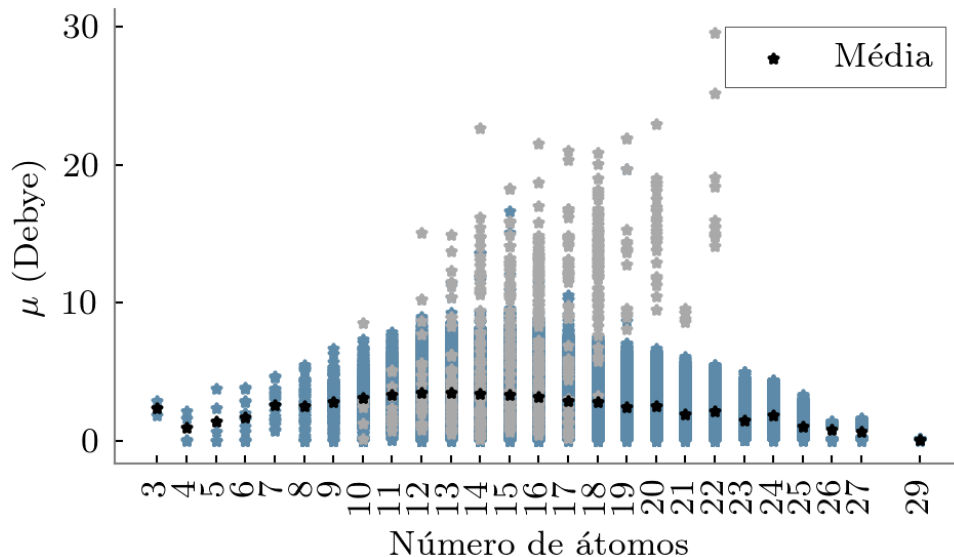
4.3.1 Pré-processamento dos dados

Com base nas análises apresentadas na seção 4.2.2, foram excluídas as moléculas que mostraram valores discrepantes na constante rotacional A; que contém múltiplas estruturas, já que a predição das propriedades de interesse pode ser prejudicada pela presença das múltiplas estruturas; e que tiveram problemas para serem interpretadas pelo RDKit. Além dessas exclusões, foram eliminadas também outras 3.054 moléculas que apresentaram falha no processo de consistência dos SMILES, como reportado em Ramakrishnan et al. (2014). A seguir serão apresentadas algumas diferenças observadas no conjunto QM9 após a eliminação de tais moléculas.

- Todas as moléculas com valor acima 20 Debye na propriedade μ foram removidas, como apresentado na Figura 4.29.
- Remoção da molécula (QM9 ID: 123126) com valor discrepante para a propriedade α . A molécula removida continha 22 átomos e apresentava

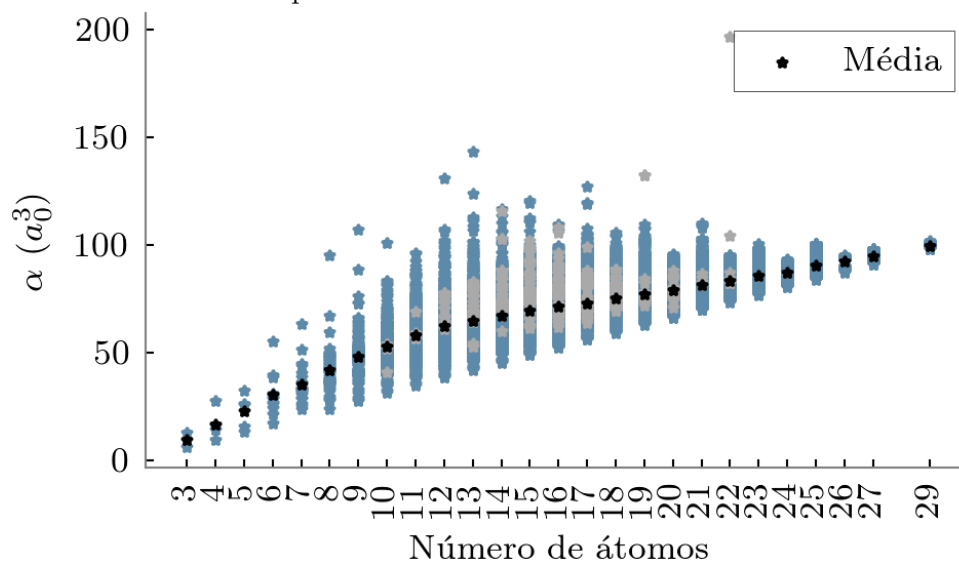
valor próximo a $200 a_0^3$, vide Figura 4.30.

Figura 4.29 - Dispersão dos valores por quantidade de átomos para a propriedade μ . Pontos em cinza representa as moléculas removidas.



Fonte: Produção do autor.

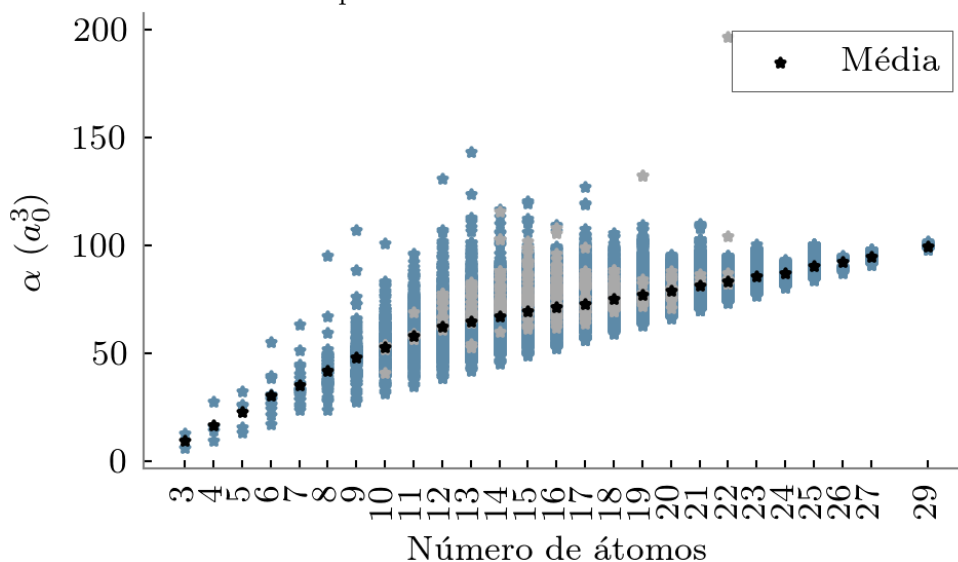
Figura 4.30 - Dispersão dos valores por quantidade de átomos para a propriedade α . Pontos em cinza representa as moléculas removidas.



Fonte: Produção do autor.

- No geral, as moléculas removidas tinham valores entre -4 eV a -6 eV na propriedade de ϵ_{HOMO} , conforme Figura 4.31.

Figura 4.31 - Dispersão dos valores por quantidade de átomos para a propriedade ϵ_{HOMO} . Pontos em cinza representa as moléculas removidas.



Fonte: Produção do autor.

Nenhuma das 24.446 moléculas que sofreram modificações no SMILES foram excluídas, já que as alterações realizadas não afetaram a estrutura molecular, mas sim uma segunda forma de expressar a mesma molécula pela representação SMILES. Dessa forma, para lidar com o problema de ambiguidade do SMILES, foi realizado um processo de extração de descritores para representar a molécula de uma única maneira.

Para o estudo comparativo da eficiência dos algoritmos de AM lidando apenas com um conjunto de dados composto por moléculas que apresentassem uma característica em específico, foi usado os subconjuntos com moléculas de apenas 17, 18 e 19 átomos. Estes subconjuntos do QM9 são os que contém a maior quantidade de moléculas, cada qual contendo, respectivamente, 16910, 17356 e 18129 moléculas.

Além disso, para cada um dos conjuntos citados anteriormente, foi feita uma divisão aleatória dos dados para a definição dos conjuntos de treino, validação e teste. O conjunto de treino é designado para o ajuste dos parâmetros do modelo e contém aproximadamente 78% dos dados, o conjunto de validação é usado para o ajuste

dos hiperparâmetros e contém 4 % dos dados, e o conjunto de teste para a avaliação final do modelo e apresenta 18 % dos dados.

4.3.2 Extração de atributos

O descritor molecular é o resultado de operações matemáticas e lógicas que codificam a representação química da molécula em um ou mais valores numéricos. Por exemplo, *fingerprints* descrevem a composição de uma molécula por meio de um conjunto de valores que apontam a presença ou quantidade das subestruturas de uma molécula. Também pode-se obter um descritor molecular por meio de resultados de experimentos laboratoriais.

Para as simulações deste trabalho, foi utilizado um pacote Python chamado Mordred (MORIWAKI et al., 2018) que oferece 1613 descritores moleculares baseados na representação em SMILES. O Mordred é um pacote baseado no objeto da molécula criado pelo RDKit, isto é, o Mordred utiliza as funções e variáveis, declaradas pelo RDKit, de cada molécula para fazer a extração dos descritores. Tais descritores descrevem inúmeras características das moléculas, desde das mais básicos, por exemplo, contagem de átomos específicos da molécula até descritores que codificam mais de uma propriedade da molécula. A seguir serão apresentados as 43 famílias de descritores oferecidas pelo Mordred.

- **Índice ABC** (do inglês, atom-bond connectivity) são descritores que se baseiam no grau dos vértices para descrever o grafo da molécula.
- **AcidBase**, realiza uma contagem de grupos ácidos e bases da molécula.
- **Matriz de adjacência**, representa um conjunto de métricas que resumizam a matriz de adjacência do grafo da molécula.
- **Aromaticidade**, realiza uma contagem de átomos e ligações aromáticas da molécula.
- **Contagem de átomos**, como o próprio nome sugere, este descritor realiza uma contagem de cada tipo de átomo da molécula. Por exemplo, quantidade de átomos de H, C, N, O ou F.
- **Autocorrelação**, descreve como uma propriedade atômica está distribuída no grafo da molécula.
- **BCUT** constrói uma versão modificada da matriz de adjacência, no qual a diagonal principal contém uma propriedade atômica e elementos não

pertencentes a diagonal principal descrevem o tipo de ligação entre dois pares de átomos da molécula. Em seguida, autovalores são usados para sumarizar tal representação matricial da molécula.

- **Índice de Balaban**, realiza o somatório da multiplicação da distância dos vértices adjacentes de todas as ligações da molécula, e normaliza o valor resultante pela quantidade de anéis.
- **Matriz de Barysz** é uma matriz de distância ponderada pela presença de heteroátomos e tipos de ligações da molécula. Descritores que estão englobados nesta categoria utilizam diferentes métricas para sumarizar essa matriz.
- **Bertz CT**, realiza a soma de um termo que descreve as ligações com um termo que representa a distribuição dos heteroátomos da molécula.
- **Contagem de ligações**, como o próprio nome sugere, este descritor realiza uma contagem de cada tipo de ligação da molécula. Por exemplo, quantidade ligação simples, dupla, tripla ou aromática.
- **Tipos de Carbono** descreve a conectividade dos átomos de carbono.
- **Chi** sumariza o grau dos vértices do grafo da molécula.
- **Constitucional**, realiza a média ou a soma de diversas propriedades atômicas considerando todos os átomos na molécula.
- **Matriz de Detour** descreve o maior caminho entre pares de vértices do grafo da molécula. Descritores que estão englobados nesta categoria utilizam diferentes métricas para sumarizar essa matriz.
- **Matriz de distância** descreve o menor caminho entre pares de vértices do grafo da molécula. Descritores que estão englobados nesta categoria utilizam diferentes métricas para sumarizar essa matriz.
- **Estado eletrotológico (E-State)** sumariza os estados eletrônicos e topológicos dos átomos na molécula através de uma contagem, soma ou valores máximos e mínimos.
- **Índice de conectividade excêntrica** sumariza as informações contidas na matriz de distância e adjacência. A matriz de distância e adjacência são utilizadas para determinar a excentricidade e grau dos átomos da molécula, respectivamente.

- **Átomo topoquímico estendido** (do inglês, *Extended Topochemical Atom*) são índices que agregam informações relacionadas ao número atômico, elétrons de valência e o número do período dos átomos da molécula.
- **Complexidade de fragmento** é um índice que sumariza os valores da quantidade de átomos, ligações e heteroátomos da molécula.
- **Framework** é um índice que descreve a proporção da quantidade de átomos em um *framework* em relação a quantidade total de átomos da molécula. Um *framework* é a subestrutura formada por dois anéis e os átomos que realizam uma ligação entre esses dois anéis.
- **Ligação de hidrogênio**, realiza uma contagem de doadores e aceptores de ligações de hidrogênio.
- **Entropia de Shannon** é uma família de descritores que utiliza a entropia de Shannon para quantificar as subestruturas da molécula.
- **Índice de morfologia Kappa** diz respeito a morfologia do grafo da molécula ao compará-la com dois grafos conexos, contendo o mesmo número de vértices, caracterizados por gerar a menor e a maior quantidade de caminhos dado um tamanho n .
- **Lipinski** utiliza de uma heurística para classificar as características de absorção e permeação da molécula. Para tanto, é usado como base a massa molecular, o $\log P$, e a contagem de doadores e aceptores de ligações de hidrogênio.
- **Volume de McGowan**, realiza a soma do volume atômico corrigido por um fator da quantidade de ligações na molécula.
- **Moe-Type**, realiza a soma de uma propriedade atômica para todos os átomos da molécula que respeitem a restrição sobre uma segunda propriedade atômica.
- **Distância de ligação molecular** sumariza informações da matriz de distância levando em consideração a valência dos átomos.
- **ID molecular**, realiza um somatória ou uma média dos pesos atribuídos para cada caminho percorrido no grafo da molécula.
- **Contagem de caminhos**, realiza a contagem de todos os caminhos restrito a um tamanho n sobre pares de vértices do grafo da molécula.

- **Polarizabilidade** descreve a polarizabilidade atômica e de ligação.
- **Contagem de anéis**, como o próprio nome sugere, este descritor realiza uma contagem dos anéis aromáticos na molécula. Por exemplo, quantidade de anéis de quatro, cinco, seis ou sete membros.
- **Ligações rotativas**, realiza uma contagem das ligações rotativas e também calcula a proporção dessas ligações em relação a todas as ligações existentes na molécula.
- **SLogP** calcula o coeficiente de partição usando o método de Wildman-Crippen.
- **Topológico área de superfície polar**, realiza a soma de superfície sobre todos os átomos polares da molécula.
- **Carga topológica**, realiza uma soma ou média da transferência de carga sobre os pares de átomos da molécula.
- **Índices topológicos**, realiza o cálculo de vários índices de grafo, como o diâmetro e raio, por exemplo.
- **Volume de van der Waals**, realiza a soma da superfície de van der Waals de cada átomo da molécula.
- **Informação do vértice adjacente** sumariza as ligações entre átomos pesados.
- **Contagem de percurso**, realiza a contagem dos percursos restrito a um tamanho n sobre os vértices do grafo da molécula.
- **Massa** diz respeito a massa molecular.
- **Índice de Wiener**, realiza a soma de todos os valores da matriz de distância.
- **Índice de Zagreb**, realiza a soma do grau dos vértices do grafo da molécula.

Após o processo de extração, foram removidos 539 descritores que tiveram variância igual a zero e 194 descritores que apresentaram valores nulos. Dentre os descritores removidos, as principais famílias encontradas foram: (i) contagem de átomos (por

exemplo, a quantidade de B, S, P e entre outros), o que faz sentido já que as moléculas do QM9 são compostas por átomos de C, H, O, N e F; (ii) contagem de anéis, neste caso, foram removidos 66 descritores de um total de 138, sendo que grande parte dos que foram removidos referem-se à moléculas com anéis compostos por mais de dez membros; (iii) autocorrelação e E-State, essas duas famílias de descritores juntos compõe mais de 900 dos 1613 descritores oferecidos pelo Mordred. Essa quantidade é justificada pelo fato de que tais descritores trabalham sobre a combinação de diversos parâmetros. Por exemplo, a contagem, soma, e o valor máximo e mínimo de uma característica da molécula. Com isso, se uma determinada característica não variar para todas as moléculas do conjunto, todas as operações realizadas sobre essa característica também terão valor igual para todas as moléculas.

Além dos descritores baseados na representação SMILES, foi extraído também a matriz de Coulomb da molécula. Diferentemente dos descritores anteriores, a matriz de Coulomb é baseada nas coordenadas cartesianas. O objetivo da utilização de tal descritor neste trabalho é comparar a eficiência de descritores que contém e não contém a dependência da geometria da molécula. Neste caso, escolheu-se a matriz de Coulomb devido a sua constante utilização em trabalhos da área.

4.3.3 Seleção de atributos

Como foi apresentado na seção 2.9, o processo de SA tem inúmeras vantagens, sendo as principais contribuições na facilitação da interpretabilidade dos dados e modelos, eliminação de atributos redundantes e custosos de serem calculados, ganho de tempo de treinamento, redução do consumo de memória do computador e aumento da acurácia do modelo.

Sendo assim, neste trabalho foram empregados cinco métodos de SA, este são: CFS, filtro por correlação usando a correlação de Pearson (Pcorr) e Spearman (Scorr), RF e LASSO. As três primeiras abordagens se configuram como sendo métodos de filtro enquanto as duas últimas como técnicas *ensembles*.

No caso dos métodos de filtro por correlação, RF e LASSO, foi necessário ajustar os hiperparâmetros dos métodos. Na abordagem filtro por correlação, o limiar δ para remoção de atributos redundantes foi definido como 0,8 em ambas métricas de correlação. O valor de δ foi estipulado baseando-se em um método iterativo, no qual, inicialmente, δ recebia o valor 1 e a cada iteração era diminuído em 0,05 e então, os descritores selecionados eram avaliados usando uma regressão linear. Posteriormente, o melhor resultado era escolhido. No geral, o valor de δ igual a 0,8

apresentou excelentes resultados.

Para a seleção feita por RF, definiu-se o número de árvores igual a 300, a quantidade de atributos selecionados aleatoriamente para cada nó igual à metade do tamanho do vetor de características e o erro quadrático médio como critério de avaliação para divisão dos nós. Essa definição foi baseada em diversos experimentos sobre os conjuntos de treino e validação. Além disso, para reduzir a chances de construir as árvores sobre atributos correlacionados, foram desconsiderados descritores com intercorrelação absoluta acima de 0,8. Na seleção da quantidade de atributos m sobre o ranqueamento feito pelo RF, foi utilizado uma regressão linear na avaliação de vários valores de m . Neste caso, foi observado que m equivalente a 140 conseguiu ponderar os objetivos da redução do erro da regressão com a minimizando os atributos selecionados.

Por fim, para o algoritmo LASSO, notou-se que λ igual a 0,02 foi suficiente para obter bons resultados. Tabela 4.3 apresenta a quantidade média de descritores selecionados por cada método de seleção em cada conjunto de dados usado neste trabalho.

4.3.4 Normalização dos dados

Métodos de normalização inserem os valores de uma variável dentro de um intervalo de valores. Em alguns casos, essa etapa do pré-processamento pode beneficiar os algoritmos de AM, como é o caso das RNAs, enquanto em outros casos a normalização não faz efeito, por exemplo, as árvores de decisão. Neste trabalho, foi empregado a normalização que modifica a média $\bar{x}$ para 0 e desvio padrão σ para 1 conforme apresentado na Equação abaixo.

$$x'_i = \frac{x_i - \bar{x}}{\sigma} \quad (4.1)$$

no qual x'_i corresponde aos valores normalizados do atributo x_i . As propriedades preditas y_i também foram normalizadas para o processo de treinamento das RNAs. Neste caso, primeiramente, as propriedades foram divididas pela quantidade de átomos n presente na molécula e posteriormente, normalizada seguindo a equação 4.1. Após o treinamento, o valor predito $\hat{y}_i$ sofria uma transformação inversa, em outras palavras, $\hat{y}_i$ era multiplicado por σ , acrescido de $\bar{y}_i$ e em seguida, multiplicado por n . Logo após, o cálculo do erro era feito sobre a transformação inversa de $\hat{y}_i$ e y_i .

Tabela 4.3 - Quantidade média de descritores selecionados por conjunto de dados. A média foi calculada sobre as seleções por propriedade. Conjunto 17, 18 e 19 representa os subconjuntos composto por moléculas com 17, 18 e 19 átomos, respectivamente e # denomina quantidade.

	Método	# descritores selecionados
QM9	CFS	75
	Pcorr	108
	Scorr	115
	RF	140
	LASSO	110
Conjunto 17	CFS	132
	Pcorr	169
	Scorr	169
	RF	140
	LASSO	100
Conjunto 18	CFS	113
	Pcorr	174
	Scorr	184
	RF	140
	LASSO	95
Conjunto 19	CFS	115
	Pcorr	167
	Scorr	175
	RF	140
	LASSO	87

Fonte: Produção do autor.

4.4 Modelagem

No processo de otimização do modelo de AM, o ajuste correto dos hiperparâmetros pode influenciar positivamente na acurácia final do algoritmo. Neste trabalho, o ajuste dos hiperparâmetros sobre os algoritmos destinados para a SA fora anteriormente explicado na seção 4.3.3. Em relação à rede Perceptron de múltiplas camadas, usou-se uma estratégia de busca para encontrar a melhor combinação sobre a quantidade de camada, quantidade de neurônio por camada e taxa de aprendizagem. Tabela 4.4 apresenta todos os possíveis valores definidos para os hiperparâmetros da rede. O funcionamento da estratégia de busca em questão é dada, inicialmente, pela definição da quantidade ótima de neurônios na primeira camada. A escolha dessa quantidade é feita dentro de um processo iterativo que analisa a acurácia do

modelo ao adicionar ou reduzir neurônios na camada. Dessa forma, se a acurácia é melhorada ao aumentar a quantidade de neurônios, então deve-se continuar aumentando o número total de neurônios e, caso contrário, deve-se diminuir a quantidade de neurônios. Após estabelecer a quantidade de neurônios para a primeira camada, essa abordagem de busca é repetida na segunda camada. Vale ressaltar que para cada uma das quantidades de neurônios analisadas é verificado qual taxa de aprendizagem apresenta o melhor resultado.

Tabela 4.4 - População de valores definidos para ajustar os hiperparâmetros das RNAs.

Hiperparâmetros	Valores
Quantidade de camadas ocultas	1, 2
Quantidade de neurônios por camadas	5 até 505 passo 25
Taxa de aprendizagem	0,001, 0,01, 0,1
Tamanho do batch	562
Função de ativação	ReLU
Otimizador	Adam - $\beta_1 = 0,9$, $\beta_2 = 0,999$

Fonte: Produção do autor.

4.5 Avaliação

Na seção anterior, o processo de otimização foi realizado através do conjunto de validação. Nesta etapa, um outro conjunto, chamado de conjunto de teste, é empregado com o objetivo de averiguar a capacidade de generalização do modelo. Essa capacidade de generalização é avaliada sobre a rede já treinada e com hiperparâmetros otimizados usando uma métrica para o cálculo do erro. Neste trabalho, a métrica estabelecida foi EAM definida na Equação 2.1.

5 RESULTADOS E DISCUSSÕES

5.1 Seleção de atributos

Neste trabalho, vários foram os descritores extraídos pelo pacote Mordred, ao todo foram 1613 baseados na representação SMILES da molécula. Desse total, 733 descritores foram removidos por terem variância igual a zero ou apresentarem valores nulos, fazendo com que apenas 880 descritores fossem aplicados no processo de SA e na predição de propriedades.

A SA foi realizada no contexto supervisionado, isto é, encontre os descritores mais relevantes dada a propriedade de interesse. Com isso, foram realizadas 45 seleções por conjunto de dados, considerando os cinco métodos de SA utilizados e as nove propriedades preditas. No apêndice são apresentados todos os descritores selecionados para cada um dos casos. Na média, foram selecionados 132 descritores por método.

Tabela 5.1 apresenta os vinte descritores mais selecionados, do conjunto QM9, agrupados pela família pertencente no pacote Mordred. O ranqueamento dos descritores baseou-se no número de ocorrência de seleção em todos os métodos de SA e propriedades moleculares utilizados. Neste contexto, notou-se que o atributo mais selecionado foi o CIT₁, contendo 35 seleções de um total de 45, e que 276 descritores não foram escolhidos por nenhuma SA.

Tabela 5.1 - Descritores mais selecionados entre todos os métodos de SA e propriedades preditas do conjunto QM9. O número entre parêntese representa a quantidade de ocorrência para cada atributo.

Classe do descritor	Descritores
Moe-Type	PEOE_VSA ₇ (34), SlogP_VSA ₁ (33), SlogP_VSA ₅ (32), SMR_VSA ₇ (31), PEOE_VSA ₆ (31), VSA_EState ₇ (30), PEOE_VSA ₁ (30), PEOE_VSA ₃ (30), PEOE_VSA ₁₀ (29), VSA_EState ₅ (29), EState_VSA ₄ (29), PEOE_VSA ₁₂ (28), PEOE_VSA ₁₀ (28)
Autocorrelação	ATSC0i (31), ATSC3dv (30), ATSC4d (30)
Entropia de Shannon	CIT ₁ (35), CIT ₀ (32)
LogS	FilterItLogS (31)
Matriz de adjacência	VR1_A (29)

Fonte: Produção do autor.

O mesmo ranqueamento feito na Tabela 5.1 foi aplicado para os subconjuntos com moléculas de 17, 18 e 19 átomos. Tabela 5.2 apresenta os descritores selecionados para este caso, note que a quantidade máxima de ocorrência é de 135 e que duas classes de atributo (Moe-Type e autocorrelação) se mantiveram em ambas Tabelas, sendo que os descritores da classe Moe-Type continuaram sendo os mais selecionados.

Tabela 5.2 - Descritores mais selecionados entre os subconjuntos com moléculas de 17, 18 e 19 átomos.

Classe do descritor	Descritores
Moe-Type	SlogP_VSA ₄ (114), VSA_EState ₇ (112), SlogP_VSA ₁ (112), PEOE_VSA ₇ (112), PEOE_VSA ₈ (110), SMR_VSA ₇ (109), SlogP_VSA ₂ (108), PEOE_VSA ₆ (108), SlogP_VSA ₅ (106), PEOE_VSA ₉ (106), SMR_VSA ₆ (104), PEOE_VSA ₁₀ (104), PEOE_VSA ₃ (102), EState_VSA ₄ (102), EState_VSA ₉ (101), EState_VSA ₈ (101), VSA_EState ₄ (100), PEOE_VSA ₂ (100)
Índice de conectividade excêntrica	ECIndex (104)
Autocorrelação	ATSC3dv (99)

Fonte: Produção do autor.

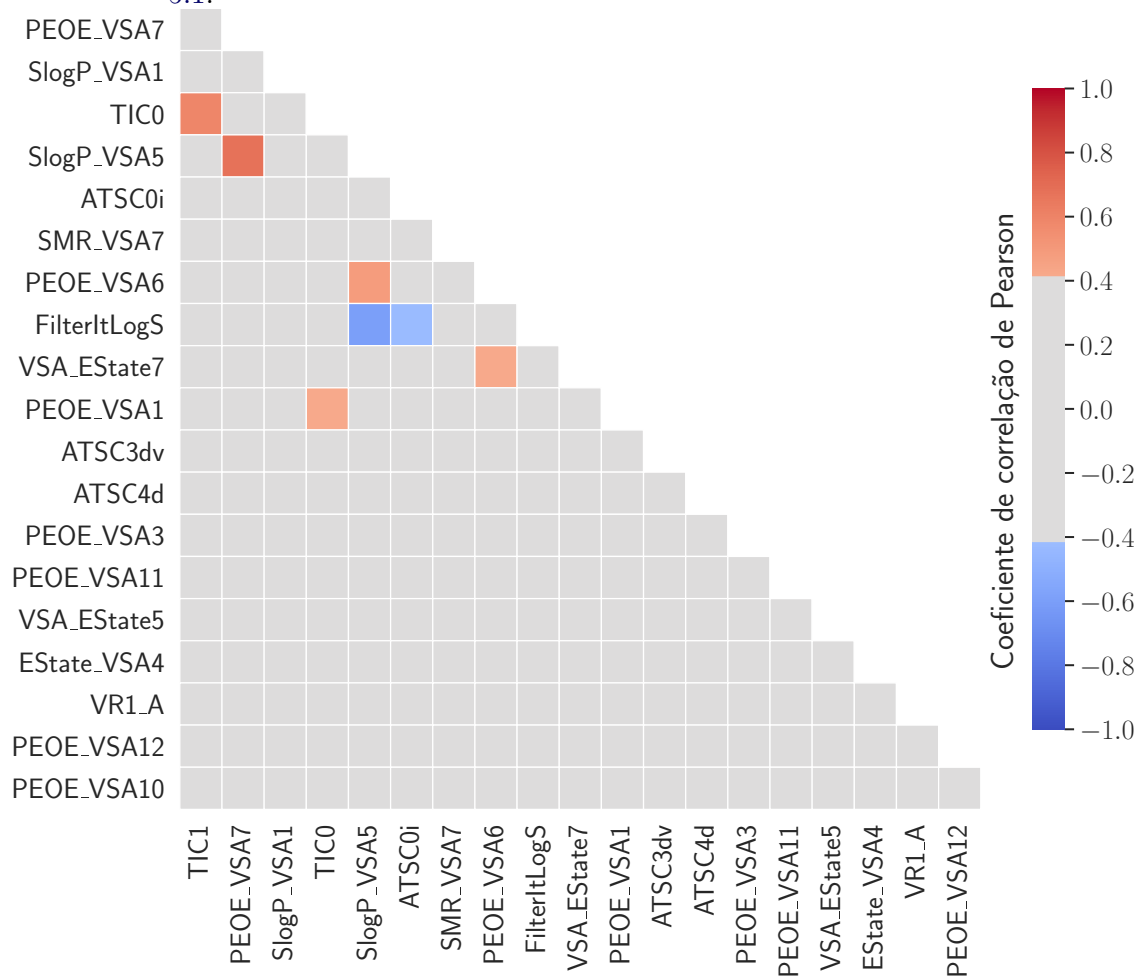
Ao analisar a correlação de Pearson entre os descritores apresentados na Tabela 5.1, observou-se que a correlação média entre cada atributo e propriedade é de 0,27 e que a intercorrelação absoluta dos descritores não passa do coeficiente 0,4 para a maior parte dos casos. Figura 5.1 mostra o mapa de calor da correlação de Pearson entre pares de descritores.

A fraca intercorrelação apresentada na Figura 5.1 sugere que a agregação de todos os descritores como entrada do modelo contribua para uma alta informatividade. Além disso, a fraca dependência linear entre os descritores e as propriedades alvos, mostra que a predição da saída desejada depende de modelos que aprendam uma função-conceito não linear entre as entradas e saída. Dessa forma, após a SA utilizou-se o modelo de rede Perceptron de múltiplas camadas para uma combinação não linear dos descritores encontrados. Os resultados obtidos foram comparados com o desempenho da mesma rede usando os 880 descritores iniciais.

5.1.1 Análise dos descritores

Os descritores mostrados nas Tabelas 5.1 e 5.2 são, em sua maioria, gerados pela combinação de duas propriedades atômicas. Por exemplo, os descritores da classe Moe-Type fazem a soma de uma propriedade atômica considerando apenas os átomos que contenham uma segunda propriedade atômica de interesse. Isto é, uma forma de quantificar a molécula selecionando apenas átomos que apresentem uma propriedade atômica similar. A seguir será descrito o processo de cálculo dos descritores da classe Moe-Type.

Figura 5.1 - Correlação de Pearson entre os pares de descritores apresentados na Tabela 5.1.



Fonte: Produção do autor.

- **PEOE_VSA_k**: Soma da superfície de van der Waals (do inglês *van der*

Waals surface area - VSA) para todos os átomos com carga parcial atômica de Gasteiger, usando o esquema de equalização parcial da eletronegatividade orbital (do inglês *partial equalization of orbital electronegativities*, PEOE) restrito as condições apresentadas na Tabela abaixo (LABUTE, 2000).

Tabela 5.3 - Restrições no somatório do atributo PEOE_VSA_k para cada condição *k*.

Valores de <i>k</i>	Restrição
PEOE_VSA ₁	$-\text{inf} < \text{PEOE} < -0,3$
PEOE_VSA ₂	$-0,30 \leq \text{PEOE} < -0,25$
PEOE_VSA ₃	$-0,25 \leq \text{PEOE} < -0,20$
PEOE_VSA ₆	$-0,10 \leq \text{PEOE} < -0,05$
PEOE_VSA ₇	$-0,05 \leq \text{PEOE} < 0$
PEOE_VSA ₈	$0 \leq \text{PEOE} < 0,05$
PEOE_VSA ₉	$0,05 \leq \text{PEOE} < 0,10$
PEOE_VSA ₁₀	$0,10 \leq \text{PEOE} < 0,15$
PEOE_VSA ₁₁	$0,15 \leq \text{PEOE} < 0,20$
PEOE_VSA ₁₂	$0,20 \leq \text{PEOE} < 0,25$

Fonte: Produção do autor.

- **SlogP_VSA_k**: Soma dos VSA para todos os átomos com a propriedade LogP, usando o método de Wildman-Crippen (SLogP), restrito as condições apresentadas na Tabela abaixo.

Tabela 5.4 - Restrições no somatório do atributo SlogP_VSA_k para cada condição *k*.

Valores de <i>k</i>	Restrição
SlogP_VSA ₁	$-\text{inf} < \text{SLogP} < -0,40$
SlogP_VSA ₂	$-0,40 \leq \text{SLogP} < -0,20$
SlogP_VSA ₄	$0 \leq \text{SLogP} < 0,10$
SlogP_VSA ₅	$0,10 \leq \text{SLogP} < 0,15$

Fonte: Produção do autor.

- **EState_VSA_k**: Soma dos VSA para todos os átomos com o índice do estado eletrotopológico (E-State) restrito as condições apresentadas na Ta-

bela a seguir (HALL; KIER, 1995).

Tabela 5.5 - Restrições no somatório do atributo EState_VSA_k para cada condição *k*.

Valores de <i>k</i>	Restrição
ESate_VSA ₄	$0,72 \leq EState < 1,17$
ESate_VSA ₈	$2,05 \leq EState < 4,69$
ESate_VSA ₉	$4,69 \leq EState < 9,17$

Fonte: Produção do autor.

- **VSA__EState_k**: Soma dos E-State para todos os átomos com o VSA restrito as condições apresentadas na Tabela a seguir.

Tabela 5.6 - Restrições no somatório do atributo VSA_EState_k para cada condição *k*.

Valores de <i>k</i>	Restrição
VSA_EState ₄	$5,41 \leq VSA < 5,74$
VSA_EState ₅	$5,74 \leq VSA < 6,00$
VSA_EState ₇	$6,07 \leq VSA < 6,45$

Fonte: Produção do autor.

- **SMR__VSA_k**: Soma dos VSA para todos os átomos com refração molar usando o método de Wildman-Crippen (SMR) restrito as condições apresentadas na Tabela abaixo.

Tabela 5.7 - Restrições no somatório do atributo SMR_VSA_k para cada condição *k*.

Valores de <i>k</i>	Restrição
SMR_VSA ₆	$2,75 \leq SMR < 3,05$
SMR_VSA ₇	$3,05 \leq SMR < 3,63$

Fonte: Produção do autor.

Segundo Labute (2000), os descritores do tipo PEOE_VSA_k capturam interações eletrostáticas, enquanto SlogP_VSA_k é responsável por capturar efeitos hidrofóbicos e hidrofílicos e SMR_VSA_k permite obter informação a respeito da polarizabilidade, no qual k representa a restrição imposta.

Os descritores relacionados a classe da autocorrelação sumarizam como uma propriedade atômica x está distribuída pela estrutura topológica da molécula, ao fazer a soma da autocorrelação de x para cada par de átomo que respeita uma distância de vizinhança d (TODESCHINI; CONSONNI, 2008). Neste contexto, as propriedades x consideradas como relevantes foram elétrons de valência (ATSC3dv), elétrons sigma (ATSC4d) e potencial de ionização (ATSC0i), com d igual a 3, 4 e 0, respectivamente.

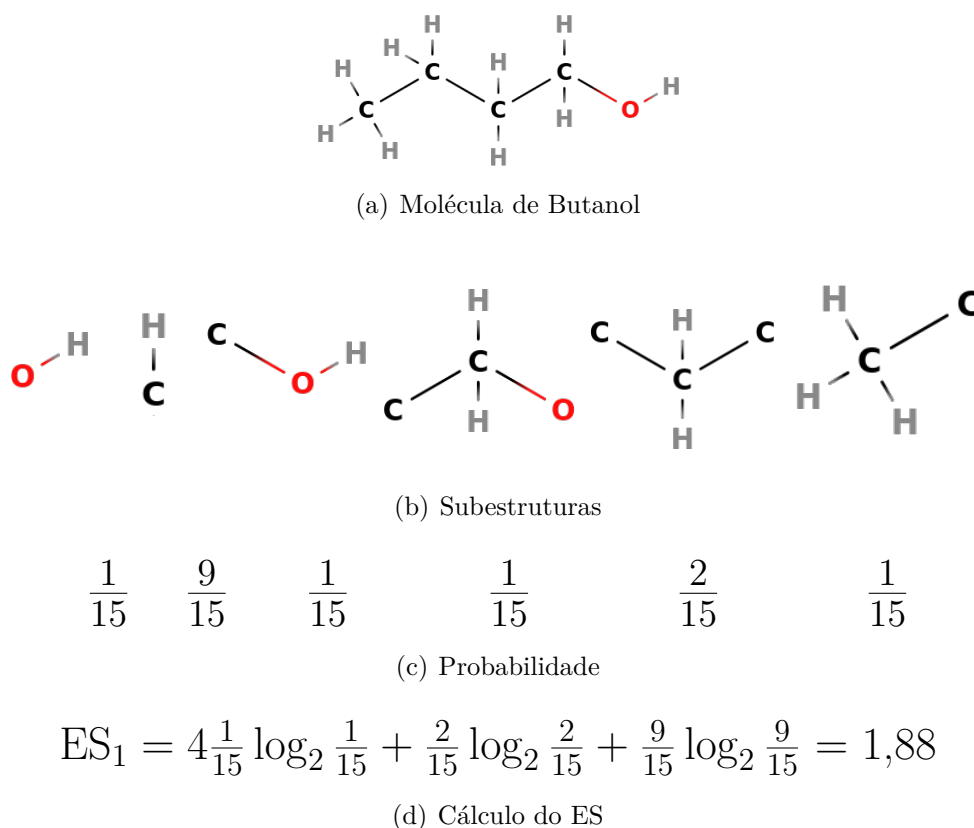
O atributo LogS estima a solubilidade dos compostos químicos baseando na contribuição de cada fragmento da molécula, essa contribuição é definida usando um somatório sobre os fragmentos moleculares ponderados por pesos (TODESCHINI; CONSONNI, 2008). Neste contexto, a solubilidade é considerada importante, pois está relacionada com algumas propriedades preditas, como o momento dipolar e polarização (μ) e polarizabilidade isotrópica (α).

O índice de conectividade excêntrica (ECIndex) é calculado sobre a matriz de distância e adjacência. A matriz de distância é utilizada para determinar a excentricidade e_i de um vértice (átomo), que é definida como o maior caminho mínimo entre o i -ésimo e qualquer outro vértice. Já a matriz de adjacência é usada para o cálculo do grau g_i do vértice i e define o número de arestas incidentes em i . A equação 5.1 apresenta a fórmula para o cálculo do índice (TODESCHINI; CONSONNI, 2008).

$$\text{ECIndex} = \sum_{i=1}^A e_i g_i \quad (5.1)$$

A entropia de Shannon (ES) mensura a quantidade de informação em uma mensagem. No caso dos descritores baseados em ES, a mensagem é a molécula e a informação são as subestruturas ou fragmentos da molécula. Com isso, se todas as subestruturas do grafo da molécula ocorrem na mesma proporção, então a ES é máxima. Caso a molécula tenha um tipo de fragmento, a ES será 0. Figura 5.2 apresenta o cálculo do ES para uma molécula de Butanol, no qual considerou subestruturas do grafo da molécula formadas por átomos vizinhos de ordem 1 (DEVILLERS; BALABAN, 2000).

Figura 5.2 - Cálculo do ES para a molécula de Butanol considerando fragmentos formados por vizinhos de ordem 1.



Fonte: Adaptada de [Devillers e Balaban \(2000\)](#).

O conteúdo da informação total (CIT) é definido na equação 5.2.

$$CIT_m = A \cdot ES_m, \quad (5.2)$$

no qual, m define o caminho mínimo entre pares de átomos para a formação dos fragmentos e A é a quantidade de átomos na molécula. Neste caso, CIT_0 e CIT_1 significam o CIT sobre cada espécie química e das subestruturas com átomos vizinhos de ordem máxima igual a 1, respectivamente.

O atributo VR1_A sumariza a matriz de adjacência seguindo a equação 5.3, no qual ℓ são os coeficientes do autovetor associado com o menor autovalor da matriz de adjacência ([TODESCHINI; CONSONNI, 2008](#)).

$$\text{VR1_A} = \sum_{i,j \in \text{ligações}} (\ell_i \ell_j)^{-\frac{1}{2}} \quad (5.3)$$

5.2 Predição das propriedades

Para a tarefa de predição foi construída uma rede neural de múltiplas camadas para cada uma das nove propriedades moleculares no menor estado de energia. Tabelas 5.8, 5.9, 5.10 e 5.11 apresentam o EAM, para todas as propriedades e métodos de SA, obtido sobre o conjunto de teste do QM9 e os subconjuntos com moléculas de 17, 18 e 19 átomos, respectivamente. O cálculo do erro levou em conta a média de cinco execuções em cada casos para a verificação da variância do modelo.

As predições para as propriedades U_0 , U , H , e G mostraram um EAM próximo. Por outro lado, a propriedade μ teve um erro de 0.5 Debye em todos os casos, o que pode ser justificado pela alta dependência relacionada as coordenadas cartesianas dos átomos da molécula. Figura 5.3 exibe o erro de predição ao mostrar os valores reais contra os valores preditos. Um bom desempenho na predição resulta em uma forte correlação linear positiva entre os valores reais e preditos.

Tabela 5.8 - Erro de generalização, em termos de EAM, sobre o conjunto QM9. Nenhum significa que todos os descritores foram usados como entrada.

Propriedades	CFS	Pcorr	Scorr	RF	LASSO	Nenhum
U_0 (eV)	0,0765 ($\pm 0,0004$)	0,0808 ($\pm 0,0014$)	0,0775 ($\pm 0,001$)	0,0685 ($\pm 0,0029$)	0,084 ($\pm 0,0028$)	0,0573 ($\pm 0,0004$)
U (eV)	0,1102 ($\pm 0,0009$)	0,082 ($\pm 0,0005$)	0,0799 ($\pm 0,0007$)	0,0722 ($\pm 0,0036$)	0,085 ($\pm 0,001$)	0,0582 ($\pm 0,001$)
H (eV)	0,1104 ($\pm 0,0011$)	0,0805 ($\pm 0,0007$)	0,0776 ($\pm 0,0005$)	0,0737 ($\pm 0,0057$)	0,0848 ($\pm 0,002$)	0,0575 ($\pm 0,001$)
G (eV)	0,0929 ($\pm 0,0013$)	0,08 ($\pm 0,0003$)	0,0812 ($\pm 0,0008$)	0,0689 ($\pm 0,0029$)	0,08 ($\pm 0,0015$)	0,0562 ($\pm 0,001$)
C_v (cal mol K ⁻¹)	0,1446 ($\pm 0,0014$)	0,1662 ($\pm 0,0024$)	0,1708 ($\pm 0,0009$)	0,1458 ($\pm 0,0062$)	0,1483 ($\pm 0,0011$)	0,1223 ($\pm 0,002$)
μ (Debye)	0,5663 ($\pm 0,0015$)	0,5339 ($\pm 0,0031$)	0,5396 ($\pm 0,0016$)	0,542 ($\pm 0,0024$)	0,5565 ($\pm 0,0036$)	0,523 ($\pm 0,006$)
α (a ₀ ³)	0,3973 ($\pm 0,0013$)	0,3747 ($\pm 0,0037$)	0,3741 ($\pm 0,0062$)	0,3734 ($\pm 0,0035$)	0,365 ($\pm 0,0028$)	0,3095 ($\pm 0,002$)
ϵ_{HOMO} (eV)	0,1071 ($\pm 0,0004$)	0,1089 ($\pm 0,0011$)	0,1012 ($\pm 0,0009$)	0,1082 ($\pm 0,0011$)	0,1179 ($\pm 0,001$)	0,0952 ($\pm 0,001$)
Δ_ϵ (eV)	0,175 ($\pm 0,001$)	0,1746 ($\pm 0,0012$)	0,185 ($\pm 0,0011$)	0,1474 ($\pm 0,0011$)	0,1666 ($\pm 0,0014$)	0,1369 ($\pm 0,001$)

Fonte: Produção do autor.

Na Figura 5.3 é interessante observar que ϵ_{HOMO} apresenta um EAM maior em relação as propriedades termodinâmicas, mesmo com um intervalo de valores em eV menor. Em contrapartida, ao comparar os resultados entre ϵ_{HOMO} e Δ_ϵ , Δ_ϵ apresenta um EAM maior e uma magnitude em eV similar as propriedades termodinâmicas. Neste contexto, propriedades termodinâmicas demonstraram-se mais fáceis de serem preditas por meio de descritores baseados em SMILES.

Tabela 5.9 - Erro de generalização, em termos de EAM, sobre o subconjunto com moléculas de 17 átomos. Nenhum significa que todos os descritores foram usados como entrada.

Propriedades	CFS	Pcorr	Scorr	RF	LASSO	Nenhum
U_0 (eV)	0,1027 ($\pm 0,002$)	0,0932 ($\pm 0,002$)	0,0961 ($\pm 0,001$)	0,1039 ($\pm 0,003$)	0,1204 ($\pm 0,004$)	0,0896 ($\pm 0,002$)
U (eV)	0,096 ($\pm 0,002$)	0,1081 ($\pm 0,004$)	0,0968 ($\pm 0,001$)	0,0928 ($\pm 0,001$)	0,1299 ($\pm 0,007$)	0,0834 ($\pm 0,001$)
H (eV)	0,1091 ($\pm 0,002$)	0,0915 ($\pm 0,001$)	0,0958 ($\pm 0,002$)	0,0987 ($\pm 0,001$)	0,1172 ($\pm 0,001$)	0,0818 ($\pm 0,002$)
G (eV)	0,1162 ($\pm 0,009$)	0,0912 ($\pm 0,001$)	0,0902 ($\pm 0,001$)	0,0932 ($\pm 0,002$)	0,122 ($\pm 0,001$)	0,0789 ($\pm 0,002$)
C_v (cal mol K ⁻¹)	0,1761 ($\pm 0,003$)	0,1556 ($\pm 0,001$)	0,1545 ($\pm 0,001$)	0,1565 ($\pm 0,001$)	0,1721 ($\pm 0,001$)	0,1482 ($\pm 0,001$)
μ (Debye)	0,690 ($\pm 0,006$)	0,6879 ($\pm 0,005$)	0,6726 ($\pm 0,006$)	0,698 ($\pm 0,007$)	0,6816 ($\pm 0,008$)	0,6758 ($\pm 0,007$)
α (a ₀ ³)	0,4647 ($\pm 0,003$)	0,4477 ($\pm 0,005$)	0,4627 ($\pm 0,005$)	0,5167 ($\pm 0,04$)	0,4813 ($\pm 0,005$)	0,4658 ($\pm 0,011$)
ϵ_{HOMO} (eV)	0,1365 ($\pm 0,002$)	0,1238 ($\pm 0,001$)	0,1249 ($\pm 0,001$)	0,1218 ($\pm 0,001$)	0,1503 ($\pm 0,002$)	0,1327 ($\pm 0,001$)
Δ_ϵ (eV)	0,227 ($\pm 0,004$)	0,1781 ($\pm 0,002$)	0,1805 ($\pm 0,003$)	0,1788 ($\pm 0,001$)	0,2118 ($\pm 0,004$)	0,1697 ($\pm 0,002$)

Fonte: Produção do autor.

Um outro ponto relevante a ser observado é que a rede treinada usando todo o conjunto de descritores teve um melhor desempenho na grande parte dos casos. Pode-se justificar tal fato levando em consideração o processo de aprendizagem de uma RNA. RNAs ajustam pesos sinápticos de forma a reduzir o erro de predição. Este ajuste pode realizar uma SA atribuindo pesos próximos a zero para atributos irrelevantes. Sendo assim, a melhoria dos resultados pode ser explicada considerando que a SA foi enviesada, em outras palavras, teve por objetivo melhorar o desempenho da rede. Contudo, interpretar o resultado de uma SA de uma rede neural de múltiplas camadas não é trivial, pois a seleção envolve processo de combinação de pesos e transformações não lineares.

Observando a Tabela 5.8, nota-se uma tendência relacionada aos resultados obtidos com a SA. Em geral, as propriedades termodinâmicas tiveram melhores desempenhos usando os descritores selecionados pelo RF juntamente com Δ_ϵ , C_v teve uma boa acurácia com CFS, μ com os descritores selecionados pelo método Pcorr, α com LASSO e ϵ_{HOMO} usando o Scorr. Em suma, é possível ranquear os desempenhos do SA na seguinte ordem: (i) RF, (ii) Scorr, (iii) LASSO, (iv) Pcorr e (v) CFS.

Tabela 5.10 - Erro de generalização, em termos de EAM, sobre o subconjunto com moléculas de 18 átomos. Nenhum significa que todos os descritores foram usados como entrada.

Propriedades	CFS	Pcorr	Scorr	RF	LASSO	Nenhum
U_0 (eV)	0,1008 ($\pm 0,003$)	0,0858 ($\pm 0,001$)	0,0812 ($\pm 0,002$)	0,0824 ($\pm 0,001$)	0,1079 ($\pm 0,003$)	0,0729 ($\pm 0,002$)
U (eV)	0,0877 ($\pm 0,002$)	0,0847 ($\pm 0,001$)	0,0855 ($\pm 0,0004$)	0,0901 ($\pm 0,003$)	0,1122 ($\pm 0,002$)	0,0726 ($\pm 0,0$)
H (eV)	0,0886 ($\pm 0,001$)	0,0963 ($\pm 0,001$)	0,0884 ($\pm 0,001$)	0,09 ($\pm 0,003$)	0,1101 ($\pm 0,002$)	0,0722 ($\pm 0,001$)
G (eV)	0,102 ($\pm 0,003$)	0,0889 ($\pm 0,001$)	0,0799 ($\pm 0,001$)	0,0821 ($\pm 0,002$)	0,1033 ($\pm 0,002$)	0,0702 ($\pm 0,001$)
C_v (cal mol K ⁻¹)	0,1613 ($\pm 0,005$)	0,1453 ($\pm 0,001$)	0,1432 ($\pm 0,001$)	0,1426 ($\pm 0,002$)	0,1711 ($\pm 0,002$)	0,1443 ($\pm 0,002$)
μ (Debye)	0,6024 ($\pm 0,015$)	0,6155 ($\pm 0,003$)	0,6148 ($\pm 0,003$)	0,5942 ($\pm 0,004$)	0,618 ($\pm 0,006$)	0,6061 ($\pm 0,005$)
α (a ₀ ³)	0,4951 ($\pm 0,001$)	0,4115 ($\pm 0,003$)	0,4087 ($\pm 0,003$)	0,4212 ($\pm 0,003$)	0,4601 ($\pm 0,013$)	0,3766 ($\pm 0,002$)
ϵ_{HOMO} (eV)	0,1369 ($\pm 0,001$)	0,1186 ($\pm 0,001$)	0,1287 ($\pm 0,001$)	0,1306 ($\pm 0,001$)	0,1414 ($\pm 0,002$)	0,1246 ($\pm 0,002$)
Δ_ϵ (eV)	0,2188 ($\pm 0,004$)	0,1651 ($\pm 0,001$)	0,1638 ($\pm 0,001$)	0,1736 ($\pm 0,004$)	0,1924 ($\pm 0,001$)	0,1616 ($\pm 0,001$)

Fonte: Produção do autor.

Na Tabela 5.12 tem-se uma comparação dos resultados obtidos pela rede usando os descritores baseados na representação SMILES e usando a matriz de Coulomb. Para todas as propriedades, os descritores derivados do SMILES tiveram um desempenho melhor.

Tabela 5.11 - Erro de generalização, em termos de EAM, sobre o subconjunto com moléculas de 19 átomos. Nenhum significa que todos os descritores foram usados como entrada.

Propriedades	CFS	Pcorr	Scorr	RF	LASSO	Nenhum
U_0 (eV)	0,1081 ($\pm 0,001$)	0,093 ($\pm 0,003$)	0,0821 ($\pm 0,001$)	0,0916 ($\pm 0,003$)	0,1022 ($\pm 0,003$)	0,0721 ($\pm 0,002$)
U (eV)	0,1003 ($\pm 0,001$)	0,0839 ($\pm 0,001$)	0,0859 ($\pm 0,002$)	0,0839 ($\pm 0,001$)	0,1002 ($\pm 0,002$)	0,0736 ($\pm 0,001$)
H (eV)	0,0948 ($\pm 0,002$)	0,0824 ($\pm 0,001$)	0,0921 ($\pm 0,003$)	0,085 ($\pm 0,002$)	0,0996 ($\pm 0,001$)	0,0741 ($\pm 0,001$)
G (eV)	0,0883 ($\pm 0,003$)	0,0821 ($\pm 0,001$)	0,088 ($\pm 0,001$)	0,0865 ($\pm 0,001$)	0,1031 ($\pm 0,002$)	0,0718 ($\pm 0,001$)
C_v (cal mol K ⁻¹)	0,1751 ($\pm 0,001$)	0,1585 ($\pm 0,002$)	0,1723 ($\pm 0,025$)	0,1483 ($\pm 0,0003$)	0,1742 ($\pm 0,002$)	0,1522 ($\pm 0,002$)
μ (Debye)	0,53 ($\pm 0,004$)	0,5248 ($\pm 0,005$)	0,5213 ($\pm 0,006$)	0,5275 ($\pm 0,002$)	0,5515 ($\pm 0,014$)	0,5245 ($\pm 0,004$)
α (a ₀ ³)	0,4537 ($\pm 0,007$)	0,3913 ($\pm 0,007$)	0,3773 ($\pm 0,002$)	0,4163 ($\pm 0,004$)	0,4214 ($\pm 0,009$)	0,397 ($\pm 0,005$)
ϵ_{HOMO} (eV)	0,1246 ($\pm 0,002$)	0,1128 ($\pm 0,001$)	0,1119 ($\pm 0,001$)	0,1159 ($\pm 0,001$)	0,1307 ($\pm 0,001$)	0,1162 ($\pm 0,001$)
Δ_ϵ (eV)	0,1861 ($\pm 0,002$)	0,161 ($\pm 0,001$)	0,1661 ($\pm 0,001$)	0,1614 ($\pm 0,001$)	0,1948 ($\pm 0,002$)	0,1607 ($\pm 0,001$)

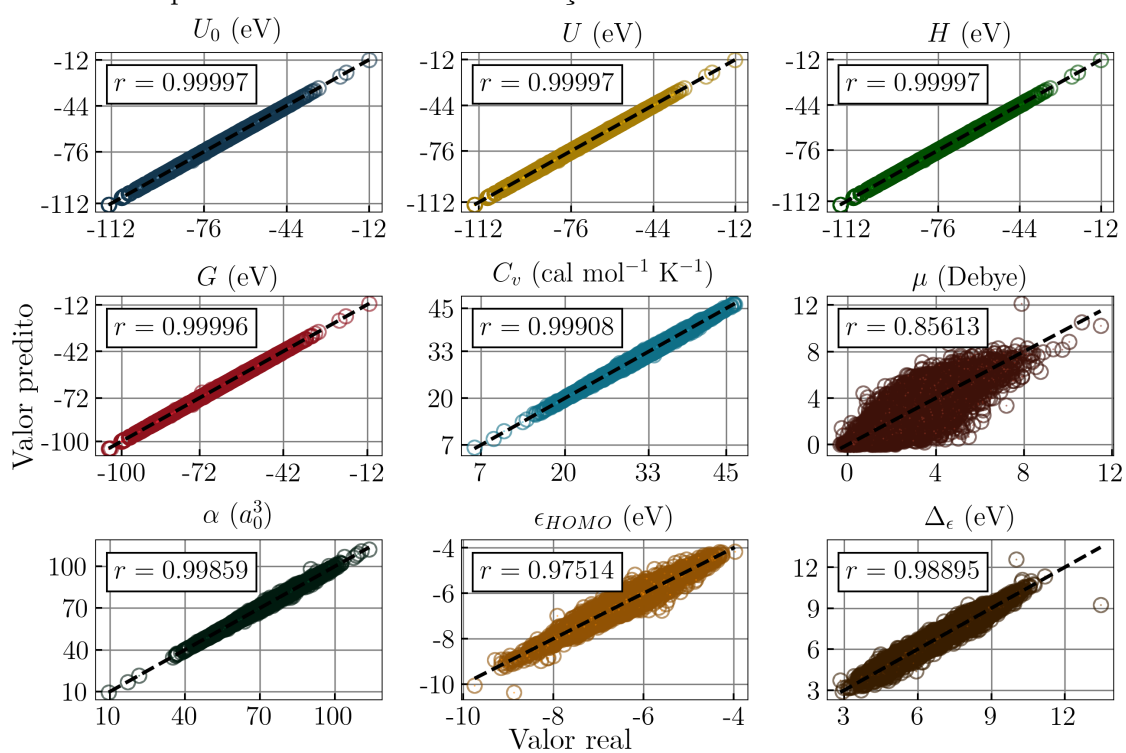
Fonte: Produção do autor.

Tabela 5.12 - Erro de generalização no conjunto QM9 usando como entrada da rede os descritores baseados no SMILES e a matriz de Coulomb.

Propriedades	Descritores SMILES	Matriz de Coulomb
U_0 (eV)	0,0573 ($\pm 0,0004$)	0,4609 ($\pm 0,0058$)
U (eV)	0,0582 ($\pm 0,001$)	0,4662 ($\pm 0,0063$)
H (eV)	0,0575 ($\pm 0,001$)	0,4609 ($\pm 0,0043$)
G (eV)	0,0562 ($\pm 0,001$)	0,4684 ($\pm 0,0061$)
C_v (cal mol K ⁻¹)	0,1223 ($\pm 0,002$)	0,3828 ($\pm 0,0061$)
μ (Debye)	0,523 ($\pm 0,006$)	0,6699 ($\pm 0,0078$)
α (a ₀ ³)	0,3095 ($\pm 0,002$)	0,9075 ($\pm 0,0103$)
ϵ_{HOMO} (eV)	0,0952 ($\pm 0,001$)	0,2284 ($\pm 0,0025$)
Δ_ϵ (eV)	0,1369 ($\pm 0,001$)	0,4381 ($\pm 0,0048$)

Fonte: Produção do autor.

Figura 5.3 - *Scatter plot* do valor real contra o valor predito para o conjunto de teste, no qual r é o coeficiente da correlação de Pearson entre ambas variáveis.



Fonte: Produção do autor.

6 CONCLUSÃO

A aplicação de algoritmos de AM tem-se mostrado como método alternativo para o cálculo de propriedades moleculares, triagem virtual de moléculas e predição da estrutura da molécula, por exemplo. Nesta direção, este trabalho visou realizar uma investigação sobre a predição de propriedades moleculares usando técnicas de AM.

Neste contexto, dois pontos foram atacados: (i) o uso de descritores que sejam derivados de uma representação computacionalmente barata; e (ii) a aplicação de técnicas que permitam realizar um estudo sobre a relação físico-química entre o composto e a propriedade de interesse. No primeiro caso, foi utilizada a representação SMILES, já que a codificação da molécula em SMILES não demanda processos que envolvam a otimização da molécula. Além do mais, SMILES é uma representação popular, contida em diversos programas e base de dados da quimioinformática. No segundo caso, foram implementadas técnicas de SA para a descoberta de descritores relevantes na predição das propriedades moleculares, já que as redes neurais convolucionais baseadas em grafo, comumente utilizadas em tarefas de predição, não têm tal capacidade de interpretação.

Para tanto, utilizou-se a rede Perceptron de múltiplas camadas na predição de nove propriedades moleculares do conjunto QM9. O QM9 é uma base de dados pública que oferece o cálculo de 15 propriedades moleculares em estado de menor energia para aproximadamente 133 mil moléculas. Além disso, foram usados cinco métodos de SA com o objetivo de investigar quais são os descritores frequentemente selecionados e realizar uma comparação entre diferentes métodos. Neste estudo, também foi considerado o uso de descritores derivados baseado no SMILES e coordenadas XYZ para fins comparativos, além de conjuntos com moléculas homogêneas e heterogêneas.

Em suma, as propriedades preditas usando os descritores baseados no SMILES apresentaram resultados melhores sobre o conjunto de atributos completo. No caso das propriedades termodinâmicas, o erro obtido foi próximo à acurácia química, que é igual a 0.043 eV. Para as propriedades C_v e ϵ_{HOMO} a predição obteve um erro do conjunto de teste duas vezes o valor da acurácia química, enquanto Δ_ϵ e α tiveram um valor três vezes maior e μ cinco vezes maior.

Foi observado que o resultado obtido com o SMILES gerou um erro menor para todas as propriedades preditas quando comparado com o descritor da matriz de Coulomb. Isso é interessante, já que se pode chegar no valor da propriedade para o estado de menor energia da molécula sem precisar fazer a otimização sobre a

mesma. Por meio da SA foi possível notar quais são os principais descritores para uma determinada propriedade assumindo um erro de predição um pouco maior. Em geral, os descritores selecionados apresentam a combinação de duas propriedades atômicas e resumizam como a propriedade está distribuída na estrutura da molécula.

Os erros obtidos mostraram que a rede neural consegue aprender melhor com uma maior variedade e quantidade de dados, visto que os menores erros foram atingidos usando o conjunto QM9 inteiro. Uma limitação deste trabalho é em relação às acurácias obtidas nos resultados. Quando comparados com os resultados apresentados na literatura sobre uma diferente divisão do conjunto de treino, teste e validação, as predições apresentam erros um pouco superiores.

Como trabalho futuro, visa-se propor o desenvolvimento de descritores e modelos de AM que possibilitem tanto o entendimento das relações entre entrada e propriedade, quanto acurácias que atinjam o estado da arte. Pretende-se também realizar estudos relacionados diretamente com a aplicação de AM para a predição de materiais no desenvolvimento de supercapacitores usando a representação SMILES.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGRAWAL, A.; CHOUDHARY, A. Perspective: materials informatics and big data: realization of the “fourth paradigm” of science in materials science. **APL Materials**, v. 4, n. 5, p. 053208, 2016. [1](#), [60](#), [61](#)
- AL-JARRAH, O. Y.; YOO, P. D.; MUHAIDAT, S.; KARAGIANNIDIS, G. K.; TAHA, K. Efficient machine learning for big data: a review. **Big Data Research**, v. 2, n. 3, p. 87–93, 2015. [7](#)
- ANG, J. C.; MIRZAL, A.; HARON, H.; HAMED, H. N. A. Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, v. 13, n. 5, p. 971–989, 2015. [48](#), [49](#)
- ASPURU-GUZI, A.; PERSSON, K. **Materials acceleration platform: accelerating advanced energy materials discovery by integrating high-throughput methods and artificial intelligence**. [S.l.]: Canadian Institute for Advanced Research, 2018. [7](#), [59](#)
- BARKER, J.; BULIN, J.; HAMAEEKERS, J.; MATHIAS, S. Localized coulomb descriptors for the gaussian approximation potential. **arXiv preprint arXiv:1611.05126**, 2016. [60](#)
- BATTAGLIA, P.; PASCANU, R.; LAI, M.; REZENDE, D. J.; KAVUKCUOGLU, K. Interaction networks for learning about objects, relations and physics. In: LEE, D. D.; SUGIYAMA, M.; LUXBURG, U. V.; GUYON, I.; GARNETT, R. (Ed.). **Advances in neural information processing systems 29**. [S.l.]: Curran Associates, 2016. p. 4502–4510. [69](#)
- BEHLER, J.; PARRINELLO, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. **Physical Review Letters**, v. 98, n. 14, p. 146401, 2007. [2](#)
- BHARTI, K. K.; SINGH, P. K. A three-stage unsupervised dimension reduction method for text clustering. **Journal of Computational Science**, v. 5, n. 2, p. 156–169, 2014. [50](#)
- BIESIADA, J.; DUCH, W. Feature selection for high-dimensional data—a pearson redundancy based filter. In: KURZYNSKI M.; PUCHALA E.; WOZNIAK M.;

- ZOLNIEREK A. (ED.). **Computer recognition systems 2**. [S.l.]: Springer, 2007. p. 242–249. [53](#)
- BISBO, M. K.; HAMMER, B. Efficient global structure optimization with a machine learned surrogate model. **arXiv preprint arXiv:1907.05741**, 2019. [1](#)
- BISHOP, C. M. **Pattern recognition and machine learning**. [S.l.]: springer, 2006. [7](#), [9](#)
- BISHOP, C. M. et al. **Neural networks for pattern recognition**. [S.l.]: Oxford university press, 1995. [31](#)
- BJERRUM, E.; SATTAROV, B. Improving chemical autoencoder latent space and molecular de novo generation diversity with heteroencoders. **Biomolecules**, v. 8, n. 4, p. 131, 2018. [3](#)
- BJERRUM, E. J. Smiles enumeration as data augmentation for neural network modeling of molecules. **arXiv preprint arXiv:1703.07076**, 2017. [3](#)
- BREIMAN, L. Bagging predictors. **Machine Learning**, v. 24, n. 2, p. 123–140, 1996. [18](#)
- _____. Out-of-bag estimation. Citeseer, 1996. [19](#)
- _____. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. [19](#), [56](#)
- BROTO, P.; MOREAU, G.; VANDYCKE, C. Molecular structures: perception, autocorrelation descriptor and sar studies: system of atomic contributions for the calculation of the n-octanol/water partition coefficients. **European Journal of Medicinal Chemistry**, v. 19, n. 1, p. 71–78, 1984. [66](#)
- BURKE, K. Perspective on density functional theory. **The Journal of Chemical Physics**, v. 136, n. 15, p. 150901, 2012. [60](#)
- BUTLER, K. T.; DAVIES, D. W.; CARTWRIGHT, H.; ISAYEV, O.; WALSH, A. Machine learning for molecular and materials science. **Nature**, v. 559, n. 7715, p. 547–555, 2018. [2](#), [68](#), [74](#)
- CAI, J.; LUO, J.; WANG, S.; YANG, S. Feature selection in machine learning: a new perspective. **Neurocomputing**, v. 300, p. 70–79, 2018. [39](#), [50](#)
- CHANDRASHEKAR, G.; SAHIN, F. A survey on feature selection methods. **Computers & Electrical Engineering**, v. 40, n. 1, p. 16–28, 2014. [40](#)

- CHAPELLE, O.; SCHOLKOPF, B.; ZIEN, A. Semi-supervised learning. **IEEE Transactions on Neural Networks**, v. 20, n. 3, p. 542–542, 2009. [51](#), [52](#)
- CHEN, C.; YE, W.; ZUO, Y.; ZHENG, C.; ONG, S. P. Graph networks as a universal machine learning framework for molecules and crystals. **Chemistry of Materials**, v. 31, n. 9, p. 3564–3572, 2019. [70](#)
- CHEN, Y.; MILLAR, J. A. Machine learning techniques in cancer prognostic modeling and performance assessment. In: MATSUI S.; CROWLEY J. (EDS). **Frontiers of biostatistical methods and applications in clinical oncology**. [S.l.], 2017. p. 193–230. [18](#)
- DASH, M.; LIU, H. Feature selection for classification. **Intelligent data analysis**, v. 1, n. 3, p. 131–156, 1997. [45](#)
- DEVILLERS, J.; BALABAN, A. T. **Topological indices and related descriptors in QSAR and QSPAR**. [S.l.]: CRC Press, 2000. [108](#), [109](#)
- DIETTERICH, T. G. Ensemble methods in machine learning. In: INTERNATIONAL WORKSHOP ON MULTIPLE CLASSIFIER SYSTEMS, 2000, Cagliari, Italy. **Proceedings...** Springer, 2000. p. 1–15. [19](#)
- DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern classification**. [S.l.]: John Wiley & Sons, 2012. [9](#), [34](#)
- DUVENAUD, D. K.; MACLAURIN, D.; IPARRAGUIRRE, J.; BOMBARELL, R.; HIRZEL, T.; ASPURU-GUZI, A.; ADAMS, R. P. Convolutional networks on graphs for learning molecular fingerprints. In: CORTES, C.; LAWRENCE, N. D.; LEE, D. D.; SUGIYAMA, M.; GARNETT, R. (Ed.). **Advances in neural information processing systems**. [S.l.: s.n.], 2015. p. 2224–2232. [2](#), [68](#)
- DY, J. G.; BRODLEY, C. E. Feature selection for unsupervised learning. **Journal of Machine Learning Research**, v. 5, p. 845–889, 2004. [50](#), [51](#)
- EISBERG, R.; RESNICK, R. **Física quântica: átomos, moléculas, sólidos, núcleos e partículas**. [S.l.]: Elsevier, 1979. [59](#)
- EKINS, S.; PUHL, A. C.; ZORN, K. M.; LANE, T. R.; RUSSO, D. P.; KLEIN, J. J.; HICKEY, A. J.; CLARK, A. M. Exploiting machine learning for end-to-end drug discovery and development. **Nature Materials**, v. 18, n. 5, p. 435, 2019. [1](#)
- ESTEVA, A.; ROBICQUET, A.; RAMSUNDAR, B.; KULESHOV, V.; DEPRISTO, M.; CHOU, K.; CUI, C.; CORRADO, G.; THRUN, S.; DEAN, J. A

guide to deep learning in healthcare. **Nature Medicine**, v. 25, n. 1, p. 24–29, 2019. [1](#)

FABER, F.; LINDMAA, A.; LILIENFELD, O. A. von; ARMIENTO, R. Crystal structure representations for machine learning models of formation energies. **International Journal of Quantum Chemistry**, v. 115, n. 16, p. 1094–1101, 2015. [65](#), [67](#), [68](#)

FABER, F. A.; CHRISTENSEN, A. S.; HUANG, B.; LILIENFELD, O. A. V. Alchemical and structural distribution based representation for universal quantum machine learning. **The Journal of Chemical Physics**, v. 148, n. 24, p. 241717, 2018. [2](#)

FABER, F. A.; HUTCHISON, L.; HUANG, B.; GILMER, J.; SCHOENHOLZ, S. S.; DAHL, G. E.; VINYALS, O.; KEARNES, S.; RILEY, P. F.; LILIENFELD, O. A. V. Prediction errors of molecular machine learning models lower than hybrid dft error. **Journal of Chemical Theory and Computation**, v. 13, n. 11, p. 5255–5264, 2017. [1](#), [69](#), [71](#), [89](#)

FAYYAD, U. M. et al. Knowledge discovery and data mining: Towards a unifying framework. In: KDD CONFERENCE, 1996. **Proceedings...** [S.l.]: AIAA, 1996. p. 82–88. [10](#)

FERREIRA, A. J.; FIGUEIREDO, M. A. Efficient feature selection filters for high-dimensional data. **Pattern Recognition Letters**, v. 33, n. 13, p. 1794–1804, 2012. [50](#)

FONTI, V.; BELITSER, E. **Feature selection using lasso**. Amsterdam: VRIJE Universiteit, 2017. [22](#), [23](#), [56](#)

FREUND, Y. et al. Experiments with a new boosting algorithm. In: CITESEER. [S.l.], 1996. [17](#)

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. **The elements of statistical learning**. [S.l.]: New York: Springer, 2001. [20](#), [21](#)

GENUER, R.; POGGI, J.-M.; TULEAU-MALOT, C. Variable selection using random forests. **Pattern Recognition Letters**, v. 31, n. 14, p. 2225–2236, 2010. [56](#)

GHIRINGHELLI, L. M.; VYBIRAL, J.; LEVCHENKO, S. V.; DRAXL, C.; SCHEFFLER, M. Big data of materials science: critical role of the descriptor. **Physical Review Letters**, v. 114, n. 10, p. 105503, 2015. [65](#)

- GILMER, J.; SCHOENHOLZ, S. S.; RILEY, P. F.; VINYALS, O.; DAHL, G. E. Neural message passing for quantum chemistry. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 34., 2017. **Proceedings...** [S.l.], 2017. p. 1263–1272. [68](#), [69](#)
- GLEN, R. C.; BENDER, A.; ARNBY, C. H.; CARLSSON, L.; BOYER, S.; SMITH, J. Circular fingerprints: flexible molecular descriptors with applications from physical chemistry to adme. **IDrugs**, v. 9, n. 3, p. 199, 2006. [3](#)
- GLOTZ, X.; BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND STATISTICS, 3., 2010. **Proceedings...** [S.l.], 2010. p. 249–256. [28](#)
- GOH, G. B.; HODAS, N. O.; SIEGEL, C.; VISHNU, A. Smiles2vec: an interpretable general-purpose deep neural network for predicting chemical properties. **arXiv preprint arXiv:1712.02034**, 2017. [3](#)
- GOH, G. B.; SIEGEL, C.; VISHNU, A.; HODAS, N. O.; BAKER, N. Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed qsar/qspr models. **arXiv preprint arXiv:1706.06689**, 2017. [70](#)
- GÓMEZ-BOMBARELLI, R.; WEI, J. N.; DUVENAUD, D.; HERNÁNDEZ-LOBATO, J. M.; SÁNCHEZ-LENGELING, B.; SHEBERLA, D.; AGUILERA-IPARRAGUIRRE, J.; HIRZEL, T. D.; ADAMS, R. P.; ASPURU-GUZI, A. Automatic chemical design using a data-driven continuous representation of molecules. **ACS Central Science**, v. 4, n. 2, p. 268–276, 2018. [3](#), [68](#), [70](#)
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [S.l.]: MIT Press, 2016. [24](#), [25](#), [26](#), [34](#), [35](#), [36](#)
- GRIFFITHS, D. J. **Introduction to quantum mechanics**. [S.l.]: Pearson, 2010. [59](#)
- GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. **Journal of Machine Learning Research**, v. 3, p. 1157–1182, 2003. [46](#)
- _____. An introduction to feature extraction. In: I., G.; M., N.; S., G.; (EDS), Z. L. (Ed.). **Feature extraction**. [S.l.]: Springer, 2006. p. 1–25. [39](#)

GUYON, I.; GUNN, S.; NIKRAVESH, M.; ZADEH, L. A. **Feature extraction: foundations and applications**. [S.l.]: Springer, 2008. 38

HALL, L. H.; KIER, L. B. Electrotological state indices for atom types: a novel combination of electronic, topological, and valence state information. **Journal of Chemical Information and Computer Sciences**, v. 35, n. 6, p. 1039–1045, 1995. 107

HALL, M. A. **Correlation-based feature selection for machine learning**. Thesis (PhD)), pages = 1999. 198,, 1999. 47, 52, 53

HAN, J.; PEI, J.; KAMBER, M. **Data mining: concepts and techniques**. [S.l.]: Elsevier, 2011. 11, 12

HANSEN, K.; BIEGLER, F.; RAMAKRISHNAN, R.; PRONOBIS, W.; LILIENFELD, O. A. V.; MÜLLER, K.-R.; TKATCHENKO, A. Machine learning predictions of molecular properties: accurate many-body potentials and nonlocality in chemical space. **The Journal of Physical Chemistry Letters**, v. 6, n. 12, p. 2326–2331, 2015. 2

HANSEN, K.; MONTAVON, G.; BIEGLER, F.; FAZLI, S.; RUPP, M.; SCHEFFLER, M.; LILIENFELD, O. A. V.; TKATCHENKO, A.; MÜLLER, K.-R. Assessment and validation of machine learning methods for predicting molecular atomization energies. **Journal of Chemical Theory and Computation**, v. 9, n. 8, p. 3404–3419, 2013. 68, 71

HARRINGTON, P. **Machine learning in action**. [S.l.]: Manning, 2012. 11, 12, 13, 22

HAYKIN, S. **Redes Neurais: princípios e prática**. [S.l.]: Bookman, 2007. 24, 25, 26, 27, 29, 30, 32

HEBB, D. O. **The organization of behavior: a neuropsychological theory**. [S.l.]: Psychology Press, 2005. 25

HUANG, B.; LILIENFELD, O. A. V. **Communication: understanding molecular representations in machine learning: the role of uniqueness and target similarity**. [S.l.]: AIP Publishing, 2016. 3

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An introduction to statistical learning**. [S.l.]: Springer, 2013. 21

- JORDAN, M. I.; MITCHELL, T. M. Machine learning: trends, perspectives, and prospects. **Science**, v. 349, n. 6245, p. 255–260, 2015. 1
- JØRGENSEN, P. B.; MESTA, M.; SHIL, S.; LASTRA, J. M. G.; JACOBSEN, K. W.; THYGESEN, K. S.; SCHMIDT, M. N. Machine learning-based screening of complex molecules for polymer solar cells. **The Journal of Chemical Physics**, v. 148, n. 24, p. 241735, 2018. 3
- KANDEL, E.; SCHWARTZ, J.; JESSELL, T.; SIEGELBAUM, S.; HUDSPETH, A. **Princípios de neurociências-5**. [S.l.]: AMGH, 2014. 23
- KEARNES, S.; MCCLOSKEY, K.; BERNDL, M.; PANDE, V.; RILEY, P. Molecular graph convolutions: moving beyond fingerprints. **Journal of Computer-Aided Molecular Design**, v. 30, n. 8, p. 595–608, 2016. 1, 69
- KHALID, S.; KHALIL, T.; NASREEN, S. A survey of feature selection and feature extraction techniques in machine learning. In: SCIENCE AND INFORMATION CONFERENCE, 2014. **Proceedings...** [S.l.]: IEEE, 2014. p. 372–378. 39
- KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE. 1995. **Proceedings...** [S.l.], 1995. 37
- _____. Wrappers for feature subset selection. **Artificial Intelligence**, v. 97, n. 1-2, p. 273–324, 1997. 48
- KOKOSKA, S.; ZWILLINGER, D. **CRC standard probability and statistics tables and formulae**. [S.l.]: CRC Press, 2000. 54
- KOLLER, D.; SAHAMI, M. **Toward optimal feature selection**. [S.l.: s.n.], 1996. 40
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F.; BURGESS, C. J. C.; BOTTOU, L.; WEINBERGER, K. Q. (Ed.). **Advances in neural information processing systems**. [s.n.], 2012. p. 1097–1105. Disponível em: <<http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>>. 1
- KUMAR, V.; MINZ, S. Feature selection: a literature review. **SmartCR**, v. 4, n. 3, p. 211–229, 2014. 40, 41, 44, 45

- LABUTE, P. A widely applicable set of descriptors. **Journal of Molecular Graphics and Modelling**, v. 18, n. 4-5, p. 464–477, 2000. [106](#), [108](#)
- LANGLEY, P. et al. Selection of relevant features in machine learning. In: . [S.l.]: AIAAm, 1994. [42](#), [43](#)
- LEDOUX, V.; DAELE, M. V.; BERGHE, G. V. Matslise: A matlab package for the numerical solution of sturm-liouville and schrödinger equations. **ACM Transactions on Mathematical Software (TOMS)**, v. 31, n. 4, p. 532–554, 2005. [59](#)
- LENT, R. **Cem bilhões de neurônios: conceitos fundamentais de neurociência**. [S.l.: s.n.], 2004. [23](#)
- LI, J.; CHENG, K.; WANG, S.; MORSTATTER, F.; TREVINO, R. P.; TANG, J.; LIU, H. Feature selection: a data perspective. **ACM Computing Surveys (CSUR)**, v. 50, n. 6, p. 1–45, 2017. [39](#)
- LI, Y.; TARLOW, D.; BROCKSCHMIDT, M.; ZEMEL, R. Gated graph sequence neural networks. **arXiv preprint arXiv:1511.05493**, 2015. [69](#)
- LIU, H.; GEGOV, A.; COCEA, M. Ensemble learning approaches. In: LIU, H.; COCEA, M.; GEGOV, A. (ED.). **Rule based systems for big Data**. [S.l.]: Springer, 2016. p. 63–73. [17](#)
- LIU, H.; MOTODA, H. **Feature selection for knowledge discovery and data mining**. [S.l.]: Springer Science & Business Media, 2012. [40](#), [41](#), [42](#), [43](#), [44](#), [45](#), [46](#), [47](#), [48](#)
- LIU, Y.; ZHAO, T.; JU, W.; SHI, S. Materials discovery and design using machine learning. **Journal of Materiomics**, v. 3, n. 3, p. 159–177, 2017. [59](#)
- LO, Y.-C.; RENSI, S. E.; TORNG, W.; ALTMAN, R. B. Machine learning in chemoinformatics and drug discovery. **Drug Discovery Today**, v. 23, n. 8, p. 1538–1546, 2018. [1](#), [2](#)
- LOUPPE, G. Understanding random forests: from theory to practice. **arXiv preprint arXiv:1407.7502**, 2014. [19](#), [56](#)
- MATTA, C.; GILLESPIE, R. Understanding and interpreting molecular electron density distributions. **Journal of Chemical Education**, v. 79, n. 9, p. 1141, 2002. [59](#)

- MAURI, A.; CONSONNI, V.; TODESCHINI, R. Molecular descriptors. **Handbook of Computational Chemistry**, p. 2065–2093, 2017. [62](#), [64](#), [66](#), [67](#)
- MELVILLE, J. L.; BURKE, E. K.; HIRST, J. D. Machine learning in virtual screening. **Combinatorial Chemistry & High Throughput Screening**, v. 12, n. 4, p. 332–343, 2009. [68](#)
- MITCHELL, T. **Machine learning**. [S.l.]: McGraw-Hill, 1997. [7](#), [8](#)
- MITRA, P.; MURTHY, C.; PAL, S. K. Unsupervised feature selection using feature similarity. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 24, n. 3, p. 301–312, 2002. [50](#)
- MOLINA, L. C.; BELANCHE, L.; NEBOT, À. Feature selection algorithms: a survey and experimental evaluation. In: INTERNATIONAL CONFERENCE ON DATA MINING, 2002. **Proceedings...** [S.l.], IEEE. 2002. p. 306–313. [41](#), [45](#), [47](#)
- MONTAVON, G.; HANSEN, K.; FAZLI, S.; RUPP, M.; BIEGLER, F.; ZIEHE, A.; TKATCHENKO, A.; LILIENFELD, A. V.; MÜLLER, K.-R. Learning invariant representations of molecules for atomization energy prediction. In: PEREIRA, F.; BURGESS, C. J. C.; BOTTOU, L.; WEINBERGER, K. Q. (ED.). **Advances in neural information processing systems**. [S.l.], 2012. p. 440–448. [67](#)
- MONTAVON, G.; RUPP, M.; GOBRE, V.; VAZQUEZ-MAYAGOITIA, A.; HANSEN, K.; TKATCHENKO, A.; MÜLLER, K.-R.; LILIENFELD, O. A. V. Machine learning of molecular electronic properties in chemical compound space. **New Journal of Physics**, v. 15, n. 9, p. 095003, 2013. [2](#), [68](#)
- MORAES, M. B. **Estudo da aplicação de nanotubos de carbono para supercapacitores**. 2013. 117 p. Dissertação (Mestrado em Engenharia e Tecnologia Espaciais / Ciência e Tecnologia de Materiais e Sensores) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2013. [1](#)
- MORIWAKI, H.; TIAN, Y.-S.; KAWASHITA, N.; TAKAGI, T. Mordred: a molecular descriptor calculator. **Journal of Cheminformatics**, v. 10, n. 1, p. 4, 2018. [95](#)
- MOTODA, H.; LIU, H. Feature selection, extraction and construction. **Communication of IICM**, v. 5, n. 2, p. 67–72, 2002. [39](#)
- MURPHY, K. P. **Machine learning: a probabilistic perspective**. [S.l.]: MIT Press, 2012. [9](#)

MURTHY, S. K. Automatic construction of decision trees from data: a multi-disciplinary survey. **Data Mining and Knowledge Discovery**, v. 2, n. 4, p. 345–389, 1998. 13

NETTO, P. O. B. **Grafos: teoria, modelos, algoritmos**. [S.l.]: Edgard Blücher, 2003. 64

NETZER, Y.; WANG, T.; COATES, A.; BISSACCO, A.; WU, B.; NG, A. Y. Reading digits in natural images with unsupervised feature learning. In: . [s.n.], 2011. Disponível em: http://ufldl.stanford.edu/housenumbers/nips2011_housenumbers.pdf. 38

NOSENGO, N. Can artificial intelligence create the next wonder material? **Nature News**, v. 533, n. 7601, p. 22, 2016. 60

OORD, A. Van den; DIELEMAN, S.; SCHRAUWEN, B. Deep content-based music recommendation. In: BURGESS, C. J. C.; BOTTOU, L.; WELLM, M.; GHAHRAMANI, Z.; WEINBERGER, K. Q. (ED.). **Advances in neural information processing systems**. [S.l.], 2013. p. 2643–2651. 1

PISE, N. N.; KULKARNI, P. A survey of semi-supervised learning methods. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE AND SECURITY, 2008. **Proceedings...** [S.l.]: IEEE, 2008. v. 2, p. 30–34. 51

POLANSKI, J.; GASTEIGER, J. Computer representation of chemical compounds. **Handbook of Computational Chemistry**, p. 1997–2039, 2017. 63, 65

PRABHU, G. R. D.; URBAN, P. L. The dawn of unmanned analytical laboratories. **TrAC Trends in Analytical Chemistry**, v. 88, p. 41–52, 2017. 59

PRONOBIS, W.; SCHÜTT, K. T.; TKATCHENKO, A.; MÜLLER, K.-R. Capturing intensive and extensive dft/tddft molecular properties with machine learning. **The European Physical Journal B**, v. 91, n. 8, p. 178, 2018. 3

RAJAN, K. Materials informatics. **Materials Today**, v. 8, n. 10, p. 38–45, 2005. 59

RAMACHANDRAN, P.; ZOPH, B.; LE, Q. V. Searching for activation functions. **arXiv preprint arXiv:1710.05941**, 2017. 29

- RAMAKRISHNAN, R.; DRAL, P. O.; RUPP, M.; LILIENFELD, O. A. V. Quantum chemistry structures and properties of 134 kilo molecules. **Scientific Data**, v. 1, p. 140022, 2014. [3](#), [92](#)
- RAMAKRISHNAN, R.; DRAL, P. O.; RUPP, M.; LILIENFELD, O. A. von. Big data meets quantum chemistry approximations: The δ -machine learning approach. **Journal of Chemical Theory and Computation**, v. 11, n. 5, p. 2087–2096, 2015. [1](#)
- RAMPRASAD, R.; BATRA, R.; PILANIA, G.; MANNODI-KANAKKITHODI, A.; KIM, C. Machine learning in materials informatics: recent applications and prospects. **NPJ Computational Materials**, v. 3, n. 1, p. 54, 2017. [61](#)
- ROKACH, L. Ensemble methods for classifiers. In: MAIMON, O.; ROKACH, L. (ED.). **Data mining and knowledge discovery handbook**. [S.l.]: Springer, 2005. p. 957–980. [17](#)
- ROKACH, L.; MAIMON, O. Decision trees. In: MAIMON, O.; ROKACH, L. (ED.). **Data mining and knowledge discovery handbook**. [S.l.]: Springer, 2005. p. 165–192. [13](#)
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 386, 1958. [25](#)
- RUPP, M. Machine learning for quantum mechanics in a nutshell. **International Journal of Quantum Chemistry**, v. 115, n. 16, p. 1058–1073, 2015. [60](#), [61](#)
- RUPP, M.; RAMAKRISHNAN, R.; LILIENFELD, O. A. von. Machine learning for quantum mechanical properties of atoms in molecules. **The Journal of Physical Chemistry Letters**, v. 6, n. 16, p. 3309–3313, 2015. [68](#)
- RUPP, M.; TKATCHENKO, A.; MÜLLER, K.-R.; LILIENFELD, O. A. V. Fast and accurate modeling of molecular atomization energies with machine learning. **Physical Review Letters**, v. 108, n. 5, p. 058301, 2012. [67](#)
- RYCZKO, K.; STRUBBE, D.; TAMBLYN, I. Deep learning and density functional theory. **arXiv preprint arXiv:1811.08928**, 2018. [60](#)
- SANCHEZ-LENGELING, B.; ASPURU-GUZI, A. Inverse molecular design using machine learning: generative models for matter engineering. **Science**, v. 361, n. 6400, p. 360–365, 2018. [1](#), [2](#), [68](#)

- SANCHEZ-LENGELING, B.; OUTEIRAL, C.; GUIMARAES, G. L.; ASPURU-GUZI, A. Optimizing distributions over molecular space: An objective-reinforced generative adversarial network for inverse-design chemistry (organic). **ChemRxiv**, 2017. [62](#)
- SARKAR, D.; BALI, R.; SHARMA, T. Practical machine learning with python. In: . [S.l.]: Springer, 2018. [11](#)
- SCHÜTT, K.; KINDERMANS, P.-J.; FELIX, H. E. S.; CHMIELA, S.; TKATCHENKO, A.; MÜLLER, K.-R. Schnet: a continuous-filter convolutional neural network for modeling quantum interactions. In: GUYON, I.; LUXBURG, U. V.; BENGIO, S.; WALLACH, H.; FERGUS, R.; VISHWANTHAN, S.; GARNETT, R. (ED.). **Advances in neural information processing systems**. [S.l.], 2017. p. 991–1001. [68](#)
- SCHÜTT, K. T.; ARBABZADAH, F.; CHMIELA, S.; MÜLLER, K. R.; TKATCHENKO, A. Quantum-chemical insights from deep tensor neural networks. **Nature Communications**, v. 8, p. 13890, 2017. [2](#), [68](#), [69](#)
- SCHÜTT, K. T.; SAUCEDA, H. E.; KINDERMANS, P.-J.; TKATCHENKO, A.; MÜLLER, K.-R. Schnet—a deep learning architecture for molecules and materials. **The Journal of Chemical Physics**, v. 148, n. 24, p. 241722, 2018. [2](#), [70](#)
- SEGAL, M. R. Regression trees for censored data. **Biometrics**, p. 35–47, 1988. [13](#), [14](#)
- SEGLER, M. H.; KOGEJ, T.; TYRCHAN, C.; WALLER, M. P. Generating focused molecule libraries for drug discovery with recurrent neural networks. **ACS Central Science**, v. 4, n. 1, p. 120–131, 2018. [70](#)
- SEGLER, M. H.; WALLER, M. P. Neural-symbolic machine learning for retrosynthesis and reaction prediction. **Chemistry—A European Journal**, v. 23, n. 25, p. 5966–5971, 2017. [68](#)
- SENI, G.; ELDER, J. **Ensemble methods in data mining: improving accuracy through combining predictions**. [S.l.]: Morgan & Claypool, 2010. [14](#), [15](#), [17](#), [18](#), [19](#)
- SHARMA, V.; GOSWAMI, R.; MADAN, A. Eccentric connectivity index: a novel highly discriminating topological descriptor for structure- property and structure-activity studies. **Journal of Chemical Information and Computer Sciences**, v. 37, n. 2, p. 273–282, 1997. [66](#)

- SHEIKHPOUR, R.; SARRAM, M. A.; GHARAGHANI, S.; CHAHOOKI, M. A. Z. A survey on semi-supervised feature selection methods. **Pattern Recognition**, v. 64, p. 141–158, 2017. [51](#)
- STUKE, A.; TODOROVIĆ, M.; RUPP, M.; KUNKEL, C.; GHOSH, K.; HIMANEN, L.; RINKE, P. Chemical diversity in molecular orbital energy predictions with kernel ridge regression. **The Journal of Chemical Physics**, v. 150, n. 20, p. 204121, 2019. [3](#)
- SWANN, E.; SUN, B.; CLELAND, D.; BARNARD, A. Representing molecular and materials data for unsupervised machine learning. **Molecular Simulation**, v. 44, n. 11, p. 905–920, 2018. [65](#)
- TANG, J.; ALELYANI, S.; LIU, H. Feature selection for classification: a review. **Data classification: Algorithms and Applications**, p. 37, 2014. [48](#), [49](#), [57](#)
- TIBSHIRANI, R. Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 58, n. 1, p. 267–288, 1996. [22](#), [23](#)
- TIPLER, P. A. **Física moderna**. [S.l.]: Reverté, 1994. [59](#)
- TODESCHINI, R.; CONSONNI, V. **Handbook of molecular descriptors**. [S.l.]: John Wiley & Sons, 2008. [108](#), [109](#)
- _____. **Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references**. [S.l.]: John Wiley & Sons, 2009. [65](#)
- TOLEDO, W. D. de. **Compósito nanoestruturado de diÓxido de titânio/Óxido de grafeno reduzido/fibra de carbono aplicado como eletrodo em supercapacitor**. 2013. 117 p. Dissertação (Mestrado em Engenharia e Tecnologia Espaciais / Ciência e Tecnologia de Materiais e Sensores) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2013. [1](#)
- VAMATHEVAN, J. et al. Applications of machine learning in drug discovery and development. **Nature Reviews Drug Discovery**, v. 18, n. 6, p. 463–477, 2019. [1](#)
- WARD, L. et al. Matminer: an open source toolkit for materials data mining. **Computational Materials Science**, v. 152, p. 60–69, 2018. [65](#)

WARD, L.; WOLVERTON, C. Atomistic calculations and materials informatics: a review. **Current Opinion in Solid State and Materials Science**, v. 21, n. 3, p. 167–176, 2017. [61](#), [65](#)

WEININGER, D. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. **Journal of Chemical Information and Computer Sciences**, v. 28, n. 1, p. 31–36, 1988. [63](#)

WIRTH, R.; HIPPEL, J. Crisp-dm: towards a standard process model for data mining. In: . [s.n.], 2020. Disponível em: <<https://pdfs.semanticscholar.org/48b9/293cfd4297f855867ca278f7069abc6a9c24.pdf>>. [74](#)

WU, X.; ZHU, X.; WU, G.-Q.; DING, W. Data mining with big data. **IEEE Transactions on Knowledge and Data Engineering**, v. 26, n. 1, p. 97–107, 2014. [7](#)

WU, Z.; RAMSUNDAR, B.; FEINBERG, E. N.; GOMES, J.; GENIESSE, C.; PAPPU, A. S.; LESWING, K.; PANDE, V. Moleculenet: a benchmark for molecular machine learning. **Chemical Science**, v. 9, n. 2, p. 513–530, 2018. [1](#), [4](#), [70](#)

XUE, L.; BAJORATH, J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. **Combinatorial Chemistry & High Throughput Screening**, v. 3, n. 5, p. 363–372, 2000. [66](#)

YE, W.; CHEN, C.; DWARAKNATH, S.; JAIN, A.; ONG, S. P.; PERSSON, K. A. Harnessing the materials project for machine-learning and accelerated discovery. **MRS Bulletin**, v. 43, n. 9, p. 664–669, 2018. [71](#)

YU, L.; LIU, H. Efficient feature selection via analysis of relevance and redundancy. **Journal of Machine Learning Research**, v. 5, p. 1205–1224, 2004. [40](#)

ZHANG, C.; MA, Y. **Ensemble machine learning: methods and applications**. [S.l.]: Springer, 2012. [16](#)

ZHAO, Z.; MORSTATTER, F.; SHARMA, S.; ALELYANI, S.; ANAND, A.; LIU, H. Advancing feature selection research. **ASU Feature Selection Repository**, p. 1–28, 2010. [52](#)

ZHU, H. Big data and artificial intelligence modeling for drug discovery. **Annual Review of Pharmacology and Toxicology**, v. 60, p. 573–589, 2020. [60](#)

APÊNDICE A - DESCRITORES SELECIONADOS

A.1 Subconjunto com moléculas de 17 átomos

A.1.1 Energia interna de atomização para a temperatura de 0 K

Metódos	Atributos selecionados
Pcorr	ATS0v, SMR, BCUTp-1h, SlogP_VSA2, SLogP, ATS3v, VSA_EState7, SpMAD_Dzp, ETA_dEpsilon_D, C2SP2, ETA_beta, ETA_dEpsilon_B, SMR_VSA7, SMR_VSA6, Vabc, PEOE_VSA1, SlogP_VSA4, GATS1are, VSA_EState5, FilterItLogS, PEOE_VSA7, SdsCH, ZMIC0, VSA_EState3, ATS4p, piPC8, nBondsKD, C3SP2, BCUTi-1h, PEOE_VSA6, AATSC1se, ECIndex, WPol, NsOH, SMR_VSA9, C1SP3, VE2_A, SlogP_VSA1, GATS1c, SlogP_VSA5, StCH, nRot, nBase, SaaCH, nBondsD, VR1_A, RNCG, TIC5, EState_VSA7, n5Ring, GATS2pe, ATS5i, ATSC3c, NssNH, VSA_EState9, GATS2i, SpMAD_A, n4AHRing, SMR_VSA4, SaasC, GATS2Z, ATSC3se, n6aRing, SdO, C3SP3, nAHRing, n5FAHRing, SdssC, BCUTv-1l, SlogP_VSA8, NdsC, EState_VSA1, NssNH, EState_VSA9, NdsN, ATSC2se, NsNH2, n6Ring, BCUTc-1l, PEOE_VSA9, EState_VSA4, MATS1Z, GGI4, VSA_EState6, TIC1, C4SP3, n3HRing, n8FRing, VSA_EState4, ATSC3i, ATSC3dv, nFRing, ATSC0i, ATSC4c, n4ARing, ATS6i, ATS4i, ATS4d, TSRW10, RPCG, Radius, SlogP_VSA10, ATSC1dv, SsssCH, BalabanJ, n7FARing, n9FRing, NssO, C1SP2, NssCH2, ZMIC1, JGI2, JGI4, Xc-6d, VSA_EState8, JGI3, nAcid, ATSC2c, AATSC2d, PEOE_VSA13, PEOE_VSA3, NsCH3, JGI5, SsssC, NaaN, PEOE_VSA12, ATSC4are, ETA_shape_x, SlogP_VSA3, ATSC8se, EState_VSA6, ATSC8c, ATSC3Z, n5HRing, ATSC7pe, AATSC1d, nX, ATS8v, ATS4dv, ATSC8i, SaasN, EState_VSA8, ATSC5are, NdNH, Xch-3d, StN, AETA_beta_ns_d, EState_VSA2, SaaNH, ATS7i, GGI7, nSpiro, PEOE_VSA8, ETA_shape_y, ATSC6dv, ATSC5d, nFaRing, ZMIC5, ATSC6d, ATSC6i, ATSC7c, GGI6, n7aRing
Scorr	ATS0v, BCUTp-1h, SMR, VE1_Dzp, SLogP, SlogP_VSA2, ATS3v, ETA_dEpsilon_D, SpMAD_Dzp, VSA_EState7, piPC6, nH, PEOE_VSA1, VE1_Dt, GATS1are, ETA_beta, C2SP2, SMR_VSA6, Vabc, ETA_dEpsilon_B, SlogP_VSA4, ETA_eta_RL, RNCG, VSA_EState3, FilterItLogS, ATS4p, PEOE_VSA7, SMR_VSA7, VSA_EState5, SlogP_VSA6, GATS1Z, ZMIC0, SlogP_VSA1, NsOH, PEOE_VSA6, C3SP2, piPC8, StsC, C2SP1, C1SP3, SMR_VSA3, nBase, ETA_eta_R, SlogP_VSA5, GATS1c, SdsCH, ATSC3c, nRot, WPol, ECIndex, VR1_A, nBondsD, GATS2i, NssNH, n5ARing, ATS5i, BCUTd-1l, WPath, AATSC2se, GATS2are, SdO, VSA_EState9, ATSC3se, SpAbs_Dzare, EState_VSA7, VE2_D, TIC5, SaaCH, n5FAHRing, n4AHRing, GATS2Z, SMR_VSA4, AATSC2c, SsssN, n6aRing, VE2_Dzare, NdsN, n6Ring, NdsC, NsNH2, RPCG, PEOE_VSA9, nAHRing, C3SP3, EState_VSA1, SdssC, EState_VSA9, SlogP_VSA8, C4SP3, ETA_shape_x, GGI4, ATSC4c, MATS1Z, EState_VSA4, ATSC3i, ATSC3dv, ATSC1dv, n8FRing, AETA_eta_RL, ATS6i, n3HRing, BCUTdv-1h, SsssCH, SlogP_VSA3, TIC2, EState_VSA10, n4ARing, n7FARing, GATS2c, PEOE_VSA3, Radius, ZMIC5, ZMIC1, ATS4d, PEOE_VSA12, PEOE_VSA13, JGI4, ATSC3Z, n9FRing, JGI1, TIC1, SlogP_VSA10, NssO, ATS4i, ETA_dPsi_B, C1SP2, GGI2, EState_VSA2, BCUTc-1l, ATSC6c, nAcid, AATSC2d, ATSC6pe, NssCH2, JGT10, ATSC7v, JGI5, AATSC1d, ATSC4are, Xc-6d, VSA_EState8, JGI3, SssCH2, NdNH, StN, n5AHRing, NsCH3, ATS7i, ATSC7m, ATS4dv, ATSC8c, SaasN, VSA_EState4, PEOE_VSA11, BCUTc-1h, nSpiro, ETA_shape_y, ATS7dv, JGI7, EState_VSA8, SaaNH, AETA_beta_ns_d, ATSC6d, MIC1, SlogP_VSA11, NaaC, ATSC6i, ATSC4d, EState_VSA6, n7Ring, n6FRing, ATS8dv, GGI6, PEOE_VSA2
RF	nC, SlogP, BCUTZ-1l, ATS0dv, SlogP_VSA2, BertzCT, nBondsS, nBondsKD, ATSC0i, VE1_A, SMR_VSA7, VE2_A, ATS3v, ATS1dv, VR1_A, BCUTi-1h, ABC, GATS1se, BCUTZ-1h, BCUTd-1l, ATSC1se, ATSC2se, ATS3dv, BCUTdv-1h, PEOE_VSA1, ATSC2c, RNCG, n5Ring, n4HRing, BCUTc-1h, SpAbs_Dzv, VSA_EState7, SpMAD_A, NdO, nO, nBondsD, nN, SpAbs_DzZ, RPCG, ETA_dEpsilon_B, C2SP2, nBase, BCUTc-1l, ECIndex, VSA_EState3, Xc-5d, SsssC, SlogP_VSA4, nBridgehead, n5FRing, ATS5d, BalabanJ, ATSC1v, ATSC1Z, ATSC2d, ZMIC2, VSA_EState6, VSA_EState4, SdssC, EState_VSA9, IC1, ATS4d, GATS2c, ATSC2dv, SlogP_VSA3, n5HRing, GGI3, Xch-3d, EState_VSA8, nRot, SlogP_VSA5, BCUTdv-1l, nSpiro, NssO, FilterItLogS, SsssCH, SssCH2, PEOE_VSA6, SlogP_VSA1, n6Ring, ZMIC1, ATSC2Z, SMR_VSA5, nHRing, ATSC3se, NdsC, NdsCH, GGI4, ATS4dv, piPC8, ATSC4d, ETA_shape_y, ATSC3dv, C1SP2, PEOE_VSA11, ATSC3c, PEOE_VSA4, PEOE_VSA3, ATSC2v, ATSC1dv, nAromAtom, PEOE_VSA9, ATSC5d, VSA_EState8, NtN, JGI4, ATS4se, PEOE_VSA12, n6aRing, n7FRing, C2SP1, ETA_shape_p, ATSC6d, PEOE_VSA2, ATSC3Z, NdsN, GGI2, ATSC4dv, ATSC4se, SMR_VSA6, ATSC4c, SMR_VSA4, ATSC3v, PEOE_VSA8, ATS6d, IC2, n6FRing, EState_VSA4, ATSC5c, ATSC5se, ATSC5dv, PEOE_VSA7, SlogP_VSA10, EState_VSA2, NsOH, ATSC6c, n3HRing, EState_VSA3, PEOE_VSA10, ATSC6dv
CFS	Mi, FCSP3, GATS1m, SlogP_VSA2, nH, AATS1are, n4Ring, ATS5d, VE2_Dzi, AATS1pe, ATSC1c, ETA_dEpsilon_D, ATSC1pe, PEOE_VSA1, VE2_Dzare, SpAbs_Dzv, Xc-3dv, AATS0i, NtCH, VE2_D, nHBAcc, AETA_eta_L, AATS1i, ATSC3pe, VE2_Dzpe, SpMAD_Dzp, ATSC1d, ATS0i, ATSC4are, VE2_Dzse, AATS1se, n4HRing, C2SP1, BCUTse-1l, BCUTc-1l, NaaO, BalabanJ, nBondsS, GATS2i, AATS0i, AATSC1pe, AATSC0c, VE2_DzZ, n3HRing, BIC0, BCUTp-1h, ATSC5se, Si, n4ARing, ATSC1v, NsOH, ATSC1se, ATSC6dv, GATS2v, VSA_EState9, C1SP3, GATS2p, AETA_beta_ns_d, AATSC2i, C2SP3, NdsCH, SlogP_VSA4, AMID_h, ETA_shape_x, ATSC3i, VE2_Dt, ETA_dAlpha_B, VSA_EState5, n5Ring, TopoPSA(NO), AATSC1c, ATSC8dv, ATSC7Z, ATS4i, SsssC, SpMAD_Dzi, AATSC2pe, GATS2are, ATSC5d, ATS6i, EState_VSA10, fmf, GATS1Z, NdsN, SpDiam_A, ATSC3d, JGI2, n5aRing, CIC1, BIC1, VE2_Dzm, ETA_epsilon_4, nBondsT, Xp-6d, nRot, VE2_A, SM1_Dzi, BCUTv-1h, n4AHRing, ETA_epsilon_2, VE2_Dzv, n3AHRing, MID_h, AATSC0v, ETA_shape_p, AETA_eta_RL, GGI1, RPCG, SMR_VSA6, GATS2i, MATS2i, ATSC1i, GATS1i, MATS1c, SlogP_VSA1, FilterItLogS, SLogP, NaaN, AETA_eta, SpMAD_D, VE2_Dzp, nHBDon, bpol, VSA_EState3, VR2_A, MATS2v, GATS2m

Metódos	Atributos seleccionados
LASSO	VR1_A, nBridgehead, ATS1dv, ATS2dv, ATS0m, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS5v, ATS6v, ATS8v, ATS2se, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC2dv, ATSC2d, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC0v, ATSC1v, ATSC2v, ATSC4v, ATSC5v, ATSC6v, ATSC7v, ATSC1i, ATSC2i, ATSC3i, BertzCT, SpDiam_Dt, DetourIndex, SaaCH, StN, SssO, ECIndex, ETA_beta, fragCpx, nHBAcc, TIC1, TIC2, TIC4, PEOE_VSA2, PEOE_VSA3, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA4, EState_VSA3, EState_VSA5, EState_VSA8, VSA_EState2, VSA_EState4, VSA_EState6, VSA_EState7, MPC3, MPC5, MPC6, MPC7, MPC8, TopoPSA(NO)

A.1.2 Energia interna de atomização para a temperatura de 298.15 K

Metódos	Atributos seleccionados
Pcorr	ATS0v, SMR, BCUTp-1h, SlogP_VSA2, SLogP, ATS3v, VSA_EState7, SpMAD_Dzp, ETA_dEpsilon_D, C2SP2, ETA_beta, ETA_dEpsilon_B, SMR_VSA7, SMR_VSA6, PEOE_VSA1, Vabc, SlogP_VSA4, GATS1are, VSA_EState5, FilterItLogS, PEOE_VSA7, SdsCH, VSA_EState3, ZMIC0, piPC8, ATS4p, BCUTi-1h, nBondsKD, AATSC1se, C3SP2, PEOE_VSA6, ECIndex, NsOH, WPol, C1SP3, SMR_VSA9, SlogP_VSA1, VE2_A, GATS1c, nRot, SlogP_VSA5, StCH, nBase, SaaCH, nBondsD, RNCG, VR1_A, TIC5, n5Ring, ATS5i, EState_VSA7, VSA_EState9, ATSC3c, NssNH, GATS2pe, GATS2i, SpMAD_A, SMR_VSA4, n4AHRing, SaasC, ATSC3se, GATS2Z, n6aRing, C3SP3, SdssC, SdO, BCUTv-1l, n5FAHRing, EState_VSA1, SlogP_VSA8, NdssC, nAHRing, NssN, EState_VSA9, NdsN, NsNH2, ATSC2se, n6Ring, MATS1Z, EState_VSA4, BCUTc-1l, PEOE_VSA9, VSA_EState6, C4SP3, GGI4, n8FRing, TIC1, nFRing, n3AHRing, VSA_EState4, ATSC3i, ATS6i, ATSC3dv, TSRW10, ATSC0i, ATSC4c, ATS4d, ATS4i, SsssCH, BalabanJ, RPCG, SlogP_VSA10, n4ARing, ATSC1dv, Radius, n7FRing, n9FRing, NssO, C1SP2, ZMIC1, JGI4, JGI2, VSA_EState8, Xc-6d, NssCH2, NsCH3, PEOE_VSA13, nAcid, PEOE_VSA3, ATSC2c, SssssC, JGI3, AATSC2d, PEOE_VSA12, NaaN, JGI5, ATSC4are, ATSC3Z, n5HRing, ATSC8se, ATSC8c, ETA_shape_x, SlogP_VSA3, ATSC7pe, nF, AATSC1d, EState_VSA6, ATSC5are, SaasN, NdNH, ATS8v, EState_VSA8, ATSC8i, AETA_beta_ns_d, ATS4dv, Xch-3d, StN, ATS7i, EState_VSA2, ETA_shape_y, PEOE_VSA8, ATSC5d, SaaNH, ZMIC5, JGI7, nSpiro, ATSC6dv, ATSC6d, n8FaHRing, ATSC6i
Scorr	ATS0v, BCUTp-1h, SMR, VE1_Dzp, SLogP, SlogP_VSA2, ATS3v, ETA_dEpsilon_D, SpMAD_Dzp, VSA_EState7, piPC6, nH, PEOE_VSA1, VE1_Dt, GATS1are, SMR_VSA6, C2SP2, ETA_beta, ETA_dEpsilon_B, Vabc, ETA_eta_RL, SlogP_VSA4, RNCG, VSA_EState3, PEOE_VSA7, FilterItLogS, SMR_VSA7, ATS4p, VSA_EState5, GATS1Z, SlogP_VSA6, ZMIC0, SlogP_VSA1, NsOH, piPC8, PEOE_VSA6, C3SP2, StsC, C1SP3, SMR_VSA3, C2SP1, nBase, SlogP_VSA5, GATS1c, ETA_eta_R, SdsCH, nRot, ATSC3c, WPol, VR1_A, ECIndex, n5ARing, nBondsD, ATS5i, GATS2i, NssNH, AATSC2se, BCUTd-1l, GATS2are, WPath, VSA_EState9, SdO, ATSC3se, EState_VSA7, SpAbs_Dzare, VE2_D, TIC5, n5FAHRing, SaaCH, SMR_VSA4, n4AHRing, AATSC2c, GATS2Z, n6aRing, SssN, NdsN, n6Ring, NsNH2, VE2_Dzare, C3SP3, NdssC, RPCG, EState_VSA1, PEOE_VSA9, SdssC, EState_VSA9, MATS1Z, nAHRing, C4SP3, SlogP_VSA8, EState_VSA4, ATSC4c, ETA_shape_x, ATSC3i, GGI4, ATSC3dv, AETA_eta_RL, n8FRing, ATS6i, ATSC1dv, SsssCH, n3HRing, BCUTdv-1h, n7FARing, TIC2, SlogP_VSA3, EState_VSA10, ZMIC5, PEOE_VSA3, GATS2c, n4ARing, ZMIC1, ATS4d, PEOE_VSA12, JGI4, PEOE_VSA13, ATSC3Z, n9FRing, JGI1, Radius, SlogP_VSA10, ATS4are, EState_VSA2, ETA_dPsi_B, GGI2, TIC1, NssO, BCUTc-1l, C1SP2, ATSC6c, nAcid, ATSC7v, AATSC2d, JGT10, ATSC6pe, NssCH2, SssCH2, ATSC4are, n5AHRing, NsCH3, AATSC1d, JGI5, ATS7i, VSA_EState8, NdNH, Xc-6d, JGI3, StN, ATSC7m, ATSC8c, BCUTc-1h, SaasN, PEOE_VSA11, VSA_EState4, ATS4dv, nSpiro, AETA_beta_ns_d, ATSC6d, EState_VSA8, ATS7dv, JGI7, SaaNH, ETA_shape_y, SlogP_VSA11, ATSC4d, nFaHRing, ATSC6i
RF	nC, SLogP, BCUTZ-1l, ATS0dv, SlogP_VSA2, BertzCT, nBondsS, nBondsKD, ATSC0i, VE1_A, SMR_VSA7, VE2_A, ATS3v, ATS1dv, VR1_A, BCUTi-1h, ABC, GATS1se, BCUTZ-1h, BCUTd-1l, ATSC1se, ATSC2se, ATS3dv, BCUTdv-1h, PEOE_VSA1, ATSC2c, RNCG, n5Ring, n4HRing, BCUTc-1h, SpAbs_Dzv, VSA_EState7, SpMAD_A, NdO, nO, nBondsD, nN, SpAbs_DzZ, RPCG, ETA_dEpsilon_B, C2SP2, nBase, BCUTc-1l, ECIndex, VSA_EState3, Xc-5d, SssssC, SlogP_VSA4, nBridgehead, n5FRing, ATS5d, BalabanJ, ATSC1v, ATSC1Z, ATSC2d, ZMIC2, VSA_EState6, VSA_EState4, SdssC, EState_VSA9, IC1, ATS4d, GATS2c, ATSC2dv, SlogP_VSA3, n5HRing, GGI3, Xch-3d, EState_VSA8, nRot, SlogP_VSA5, BCUTdv-1l, nSpiro, NssO, FilterItLogS, SsssCH, SssCH2, PEOE_VSA6, SlogP_VSA1, n6Ring, ZMIC1, ATSC2Z, SMR_VSA5, nHRing, ATSC3se, NdssC, NdsCH, GGI4, ATS4dv, piPC8, ATSC4d, ETA_shape_y, ATSC3dv, C1SP2, PEOE_VSA11, ATSC3c, PEOE_VSA4, PEOE_VSA3, ATSC2v, ATSC1dv, nAromAtom, PEOE_VSA9, ATSC5d, VSA_EState8, NtN, JGI4, ATS4se, PEOE_VSA12, n6aRing, n7FRing, C2SP1, ETA_shape_p, ATSC6d, PEOE_VSA2, ATSC3Z, NdsN, GGI2, ATSC4dv, ATSC4se, SMR_VSA6, ATSC4c, SMR_VSA4, ATSC3v, PEOE_VSA8, ATS6d, IC2, n6FRing, EState_VSA4, ATSC5c, ATSC5se, ATSC5dv, PEOE_VSA7, SlogP_VSA10, EState_VSA2, NsOH, ATSC6c, n3HRing, EState_VSA3, PEOE_VSA10, ATSC6dv

Metódos	Atributos seleccionados
CFS	Mi, FCSP3, GATS1m, SlogP_VSA2, nH, AATS1are, n4Ring, ATS5p, VE2_D, AATS1pe, AATSC1c, ETA_dEpsilon_D, bpol, ATSC1i, AATS2i, AETA_eta_L, VMcGowan, Xc-3dv, ATSC2d, n3Ring, AATS0i, VE2_Dzare, PEOE_VSA1, n4ARing, mZagreb2, VE2_Dzpe, nHBAcc, VSA_EState7, AATSC0d, ATSC0i, VSA_EState5, GATS1v, AATS2v, BIC0, VE2_DzZ, Mv, SpMAD_Dzp, VE2_Dzse, VR1_Dzp, ETA_dBeta, nBondsS, AATS1i, VE2_Dzm, ETA_dPsi_A, MATS1se, nHBDon, n4HRRing, nRot, AETA_eta, SMR_VSA6, GATS1Z, n3HRRing, ETA_dAlpha_B, AATSC1se, NsOH, AATSC1dv, GATS2m, SlogP_VSA1, FilterItLogS, SlogP_VSA5, AETA_beta_ns_d, ATSC1c, AATS2se, VE2_Dzi, ETA_eta_R, SMR_VSA4, GATS1i, BCUTv-1h, BCUTpe-1l, BCUTp-1h, Si, SMR_VSA5, PEOE_VSA6, AMID_h, VE2_Dzv, n4AHRing, SpMAD_Dzv, ATSC1p, GATS1p, ETA_dEpsilon_C, AATS1se, MATS1c, TopoPSA(NO), VE2_Dt, nSpiro, SM1_Dzi, SlogP_VSA4, VE2_Dzp, AATS2are, BCUTc-1l, SaaO, NdsN, AATSC0p, NdO, GATS1se, SsssN, SlogP_VSA3, VSA_EState3, ATSC1se, SsssNH, SlogP_VSA11, CIC0, BCUTpe-1h, SsNH2, PEOE_VSA3, AATSC0v, ETA_epsilon_4, SM1_Dzp, ETA_epsilon_2, SM1_Dzv, SaaNH, SpMAD_Dzse, VSA_EState9, C1SP3, nFRing, SsOH, SdNH, n7HRRing, CIC5, BIC4, AATSC1are, NsNH2, C2SP3, ETA_shape_p, NsNH, GATS2pe, BCUTse-1l, RPCG, PEOE_VSA5, Xc-5dv, nAHRing, AMID_O, BCUTp-1l, BCUTare-1h, GATS2Z, BCUTi-1h, JGI1, BCUTv-1l, AATSC0Z, VE2_A, NtsC, BCUTare-1l, SdssC, GATS2i, MATS2i, ATS5d, C3SP3, ATSC3se, ATSC4se, ATSC5se, MATS2pe, ATSC6se, ATSC3dv, BCUTdv-1h, ETA_psi_1, AATSC0m, SaasN, ATSC4dv, ATSC7dv, ATSC5dv, AATSC2dv, ATSC4pe, ATSC3are, CIC4, BIC5, VSA_EState8, GATS2p, MATS2p, BCUTdv-1l, NsCH3, NssCH2, ATSC6m, ATS5dv, nBridgehead, SlogP_VSA10, JGI2, ATSC3d, Xpc-5dv, SdsN, SMR_VSA3, nO, AETA_alpha, BalabanJ, GATS2v, SsCH3, fMF, ATSC8Z, EState_VSA9, NsssN, NdsCH, ATSC7m, AATSC1i, PEOE_VSA2
LASSO	VR1_A, nBridgehead, ATS1dv, ATS2dv, ATS0m, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS6v, ATS7v, ATS8v, ATS2se, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC2dv, ATSC2d, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC0v, ATSC1v, ATSC2v, ATSC4v, ATSC5v, ATSC6v, ATSC7v, ATSC1i, ATSC2i, ATSC3i, BertzCT, SpDiam_Dt, DetourIndex, SaaCH, StN, SssO, ECIndex, ETA_beta, fragCpx, nHBAcc, TIC1, TIC2, TIC4, PEOE_VSA2, PEOE_VSA3, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA4, EState_VSA3, EState_VSA5, EState_VSA8, VSA_EState2, VSA_EState4, VSA_EState6, VSA_EState7, MPC3, MPC5, MPC6, MPC7, MPC8, TopoPSA(NO)

A.1.3 Entalpia de atomização para a temperatura de 298.15 K

Metódos	Atributos seleccionados
Pcorr	ATS0v, SMR, BCUTp-1h, SlogP_VSA2, SLogP, ATS3v, VSA_EState7, SpMAD_Dzp, ETA_dEpsilon_D, C2SP2, ETA_beta, ETA_dEpsilon_B, SMR_VSA7, SMR_VSA6, PEOE_VSA1, Vabc, SlogP_VSA4, GATS1are, VSA_EState5, FilterItLogS, PEOE_VSA7, SdsCH, VSA_EState3, ZMIC0, piPC8, ATSC4p, BCUTi-1h, nBondsKD, AATSC1se, C3SP2, PEOE_VSA6, ECIndex, NsOH, WPol, C1SP3, SMR_VSA9, SlogP_VSA1, VE2_A, GATS1c, nRot, SlogP_VSA5, StCH, nBase, SaaCH, nBondsD, RNCG, VR1_A, TIC5, n5Ring, ATS5i, EState_VSA7, VSA_EState9, ATSC3c, NssNH, GATS2pe, GATS2i, SpMAD_A, SMR_VSA4, n4AHRing, SaasC, ATSC3se, GATS2Z, n6aRing, C3SP3, SdssC, SdO, BCUTv-1l, n5FAHRRing, EState_VSA1, SlogP_VSA8, NdssC, nAHRing, NsssN, EState_VSA9, NdsN, NsNH2, ATSC2se, n6Ring, MATS1Z, EState_VSA4, BCUTc-1l, PEOE_VSA9, VSA_EState6, C4SP3, GGI4, n8FRing, TIC1, nFRing, n3AHRing, VSA_EState4, ATSC3i, ATS6i, ATSC3dv, TSRW10, ATSC0i, ATSC4c, ATS4d, ATS4i, SsssCH, BalabanJ, RPCG, SlogP_VSA10, n4ARing, ATSC1dv, Radius, n7FRing, n9FRing, NssO, C1SP2, ZMIC1, JGI4, JGI2, VSA_EState8, Xc-6d, NssCH2, NsCH3, PEOE_VSA13, nAcid, PEOE_VSA3, ATSC2c, SsssC, JGI3, AATSC2d, PEOE_VSA12, NaaN, JGI5, ATSC4are, ATSC3Z, n5HRRing, ATSC8se, ATSC8c, ETA_shape_x, SlogP_VSA3, ATSC7pe, nF, AATSC1d, EState_VSA6, ATSC5are, SaasN, NdNH, ATS8v, EState_VSA8, ATSC8i, AETA_beta_ns_d, ATS4dv, Xch-3d, StN, ATS7i, EState_VSA2, ETA_shape_y, PEOE_VSA8, ATSC5d, SaaNH, ZMIC5, JGI7, nSpiro, ATSC6dv, ATSC6d, n8FAHRRing, ATSC6i
Scorr	ATS0v, BCUTp-1h, SMR, VE1_Dzp, SLogP, SlogP_VSA2, ATS3v, ETA_dEpsilon_D, SpMAD_Dzp, VSA_EState7, piPC6, nH, PEOE_VSA1, VE1_Dt, GATS1are, SMR_VSA6, C2SP2, ETA_beta, ETA_dEpsilon_B, Vabc, ETA_eta_RL, SlogP_VSA4, RNCG, VSA_EState3, PEOE_VSA7, FilterItLogS, SMR_VSA7, ATSC4p, VSA_EState5, GATS1Z, SlogP_VSA6, ZMIC0, SlogP_VSA1, NsOH, piPC8, PEOE_VSA6, C3SP2, StsC, C1SP3, SMR_VSA3, C2SP1, nBase, SlogP_VSA5, GATS1c, ETA_eta_R, SdsCH, nRot, ATSC3c, WPol, VR1_A, ECIndex, n5ARing, nBondsD, ATS5i, GATS2i, NssNH, AATSC2se, BCUTd-1l, GATS2are, WPath, VSA_EState9, SdO, ATSC3se, EState_VSA7, SpAbs_Dzare, VE2_D, TIC5, n5FAHRRing, SaaCH, SMR_VSA4, n4AHRing, AATSC2c, GATS2Z, n6aRing, SsssN, NdsN, n6Ring, NsNH2, VE2_Dzare, C3SP3, NdssC, RPCG, EState_VSA1, PEOE_VSA9, SdssC, EState_VSA9, MATS1Z, nAHRing, C4SP3, SlogP_VSA8, EState_VSA4, ATSC4c, ETA_shape_x, ATSC3i, GGI4, ATSC3dv, AETA_eta_RL, n8FRing, ATS6i, ATSC1dv, SsssCH, n3HRRing, BCUTdv-1h, n7FARing, TIC2, SlogP_VSA3, EState_VSA10, ZMIC5, PEOE_VSA3, GATS2c, n4ARing, ZMIC1, ATS4d, PEOE_VSA12, JGI4, PEOE_VSA13, ATSC3Z, n9FRing, JGI1, Radius, SlogP_VSA10, ATS4are, EState_VSA2, ETA_dPsi_B, GGI2, TIC1, NssO, BCUTc-1l, C1SP2, ATSC6c, nAcid, ATSC7v, AATSC2d, JGT10, ATSC6pe, NssCH2, SssCH2, ATSC4are, n5AHRing, NsCH3, AATSC1d, JGI5, ATS7i, VSA_EState8, NdNH, Xc-6d, JGI3, StN, ATSC7m, ATSC8c, BCUTc-1h, SaasN, PEOE_VSA11, VSA_EState4, ATS4dv, nSpiro, AETA_beta_ns_d, ATSC6d, EState_VSA8, ATS7dv, JGI7, SaaNH, ETA_shape_y, SlogP_VSA11, ATSC4d, nFAHRRing, ATSC6i

Metódos	Atributos selecionados
RF	nC, SLogP, BCUTZ-1l, ATS0dv, SlogP_VSA2, BertzCT, nBondsKD, nBondsS, ATS3v, ATSC0i, VE1_A, SMR_VSA7, VE2_A, ABC, ATS1dv, GATS1se, BCUTi-1h, VR1_A, BCUTZ-1h, BCUTd-1l, ATSC1se, ATSC2se, ATS3dv, ATSC2c, PEOE_VSA1, BCUTdv-1h, SpAbs_Dzv, n5Ring, n4HRRing, RNCG, BCUTc-1h, ETA_dEpsilon_B, nN, SpMAD_A, SpAbs_DzZ, nBondsD, NdO, VSA_EState3, RPCG, C2SP2, Xc-5d, nO, BCUTc-1l, nBase, nBridgehead, VSA_EState7, SssssC, SlogP_VSA4, ATSSd, ATSC2d, BalabanJ, VSA_EState6, SlogP_VSA1, ATSC1Z, ATSC1v, ECIndex, n5FRing, EState_VSA9, VSA_EState4, n5HRRing, ATSC2dv, IC1, ATS4d, GATS2c, SdssC, Xch-3d, ZMIC2, GGI3, nSpiro, SlogP_VSA5, SlogP_VSA3, SMR_VSA6, SsssCH, nRot, PEOE_VSA6, NdsCH, BCUTdv-1l, FilterItLogS, n6Ring, ZMIC1, NdssC, EState_VSA8, SssCH2, ATSC3se, ATS4dv, nHRRing, SMR_VSA5, PEOE_VSA4, ATSC2Z, ETA_shape_y, GGI4, piPC8, ATSC4d, NssO, C1SP2, ATSC3dv, ATSC1dv, PEOE_VSA11, n6aRing, ATSC4se, ATSC5d, PEOE_VSA3, PEOE_VSA12, C2SP1, VSA_EState8, PEOE_VSA9, ATSC2v, ATSC3c, PEOE_VSA2, nAromAtom, JGI4, NtN, SMR_VSA4, n7FRing, GGI2, ATSC3Z, ATSC6d, ATSC6d, ETA_shape_p, ATSC4dv, ATSC4se, ATSC3v, ATSC4c, PEOE_VSA8, NdsN, EState_VSA4, PEOE_VSA7, ATSC5c, n6FRing, ATSC5dv, ATSC5se, IC2, EState_VSA2, ATSC6c, EState_VSA3, PEOE_VSA10, NaaN, n3HRRing, ATSC6dv, n8FRing
CFS	Mi, FCSP3, GATS1m, SlogP_VSA2, nH, AATS1are, n4Ring, ATS5p, VE2_D, AATS1pe, AATSC1c, ETA_dEpsilon_D, bpol, ATSC1i, AATS2i, AETA_eta_L, VMcGowan, Xc-3dv, ATSC2d, n3Ring, AATS0i, VE2_Dzare, PEOE_VSA1, n4ARing, mZagreb2, VE2_Dzpe, nHBAcc, VSA_EState7, AATSC0d, ATS0i, VSA_EState5, GATS1v, AATS2v, BIC0, VE2_DzZ, Mv, SpMAD_Dzp, VE2_Dzse, VR1_Dzp, ETA_dBeta, nBondsS, AATS1i, VE2_Dzm, ETA_dPsi_A, MATS1se, nHBDOn, n4HRRing, nRot, AETA_eta, SMR_VSA6, GATS1Z, n3HRRing, ETA_dAlpha_B, AATSC1se, NsOH, AATSC1dv, GATS2m, SlogP_VSA1, FilterItLogS, SlogP_VSA5, AETA_beta_ns_d, ATSC1c, AATS2se, VE2_Dzi, ETA_eta_R, SMR_VSA4, GATS1i, BCUTv-1h, BCUTpe-1l, BCUTp-1h, Si, SMR_VSA5, PEOE_VSA6, AMID_h, VE2_Dzv, n4AHRing, SpMAD_Dzv, ATSC1p, GATS1p, ETA_dEpsilon_C, AATS1se, MATS1c, TopoPSA(NO), VE2_Dt, nSpiro, SM1_Dzi, SlogP_VSA4, VE2_Dzp, AATS2are, BCUTc-1l, SaaO, NdsN, AATSC0p, NdO, GATS1se, SsssN, SlogP_VSA3, VSA_EState3, ATSC1se, SssNH, SlogP_VSA11, CIC0, BCUTpe-1h, SsNH2, PEOE_VSA3, AATSC0v, ETA_epsilon_4, SM1_Dzp, ETA_epsilon_2, SM1_Dzv, SaaNH, SpMAD_Dzse, VSA_EState9, C1SP3, nFRing, SsOH, SdNH, n7HRRing, CIC5, BIC4, AATSC1are, NsNH2, C2SP3, ETA_shape_p, NsNH, GATS2pe, BCUTse-1l, RPCG, PEOE_VSA5, Xc-5dv, nAHRing, AMID_O, BCUTp-1l, BCUTare-1h, GATS2Z, BCUTi-1h, JGI1, BCUTv-1l, AATSC0Z, VE2_A, NtsC, BCUTare-1l, SdssC, GATS2i, MATS2i, ATSSd, C3SP3, ATSC3se, ATSC4se, ATSC5se, MATS2pe, ATSC6se, ATSC3dv, BCUTdv-1h, ETA_psi_1, AATSC0m, SaasN, ATSC4dv, ATSC7dv, ATSC5dv, AATSC2dv, ATSC4pe, ATSC3dv, BIC5, VSA_EState8, GATS2p, MATS2p, BCUTdv-1l, NsCH3, NssCH2, ATSC6m, ATSSdv, nBridgehead, SlogP_VSA10, JGI2, ATSC3d, Xpc-5dv, SdsN, SMR_VSA3, nO, AETA_alpha, BalabanJ, GATS2v, SsCH3, fMF, ATSC8Z, EState_VSA9, NssssN, NdsCH, ATSC7m, AATSC1i, PEOE_VSA2
LAGSSO	VR1_A, nBridgehead, ATS1dv, ATS2dv, ATS0m, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS6v, ATS7v, ATS8v, ATS2se, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATSSi, ATSC2dv, ATSC2d, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC0v, ATSC1v, ATSC2v, ATSC4v, ATSC5v, ATSC6v, ATSC7v, ATSC1i, ATSC2i, ATSC3i, BertzCT, SpDiam_Dt, DetourIndex, SaaCH, StN, SssO, ECIndex, ETA_beta, fragCpx, nHBAcc, TIC1, TIC2, TIC4, PEOE_VSA2, PEOE_VSA3, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA4, EState_VSA3, EState_VSA5, EState_VSA8, VSA_EState2, VSA_EState4, VSA_EState6, VSA_EState7, MPC3, MPC5, MPC6, MPC7, MPC8, TopoPSA(NO)

A.1.4 Energia livre de Gibbs de atomização para a temperatura de 298.15 K

Metódos	Atributos selecionados
Pcorr	ATS0v, SMR, SlogP_VSA2, BCUTp-1h, SLogP, VSA_EState7, ATS3v, ETA_dEpsilon_D, SpMAD_Dzp, C2SP2, ETA_dEpsilon_B, ETA_beta, Vabc, SMR_VSA6, PEOE_VSA1, SMR_VSA7, SlogP_VSA4, GATS1are, VSA_EState5, FilterItLogS, PEOE_VSA7, ATS4p, SdsCH, VSA_EState3, nBondsKD, ZMIC0, PEOE_VSA6, BCUTi-1h, C3SP2, AATSC1se, piPC8, SMR_VSA9, NsOH, ECIndex, C1SP3, WPOL, SlogP_VSA1, VE2_A, SlogP_VSA5, StCH, GATS1c, nBase, SaaCH, nBondsD, nRot, RNCG, VR1_A, EState_VSA7, ATSC3c, NssNH, GATS2pe, TIC5, GATS2i, VSA_EState9, n5Ring, ATS5i, n4AHRing, SaasC, SMR_VSA4, SpMAD_A, ATSC3se, GATS2Z, nAHRing, SdO, n6aRing, BCUTv-1l, n5FAHRRing, NdssC, NssssN, SdssC, SlogP_VSA8, EState_VSA1, C3SP3, ATSC2se, EState_VSA9, NdsN, NsNH2, PEOE_VSA9, BCUTc-1l, n6Ring, EState_VSA4, VSA_EState6, GGI4, C4SP3, n3AHRing, TIC1, VSA_EState4, MATS1Z, n4ARing, ATS4i, n8FRing, Radius, ATSC4c, ATSC3i, RPCG, ATSC3dv, ATSC0i, SlogP_VSA10, ATSC1dv, ATSC6i, NssO, nFRing, ATS4d, TSRW10, C1SP2, SsssCH, Xc-6d, JGI3, n7FRing, BalabanJ, JGI2, n9FRing, AATSC2d, NssCH2, PEOE_VSA3, ATSC2c, nAcid, PEOE_VSA13, JGI5, JGI4, SssssC, NaaN, ZMIC1, VSA_EState8, ATSC4are, ETA_shape_x, PEOE_VSA12, AATSC1d, EState_VSA6, ATSS8v, NsCH3, SlogP_VSA3, ATSC7pe, ATSC8se, ATSC8i, ATSC8c, ATS4dv, nF, EState_VSA8, ATSC3Z, Xch-3d, SaasN, n5HRRing, ATSC5are, NdNH, StN, AETA_beta_ns_d, EState_VSA2, nSpiro, GGI7, SaaNH, ATSC6dv, ZMIC5, GGI6, ATSC6d, ATS7m

Metodos	Atributos seleccionados
Scorr	ATS0v, BCUTp-1h, SMR, SLogP, SLogP_VSA2, VE1_Dzp, ETA_dEpsilon_D, ATS3v, VSA_EState7, SpMAD_Dzp, nH, piPC6, PEOE_VSA1, GATS1are, ETA_beta, Vabc, SMR_VSA6, C2SP2, VE1_Dt, ETA_dEpsilon_B, SlogP_VSA4, ETA_eta_RL, RNCG, VSA_EState3, FilterItLogS, ATS4p, PEOE_VSA7, VSA_EState5, SMR_VSA7, SlogP_VSA6, SlogP_VSA1, ZMIC0, GATS1Z, PEOE_VSA6, StsC, NsOH, C2SP1, C3SP2, SMR_VSA3, nBase, SlogP_VSA5, piPC8, C1SP3, ETA_eta_R, SdsCH, GATS1c, ATSC3c, nBondsD, nRot, GATS2i, ECIndex, BCUTd-1l, WPol, NssNH, WPath, VR1_A, AATSC2se, GATS2are, n5ARing, SdO, ATS5i, SpAbs_Dzare, ATSC3se, VSA_EState9, EState_VSA7, VE2_D, n4HRing, n5FHRing, SaaCH, TIC5, AATSC2c, GATS2Z, SsssN, SMR_VSA4, n6aRing, VE2_Dzare, nAHRing, NdsN, PEOE_VSA9, NdssC, RPCG, NsNH2, n6Ring, C3SP3, EState_VSA1, ETA_shape_x, SdssC, EState_VSA9, SlogP_VSA8, C4SP3, ATSC4c, GGI4, EState_VSA4, ATSC3i, ATSC1dv, MATS1Z, ATSC3dv, n3HRing, AETA_eta_RL, n8FRing, n4Ring, BCUTdv-1h, SlogP_VSA3, EState_VSA10, ATS6i, TIC2, PEOE_VSA3, Radius, GATS2c, SsssCH, ZMIC5, n7FRing, PEOE_VSA12, PEOE_VSA13, ZMIC1, NssO, ATS4i, JGI4, TIC1, ATS4d, SlogP_VSA10, JGI1, C1SP2, ATSC3Z, ETA_dPsi_B, n9FRing, BCUTc-1l, EState_VSA2, ATSC6c, AATSC2d, nAcid, ATSC6pe, AATSC1d, JGI5, JGI2, JGT10, Xc-6d, NssCH2, JGI3, ATSC4are, ATSC7v, StN, ATS4dv, VSA_EState8, NdNH, SssCH2, ATSC7m, ETA_shape_y, n5AHRing, NsCH3, ATSC7dv, nSpiro, VSA_EState4, SaasN, ATSC8c, ATS7i, PEOE_VSA11, BCUTc-1h, JGI7, EState_VSA8, SaaNH, ATSC6d, AETA_beta_ns_d, SlogP_VSA11, NaaaC, GGI6, ATSC6i, EState_VSA6, MIC1, ATSC4d, ATSC8dv, n7Ring, n6FAHRing, PEOE_VSA2
RF	nC, SLogP, BCUTZ-1l, SlogP_VSA2, BertzCT, ATS0dv, nBondsS, nBondsKD, ATS3v, ATSC0i, VE1_A, GATS1se, VE2_A, SMR_VSA7, BCUTd-1l, ATSC1se, BCUTi-1h, ATS1dv, BCUTZ-1h, ABC, VR1_A, ATS3dv, ATSC2se, BCUTdv-1h, RNCG, ATSC2c, n4HRing, SpAbs_Dzv, ETA_dEpsilon_B, PEOE_VSA1, n5Ring, nN, VSA_EState7, BCUTc-1h, SpMAD_A, NdO, SpAbs_DzZ, nBondsD, RPCG, SsssCH, nBase, nO, BCUTc-1l, VSA_EState3, Xc-5d, ECIndex, ATSC1Z, ATSC2d, SsssC, nBridgehead, ATSC1v, SlogP_VSA1, C2SP2, VSA_EState6, BalabanJ, ATS5d, IC1, n5FRing, VSA_EState4, EState_VSA9, SlogP_VSA4, ATS4d, GATS2c, Xch-3d, SdssC, n5HRing, GGI3, SlogP_VSA3, ATSC2dv, ZMIC2, BCUTdv-1l, nSpiro, FilterItLogS, EState_VSA8, PEOE_VSA6, ZMIC1, nRot, SlogP_VSA5, SssCH2, ATSC2Z, PEOE_VSA11, NssO, NdsCH, ATSC3se, SMR_VSA5, nHRing, ATS4dv, ETA_shape_y, NdssC, GGI4, piPC8, PEOE_VSA4, ATSC3dv, ATSC4d, n6Ring, ATSC1dv, C1SP2, C2SP1, n6aRing, PEOE_VSA12, PEOE_VSA13, ATS4se, JGI4, ATSC5d, ATSC2v, VSA_EState8, ATSC3c, nAromAtom, SMR_VSA4, ETA_shape_p, GGI2, PEOE_VSA9, PEOE_VSA2, n7FRing, ATSC6d, ATSC3Z, ATSC3v, NtN, ATSC4dv, ATSC4se, NdsN, PEOE_VSA7, ATSC4c, ATSC6d, PEOE_VSA8, IC2, ATSC5se, EState_VSA4, ATSC5c, ATSC5dv, NsOH, SMR_VSA6, EState_VSA2, PEOE_VSA10, n6FRing, ATSC6c, NaaN, n3HRing, ATSC6dv, EState_VSA3
CFS	Mi, FCSP3, GATS1m, SlogP_VSA2, nH, AATS1are, n4Ring, AATS2i, ATSC1i, GATS1v, Xch-3d, ETA_dEpsilon_B, BCUTi-1l, BCUTp-1h, AATS0i, n4ARing, VE2_Dzi, nHBAcc, VSA_EState7, ETA_dAlpha_B, AATS0p, ATS0i, AETA_eta_L, GATS1p, GATS1Z, VR1_Dzp, nRot, VE2_Dt, AATS1pe, Mp, VSA_EState8, FilterItLogS, SssNH, CIC5, BIC5, CIC4, BIC4, VSA_EState9, AATSC1i, NssCH2, NsCH3, PEOE_VSA1, ATSC1se, ETA_dEpsilon_D, SsssN, VSA_EState1, VE2_Dzse, TopoPSA(NO), VSA_EState5, BCUTm-1l, MATS1c, nHBDDon, NdsN, n4HRing, VE2_Dzp, SpMAD_Dzp, BCUTc-1l, NaaO, ETA_dPsi_A, AETA_alpha, NtCH, ETA_epsilon_4, VE2_D, AMID_h, Xpc-4dv, ATSC3d, ATSC1d, ATSC2d, ATS3p, SpMAD_Dzare, NaasN, VE2_Dzare, AATS1se, nSpiro, NaaNH, nBondsS, ETA_shape_p, AETA_eta_RL, JGI1, AATS1i, AATSC0p, ATSC3dv, VE2_A, BIC0, VR2_A, SsCH3, SsOH, VE2_Dzv, MID_h, ATSC1c, SsNH2, SpMAD_Dzv, nARing, VE2_DzZ, AATS2are, C2SP1, ATSC5d, NaaN, nBondsD, VSA_EState6, ATSC0p, SM1_Dzare, ETA_dEpsilon_C, ETA_epsilon_2, SM1_Dzp, SM1_Dzpe, VE2_Dzm, LabuteASA, VE2_Dzpe, SM1_Dzi, AATS0v, ATSC7d, AETA_eta, ATSC6p, Diameter, ATSC5p, SaaO, ATSC4i
LASSO	VR1_A, nBridgehead, ATS1dv, ATS2dv, ATS0m, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS5v, ATS6v, ATS8v, ATS2se, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC2dv, ATSC2d, ATSC3d, ATSC4d, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC0v, ATSC1v, ATSC2v, ATSC4v, ATSC5v, ATSC6v, ATSC1i, ATSC2i, ATSC3i, BertzCT, SpDiam_Dt, DetourIndex, SaaCH, StN, SssO, ECIndex, ETA_beta, fragCpx, nHBAcc, TIC1, TIC2, TIC4, PEOE_VSA2, PEOE_VSA3, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA4, EState_VSA3, EState_VSA5, EState_VSA8, EState_VSA9, VSA_EState2, VSA_EState4, VSA_EState6, VSA_EState7, MPC3, MPC5, MPC6, MPC7, MPC8, TopoPSA(NO)

A.1.5 Capacidad térmica para a temperatura de 298.15 K

Metodos	Atributos seleccionados
Pcorr	ATS2i, AETA_eta_R, MATS1m, LogEE_D, ATS4Z, SpMAD_A, SsssCH, ATS5pe, VE2_A, AETA_beta_s, BalabanJ, SssCH2, nHRing, nFARing, SMR_VSA9, ETA_dEpsilon_A, SpMax_Dzv, BIC1, BCUTd-1l, TopoPSA, EState_VSA4, RNCG, MATS1pe, n5ARing, NsssCH, piPC8, ETA_shape_y, EState_VSA1, nRot, C2SP3, GATS2m, BCUTc-1h, NtCH, nBridgehead, SsssC, n5AHRing, ETA_dEpsilon_B, EState_VSA9, n9FARing, bpol, n8FARing, Radius, WPol, TSRW10, VSA_EState3, BCUTse-1l, NsCH3, ATSC2v, NssCH2, NsOH, PEOE_VSA8, ZMIC5, ATSC5v, n7FARing, n4HRing, SMR_VSA5, VE1_A, GATS1Z, EState_VSA5, NsNH2, ATSC2dv, ATSC3m, GATS2pe, SssO, SsssN, NdO, ATSC5d, ATSC3v, JGI2, C3SP3, EState_VSA3, GATS1pe, n6ARing, ATSC1v, EState_VSA2, ATSC4d, AATS2dv, ATSC4dv, GGI4, SMR_VSA4, SdssC, PEOE_VSA7, GGI5, EState_VSA6, nSpiro, SLogP, PEOE_VSA1, SMR_VSA2, C1SP2, ETA_beta_ns_d, PEOE_VSA11, ATSC4v, ATSC8i, BIC2, VR2_A, nBondsKD, PEOE_VSA12, GGI6, SlogP_VSA4, ATS4pe, SdsCH, nBondsD, BCUTp-1l, PEOE_VSA9, JGI4, NdssC, PEOE_VSA4, ATSC7p, VSA_EState9, ZMIC1, PEOE_VSA10, PEOE_VSA13, ATSC3dv, ECIndex, ATSC3pe, CIC5, SlogP_VSA1, SlogP_VSA10, SlogP_VSA11, JGI3, ATS4d, n3HRing, ATSC6d, SdNH, VSA_EState4, SMR_VSA7, GGI7, ATSC3c, SMR_VSA3, ATSC8c, n6FAHRing, PEOE_VSA2, PEOE_VSA6, VSA_EState6, n7Ring, SlogP_VSA2, EState_VSA7, ATSC8dv, NaaNH, ATSC5c, SlogP_VSA8, ATSC6se, RPCG, n5FARing, ATSC1i, NaaO, VSA_EState7, nBase, SssNH, NsF, SMR_VSA6, ATSC4c, NdsN, SlogP_VSA5, EState_VSA8, BCUTdv-1h, SlogP_VSA3, SaasN, VSA_EState8, ATSC7c

Metodos	Atributos seleccionados
Scorr	ATS2i, AETA_eta_R, ATSC1m, VE2_D, LogEE_D, ETA_eta_F, ATS4Z, VE2_Dzse, SpMAD_A, ATS4v, SsssCH, ATS5pe, BCUTd-1l, AETA_beta_s, SssCH2, AETA_eta_RL, nHRRing, ETA_dEpsilon_A, MPC8, LogEE_Dzv, EState_VSA4, TopoPSA(NO), SMR_VSA9, BIC1, AATSC1pe, RNCG, EState_VSA1, n5ARing, nRot, C2SP1, MIC1, BCUTc-1h, NsssCH, EState_VSA9, JGI1, C2SP3, GATS2m, SsssC, ETA_shape_y, AATSC1c, GATS1m, ZMIC5, nBridgehead, n8FARing, ATS6dv, n5AHRing, bpol, NsCH3, Radius, MIC2, ETA_dEpsilon_B, n9FRRing, ATSC2v, NssCH2, n7FARing, ATS8m, SssO, ATS3dv, VSA_EState3, nH, SMR, NsOH, EState_VSA2, VE1_Dzse, ATSC5v, VR2_A, WPol, NsNH2, ATSC2dv, ATSC7i, GATS2pe, n4AHRing, ATSC3m, ATSC1v, Xp-1d, EState_VSA6, GATS1pe, PEOE_VSA1, ATSC3v, SdssC, PEOE_VSA8, SMR_VSA5, n6ARing, NdO, BCUTse-1l, C3SP3, JGI2, BCUTm-1h, ATS7dv, SsssN, GATS2c, ATSC4dv, ATSC4d, GGI4, nSpiro, EState_VSA10, C1SP2, GGI5, CIC5, EState_VSA5, LabuteASA, EState_VSA3, ATS4pe, ZMIC0, SMR_VSA2, SLogP, ETA_beta_ns_d, GGI6, AATS2dv, VSA_EState7, JGI4, SMR_VSA4, ATSC4v, SlogP_VSA4, VSA_EState9, PEOE_VSA12, ATSC6d, PEOE_VSA13, PEOE_VSA11, BIC2, SlogP_VSA1, PEOE_VSA10, PEOE_VSA9, ATSC3pe, NdssC, ATSC8dv, ATS4d, BCUTi-1h, SlogP_VSA10, SMR_VSA7, ATSC3dv, SlogP_VSA11, PEOE_VSA7, NdNH, n3AHRing, ATSC7dv, VSA_EState6, ZMIC1, n6FHRing, VSA_EState4, JGI3, GGI7, ATSC3c, ECIndex, ATSC8dv, SlogP_VSA6, PEOE_VSA6, n7Ring, EState_VSA7, Xc-6dv, NaasC, SlogP_VSA8, SlogP_VSA2, NaaNH, nBase, n5FARing, BCUTc-1l, ATSC6se, ATSC1i, ATSC5c, VSA_EState5, RPCG, NaaO, MATS2c, piPC1, ATSC6i, SdsCH, ATSC6c, PEOE_VSA2, NdsN, VSA_EState8, SssNH, SsF, ATSC4c, C4SP3, EState_VSA8, n8Ring, n6aRing, C1SP3, ETA_dPsi_B, SaasN, SlogP_VSA5, SlogP_VSA3, SMR_VSA6, FilterItLogS, C2SP2, nAcid, n9Ring, n7aRing, n9aRing, C3SP2, PEOE_VSA3, SddC
RF	ETA_shape_p, ATSC1Z, VE2_A, BalabanJ, SsssCH, ETA_shape_y, BCUTZ-1l, SpAbs_DzZ, C2SP1, BCUTd-1l, ETA_dEpsilon_B, nFRing, GGI4, SpMAD_A, ATS5d, Xch-3d, ABC, SpAbs_Dzv, VR1_A, VE1_A, nBondsS, ATS6d, ATS0dv, ATS3dv, ATS4dv, SssCH2, BertzCT, n7FRRing, GATS1se, ATS4d, BCUTdv-1l, IC1, ECIndex, JGI4, BCUTi-1h, n5Ring, n9FRRing, PEOE_VSA1, ATS1dv, ATSC1se, ATS3v, EState_VSA9, ETA_beta_ns_d, SsssC, GGI5, GGI3, n6FRRing, ATS5dv, NsssCH, ATSC4d, ATSC2Z, piPC8, VSA_EState4, ATSC4dv, BCUTc-1h, VSA_EState3, ATSC1v, ATSC2d, RPCG, BCUTdv-1h, FilterItLogS, ATSC2v, ATS7Z, RNCG, NtCH, ATS4se, SlogP_VSA4, VSA_EState7, BCUTc-1l, ATSC1dv, GGI2, BCUTZ-1h, ZMIC2, ATSC3Z, nBondsT, n8FRRing, nRot, nBridgehead, n5FRRing, SMR_VSA7, SlogP, ATS7d, ZMIC1, ATSC2se, ATSC5d, GATS2c, ATSC6c, VSA_EState8, Xc-5d, EState_VSA2, ATSC6d, nHRRing, ATSC4se, SdssC, ATSC3v, ATSC2c, ATSC3dv, SlogP_VSA2, ATSC4c, EState_VSA8, ATSC5dv, ATSC5c, SMR_VSA5, n4HRRing, ATSC2dv, ATSC3c, n5HRRing, ATSC3se, ATSC5se, NsOH, ATSC7c, EState_VSA4, EState_VSA3, n3HRRing, ATSC6dv, SlogP_VSA3, NdsN, EState_VSA1, VSA_EState6, IC2, PEOE_VSA9, nC, PEOE_VSA8, n6Ring, SMR_VSA4, PEOE_VSA10, SlogP_VSA5, GGI6, ATS8Z, PEOE_VSA3, NsNH2, ATSC0i, nSpiro, PEOE_VSA7, ATSC8c, NssCH2, SMR_VSA6, PEOE_VSA11, EState_VSA5, PEOE_VSA2
CFS	AETA_eta_R, SpMAD_D, ETA_shape_p, SMR_VSA9, nHBDOn, SpMax_Dzi, AATSC1d, AETA_eta_F, ATS5pe, JGI1, C2SP1, nO, VMcGowan, Xch-3dv, ETA_eta_F, AETA_eta, ATS6pe, ATS4pe, ATS7p, ATSC1p, ATS4v, BalabanJ, ATS5v, GGI3, NdsN, AETA_beta_ns_d, SpMAD_Dzi, Xc-4d, AETA_dBeta, FilterItLogS, ETA_dEpsilon_B, BCUTse-1l, NtCH, NsCH3, TopoPSA(NO), AATSC0p, TopoPSA, Vabc, Xch-3d, n5aRing, EState_VSA10, nBondsT, AATS1se, SM1_Dzv, ETA_dPsi_A, ETA_dEpsilon_C, AATSC0dv, GATS1Z, n4Ring, AETA_eta_FL, JGI2, SpMax_D, VE2_A, Xp-1dv, SdssC, ETA_dEpsilon_A, VE2_Dzp, ETA_eta_R, NtsC, BCUTc-1l, ATSC0dv, SM1_Dzp, AATSC0c, ATS5p, NssssC, ETA_dBeta, SpMAD_Dzse, ATS3m, VE2_Dt, Xp-6d, ATS5are, NsNH2, ATSC0p
LAGSSO	VR1_A, ATS1dv, ATS2dv, ATS3dv, ATS2d, ATS3d, ATS4d, ATS0m, ATS1m, ATS3m, ATS5m, ATS6m, ATS7m, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS8v, ATS2se, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC1dv, ATSC3dv, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC1v, ATSC2v, ATSC3v, ATSC5v, ATSC6v, ATSC7v, ATSC8v, SpAbs_Dzp, SpAD_Dzp, BertzCT, SpAbs_Dt, SpDiam_Dt, SpAD_Dt, StCH, SaaCH, SdNH, SaaNH, SaaN, ECIndex, fragCpx, TIC1, TIC2, PEOE_VSA2, PEOE_VSA3, PEOE_VSA5, PEOE_VSA7, PEOE_VSA10, PEOE_VSA11, SMR_VSA3, SMR_VSA4, SMR_VSA5, SMR_VSA7, SMR_VSA9, SlogP_VSA4, SlogP_VSA5, SlogP_VSA6, SlogP_VSA10, EState_VSA2, EState_VSA4, EState_VSA5, EState_VSA6, EState_VSA10, VSA_EState3, VSA_EState4, VSA_EState7, MPC3, MPC7, TopoPSA(NO), Vabc, TSRW10, Zagreb2

A.1.6 Momento dipolar

Metodos	Atributos seleccionados
Pcorr	StN, GATS2se, MIC1, PEOE_VSA4, TopoPSA, nN, TIC0, BCUTi-1h, ATS4Z, BCUTc-1h, SdO, VMcGowan, BCUTd-1l, ETA_beta, SpMax_D, VSA_EState6, ATSC2se, VSA_EState7, ETA_eta_R, VSA_EState5, VE2_A, StCH, ETA_dEpsilon_C, ZMIC1, ATSC6dv, GATS1c, GGI4, PEOE_VSA2, SlogP_VSA1, GATS1se, VSA_EState1, ATS4pe, BCUTv-1h, NssNH, ETA_dPsi_B, SpMAD_A, VE1_Dzp, ATSC5dv, Radius, SsOH, NdssC, RPCG, nBondsD, ATSC2c, ATSC3dv, GATS1i, n9FARing, ETA_dEpsilon_B, nBridgehead, ATSC7dv, ATSC1m, GATS2Z, SsssCH, ATSC1v, ATSC0i, C1SP1, PEOE_VSA6, PEOE_VSA3, WPol, ATS5are, TIC2, n5Ring, NsssCH, PEOE_VSA12, nFRing, PEOE_VSA1, AATSC1dv, BCUTc-1l, ECIndex, ATSC3m, C3SP3, nRot, piPC3, SlogP_VSA11, AMID_O, EState_VSA9, TSRW10, GGI5, VSA_EState3, nHRRing, ATSC3c, EState_VSA7, nBondsKD, piPC8, ATSC8dv, ETA_shape_x, ATSC3v, ATS4d, RNCG, FilterItLogS, EState_VSA5, ATSC7d, BCUTdv-1l, nBase, nSpiro, ATS6dv, SlogP, ZMIC2, JGI4, ATSC4m, BCUTdv-1h, NdsN, VSA_EState4, C1SP2, n5HRRing, ATSC7c, BalabanJ, ETA_shape_y, n4AHRing, SMR_VSA5, NsNH2, SaaO, SlogP_VSA5, GATS1Z, C2SP3, PEOE_VSA10, NsssN, C2SP2, EState_VSA2, ATSC8c, SMR_VSA4, SlogP_VSA2, MATS1are, ATSC6d, PEOE_VSA8, PEOE_VSA13, JGI3, ATSC6c, AETA_beta_ns_d, n8FRRing, ATSC4c, ATSC6i, SlogP_VSA6, ZMIC5, SaasN, ATS3v, SdsCH, EState_VSA6, SaaCH, VSA_EState8, GGI2, ATSC2p, C1SP3, PEOE_VSA9, GGI6, n6AHRing, SlogP_VSA8, SssCH2, ATSC4i, EState_VSA4, ATS7dv, ATSC3se, ATSC5d, SsF, n9Ring, SdssC, n7FARing, n6HRRing, ATSC5c, n7HRRing, ATSC5i, SlogP_VSA4, C4SP3, ATS8dv, SaaC, VSA_EState9, VR1_A, n9aRing, n3AHRing, NaaN, EState_VSA3, PEOE_VSA7, SMR_VSA6, NssCH2, n6FRRing, SaaNH, n7aRing, ATS8m, SddC, EState_VSA8

Metodos	Atributos seleccionados
Scorr	NtN, BCUTi-1h, BCUTd-1l, GATS2se, NdO, TopoPSA(NO), MIC1, AETA_beta, ATS4dv, ATSC2se, BCUTdv-1h, VSA_EState6, ATSC6dv, SpMAD_D, GATS1se, BCUTc-1h, VSA_EState1, StCH, GATS1c, ATS4pe, SlogP_VSA1, GG14, ATSC5dv, PEOE_VSA2, NssNH, ATS4v, ETA_dEpsilon_C, VSA_EState5, AETA_eta_R, VE3_Dzp, NdssC, VE2_A, BCUTv-1h, ZMIC1, ETA_dPsi_B, RPCG, ATSC7dv, SsOH, Radius, VSA_EState7, MIC2, n9FARing, SpMAD_A, ATSC3dv, GATS2Z, GATS1i, ETA_dEpsilon_B, nBridgehead, C1SP1, piPC3, MATS1m, PEOE_VSA3, ATSC1v, GATS2c, BCUTdv-1l, EState_VSA7, ATSC6c, PEOE_VSA12, AATSC1dv, EState_VSA10, BIC1, nBase, ATSC1d, ATSC3m, nRot, TIC2, ATS5are, AETA_eta_RL, FilterItLogS, VE2_Dzare, ATSC8Z, C1SP2, ATSC7v, nH, n5Ring, JGI4, SlogP_VSA11, GG15, WPol, Xc-6d, ATS4d, ECIndex, SRW09, PEOE_VSA6, MPC8, ATSC3v, NsssCH, SlogP_VSA5, SsssCH, VSA_EState3, ATSC7d, LabuteASA, EState_VSA5, ATSC3c, NdsN, ZMIC5, nSpiro, EState_VSA9, PEOE_VSA1, ATSC4m, EState_VSA2, BCUTc-1l, C3SP3, nHRing, MATS1pe, PEOE_VSA8, ATS6dv, SLogP, ATSSd
RF	BCUTi-1h, NtN, BCUTd-1l, NdO, BCUTdv-1h, PEOE_VSA4, BCUTZ-1h, RPCG, PEOE_VSA3, BCUTc-1l, BCUTc-1h, RNCG, nN, ATSC2se, SMR_VSA7, VSA_EState4, ATSC2c, GATS2c, SdssC, ATSC4c, ATSC6dv, VSA_EState6, BCUTZ-1l, ATSC5c, SlogP_VSA2, VSA_EState3, ATSC4se, ATSC3c, ATSC5se, ATSC3v, ATSC3se, SpAbs_Dzv, SLogP, ATSC3dv, ATS4dv, ATSC6c, ATSC1se, ATSC2dv, ATSC4dv, ATSC4se, ATSC3dv, ATSC5dv, ATSC7dv, PEOE_VSA10, BertzCT, ATSC4d, SpAbs_DzZ, ATSC3v, VSA_EState7, ATSC3Z, ATSC2d, ATSC2v, PEOE_VSA2, ATSC5d, ATS4d, FilterItLogS, BCUTdv-1l, PEOE_VSA9, VE1_A, VR1_A, ZMIC2, SpMAD_A, ATSC1dv, SsssCH, SssCH2, PEOE_VSA12, VSA_EState8, ATSC2Z, nBase, PEOE_VSA8, EState_VSA2, NtCH, EState_VSA3, ZMIC1, PEOE_VSA1, PEOE_VSA11, BalaabanJ, ATSC6d, VE2_A, ATS6d, GATS1se, ATS5d, NaaN, GG13, JGI4, ATSC0i, ATS1dv, ATS5dv, SlogP_VSA1, SMR_VSA5, Xc-5d, ATSC7c, IC1, EState_VSA4, ATS0dv, ABC, ATSC1v, GG14, ETA_shape_p, EState_VSA9, EState_VSA8, ECIndex, ATSC1Z, IC2, nBondsD, GG12, ETA_dEpsilon_B, ETA_shape_y, EState_VSA5, SsssC, nBondsS, SlogP_VSA4, EState_VSA6, PEOE_VSA7, n3HRing, Xch-3d, SMR_VSA6, EState_VSA1, C1SP2, SlogP_VSA3, EState_VSA7, piPC8, SMR_VSA4, SlogP_VSA5, nHRing, NdsN, C2SP2, ATS7d, PEOE_VSA13, VSA_EState9, ATS6dv, nBondsKD, ETA_beta_ns_d, GG15, C2SP3, nRot, NdssC, PEOE_VSA6, n4HRing, ATSZ
CFS	StN, VSA_EState2, nN, BCUTc-1h, ATS3p, NtN, RPCG, BCUTdv-1h, SMR_VSA3, ATSC4dv, ATSC3m, ATS4i, Diameter, SlogP_VSA5, ETA_shape_y, MIC1, SMR_VSA2, GATS1c, CIC0, SlogP_VSA3, SlogP_VSA1, ATSC4are, ATSC5dv, ATSC3pe, SsCH3, NssCH2, SMR_VSA7, FCSP3, SdO, VSA_EState3, ETA_dEpsilon_C, ATSC0dv, ATSC2c, GATS2c, AATSC1i, SaasN, ATSC4Z, ATS5i, MATS2se, NsssCH, SMR_VSA6, BCUTv-1h, BCUTi-1l, AETA_dBeta, AETA_eta, SlogP_VSA6, SlogP_VSA2, StsC, PEOE_VSA3, NdO, RNCG, TopoPSA(NO), AATSC1c, AMID_h, SM1_Dzv, PEOE_VSA1, ATSC2dv, CIC1, TIC1, AMID_C, GATS1i, ATSC1i, bpol, nBondsT, PEOE_VSA2, MATS1v, AETA_eta_FL, ATS3i, GATS1v, piPC6, GATS1m, Xch-4d, AATSC0dv, ETA_dEpsilon_B, EState_VSA4, TopoPSA, CIC2, SIC3, CIC4, EState_VSA3, SdsCH, SlogP_VSA4, Mi, SMR_VSA5, EState_VSA6, ATSC3dv, ATSC5are, EState_VSA7, MIC2, EState_VSA5, JGI3, GATS1p, BCUTp-1h, SaaNH, ATS3se, ATS4m, VE2_Dzp, SdssC, NssO, PEOE_VSA8, AETA_eta_R, PEOE_VSA7, SpMAD_Dzv, SssNH, SMR_VSA4, NaaN, EState_VSA2, SssCH2, VSA_EState6, ETA_shape_p, NdsCH, SssO, NssN, SaasC
LASSO	VR1_A, ATS4dv, ATS5dv, ATS6dv, ATS2d, ATS5d, ATS6d, ATS1m, ATS2m, ATS3m, ATS6m, ATS7m, ATS8m, ATS0v, ATS1v, ATS2v, ATS3v, ATS5v, ATS7v, ATS8v, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC0dv, ATSC3d, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC0v, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC6v, ATSC7v, ATSC8v, ATSC1i, ATSC2i, ATSC3i, ATSC4i, ATSC5i, ATSC6i, BCUTv-1l, SpAbs_Dzv, SpDiam_Dzare, BertzCT, SpDiam_Dt, StCH, SdsCH, SaaCH, StN, SdsN, SsOH, SaaO, ECIndex, TIC2, TIC3, TIC5, ZMIC0, ZMIC1, PEOE_VSA2, PEOE_VSA3, PEOE_VSA4, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA12, PEOE_VSA13, SMR_VSA3, SMR_VSA4, SMR_VSA6, SMR_VSA7, SlogP_VSA1, SlogP_VSA3, SlogP_VSA4, SlogP_VSA6, SlogP_VSA11, EState_VSA1, EState_VSA2, EState_VSA4, EState_VSA5, EState_VSA6, EState_VSA7, EState_VSA10, VSA_EState1, VSA_EState3, VSA_EState5, VSA_EState7, MPC7, MPC8, TMPC10, TopoPSA(NO), Vabc, TMWC10, TSRW10, Zagreb2

A.1.7 Polarizabilidade isotrópica

Metodos	Atributos seleccionados
Pcorr	SMR, Sp, piPC1, ETA_dEpsilon_B, BCUTse-1l, ETA_beta, ETA_epsilon_4, StsC, VSA_EState5, ATS4p, SlogP_VSA2, VSA_EState7, WPath, GATS1p, EState_VSA8, ATS3are, ETA_epsilon_5, ATS2are, SMR_VSA7, ECIndex, SlogP_VSA4, PEOE_VSA1, SLogP, FilterItLogS, PEOE_VSA5, VE2_A, ATSC1se, TIC1, PEOE_VSA6, NssO, BCUTZ-1h, GATS2p, ATS4se, C2SP2, ATSC0i, SaaCH, SlogP_VSA3, SpMAD_Dzp, SaasC, NsOH, C3SP2, EState_VSA1, BCUTdv-1h, nBondsKD, VR1_A, EState_VSA7, EState_VSA2, Radius, PEOE_VSA7, GG15, piPC8, SMR_VSA5, SdssC, C2SP3, NsssCH, ATSC4d, VSA_EState6, GG16, EState_VSA10, SMR_VSA6, ATSC8i, ATSC5d, nAHRing, n6aRing, NssCH2, ATSC2d, Xc-3d, GG14, EState_VSA6, AETA_eta_RL, MID_N, SdsCH, AATSC1dv, ATS3Z, Xc-5d, TIC2, C1SP3, SsssC, ATSC3c, VSA_EState3, GG17, ATS4m, ATSC2dv, PEOE_VSA10, PEOE_VSA9, SaaNH, ZMIC1, JGI3, ATSC2se, ETA_shape_y, ATSC3v, SaasN, SlogP_VSA8, SaaN, n5AHRing, C1SP2, ATSC4p, JGI2, BCUTc-1h, nFAHRing, ATSC2c, ATSC7p, nRot, ATSC3m, EState_VSA4, PEOE_VSA8, ATS4d, ETA_beta_ns_d, RNCG, SlogP_VSA5, AATS1dv, VSA_EState4, ZMIC2, EState_VSA3, SMR_VSA2, SsCH3, PEOE_VSA4, nBridgehead, nAHRing, ATSC4se, BCUTd-1l, n6AHRing, n5ARing, GATS1se, nF, GATS2pe, EState_VSA9, ATSC7dv, PEOE_VSA11, ATSC6d, ZMIC0, n5FAHRing, NdsN, SaaO, ATSC4c, C3SP3, RPCG, ATS8dv, n7FHRing, n9Ring, ATSC6Z, SlogP_VSA10, n8FAHRing, C4SP3, SssCH2, ATSC6i, SdNH, ATSC3se, SlogP_VSA1, EState_VSA5, nBase, n3HRing, n6FHRing, nSpiro, SsssCH, nBondsD, NaaaC, n7aRing, VSA_EState8, SlogP_VSA11, PEOE_VSA2, ATSC7c, PEOE_VSA13, ATSC1m, SsNH2, ATSC5se, n7HRing, n9AHRing, MPC8, ATSC6c, VSA_EState9, ATSC3dv, ATSC8c, NdssC, SpMAD_A, SMR_VSA4, BalabanJ, PEOE_VSA3, nAcid, NsssN, n8AHRing, JGI4, PEOE_VSA12, n9FHRing, ATSC5c, SssNH, SddC

Metodos	Atributos seleccionados
Scorr	SMR, piPC1, LabuteASA, ETA_dEpsilon_B, ETA_beta, BCUTse-1l, AMID_O, FCSP3, ETA_epsilon_-4, ATS4p, SlogP_VSA2, StsC, VSA_EState7, GATS1p, C2SP1, EState_VSA8, VSA_EState5, WPath, nH, ATS5p, SlogP, VE2_Dzare, SMR_VSA7, ETA_epsilon_5, SlogP_VSA6, SlogP_VSA4, PEOE_VSA1, ATSC1se, FilterItLogS, ETA_eta_RL, TIC1, ECIndex, NssO, VE3_Dzp, VE2_D, GATS2p, SpMAD_-Dzp, PEOE_VSA6, BCUTm-1h, SaaCH, C2SP2, NsOH, EState_VSA2, GATS1i, BCUTd-1l, EState_VSA1, C3SP2, SlogP_VSA3, SpAbs_Dzare, ATS4se, EState_VSA7, ATSC2se, ATSC4d, GGI5, VSA_EState4, NssCH, ETA_eta_R, VR1_A, SMR_VSA6, EState_VSA10, PEOE_VSA7, piPC8, SpAD_A, Radius, C2SP3, GGI4, ATSC2dv, SMR_VSA5, nAHRing, VE3_A, VSA_EState6, SdssC, n6aRing, ETA_shape_x, ETA_shape_y, ATS8p, AATSC1dv, MIC1, ATSC3c, VE3_Dt, ATSC2d, ZMIC1, RNCG, TIC2, NssCH2, PEOE_VSA9, BIC5, ATS3Z, ATS4d, AETA_eta_RL, C1SP3, PEOE_VSA4, ATS6dv, SdsCH, ATS4m, SaaNH, ATSC6d, EState_VSA6, ATSC7i, ETA_beta_ns_d, GGI6, ZMIC2, Xc-5d, PEOE_VSA10, BCUTc-1h, SaasN, NdO, ATSC4p, JGI3, EState_VSA4, ATSC7dv, JGI1, VSA_EState3, n5AHRing, JGI2, PEOE_VSA8, SssssC, SlogP_VSA8, EState_VSA3, n4ARing, Xc-6d, ATSC3v, C1SP2, nFAHRing, GATS1se, nBridgehead, ATSC7dv, GATS2se, ATSC3m, n4HRing, SsCH3, JGI7, SlogP_VSA5, ATSC4se, ATSC8dv, nRot, C3SP3, n6AHRing, SssCH, PEOE_VSA11, ATSC4c, NdsN, SaaO, SlogP_VSA10, SlogP_VSA1, ETA_dPsi_B, RPCG, PEOE_VSA2, BCUTc-1l, n5ARing, n5FHRing, ATS8dv, EState_VSA5, SlogP_VSA11, GATS2c, ATSC3se, ATSC6se, ATSC6i, C4SP3, nBase, SdNH, NdssC, n8FAHRing, n7FAHRing, SRW09, GATS1Z, NssN, ATSC6c, n6FAHRing, SsNH2, SssNH, nSpiro, ZMIC0, NaaaC
RF	ETA_dEpsilon_B, ATSC0i, BCUTZ-1l, nC, nN, BertzCT, C2SP1, SMR_VSA7, nO, nBondsKD, ATS0dv, VE2_A, SLogP, BCUTi-1h, nBondsS, EState_VSA8, BCUTZ-1h, GATS1se, ATSC1se, ABC, BCUTd-1l, SlogP_VSA2, SpAbs_DzZ, SdssC, ECIndex, EState_VSA9, nBondsD, RNCG, ATS6d, VR1_A, PEOE_VSA1, PEOE_VSA7, SlogP_VSA4, SpAbs_Dzv, VSA_EState7, BCUTdv-1h, ATS7d, ATSC2se, VE1_A, ATSC6d, BCUTc-1h, ATS3dv, ATSC5d, ATS7Z, BCUTdv-1l, ATSC2d, ATS5d, IC1, FilterItLogS, ATSC2dv, ATS1dv, ETA_shape_p, RPCG, SssCH, NdsN, ZMIC2, BCUTc-1l, ZMIC1, ATS3v, ATS4dv, SssssC, VSA_EState4, ETA_shape_y, ATS5dv, nFRing, BalabanJ, ATS4se, ATSC6dv, NssCH, SpMAD_A, VSA_EState6, NdsCH, ATSC4d, nBondsT, ATSC5dv, C2SP2, ATS8d, GGI4, ATSC7dv, GGI3, ATSC1v, ATSC1dv, ATSC1Z, SssCH2, JGI4, SMR_VSA5, EState_VSA2, PEOE_VSA6, EState_VSA7, ATSC2c, NdssC, GGI2, GATS2c, ATS4d, VSA_EState3, ATSC2v, n5Ring, nHRing, ATSC3dv, EState_VSA4, SlogP_VSA3, PEOE_VSA11, ATSC3Z, ATSC3v, ATSC4dv, ATSC5se, VSA_EState8, ATSC6c, ATSC2Z, ATSC4c, ATSC3c, ATSC5c, piPC8, ATSC3se, PEOE_VSA4, EState_VSA3, nSpiro, PEOE_VSA9, SlogP_VSA10, ATSC4se, Xc-5d, Xch-3d, EState_VSA5, SlogP_VSA5, nBase, PEOE_VSA2, ATS8Z, PEOE_VSA8, ATSC7c, SMR_VSA4, SlogP_VSA1, SMR_VSA6, EState_VSA6, IC2, VSA_EState9, C3SP2, NdO, NtCH, PEOE_VSA10, n5FRing
CFS	SMR, Sp, ATS6p, ETA_alpha, ETA_dEpsilon_B, apol, nN, ATS8v, PEOE_VSA7, ATS4v, SdssC, BertzCT, C2SP1, nBondsD, n6aRing, Mp, AATS1i, ATS0p, ATS6v, AATS2p, ATSC2p, ATSC3p, AATSC1p, AATSC2p, ECIndex, SM1_Dzp, Vabc, SMR_VSA7, nSpiro, StsC, BCUTse-1l, ATSC1p, GATS1p, Sv, GATS1m, ATS1dv, n7Ring, VSA_EState5, nBondsKD, ATSC1c, SlogP_VSA1, JGT10, ATS5v, SssCH2, n6AHRing, SsOH, MATS2p, AATS0p, ATSC3v, n5AHRing, PEOE_VSA8, VMcGowan, AETA_eta, ATS7p, NdsCH, SaasC, GGI3, ATSC3dv, VSA_EState4, SlogP_VSA5, PEOE_VSA9, NdNH, NdsN, nC, VSA_EState9, BCUTc-1l, VSA_EState8, WPath, piPC6, GATS1Z, ATS0v, NdO, GATS1se, MATS1pe, PEOE_VSA1, ETA_psi_1, BCUTc-1h, EState_VSA4, JGI1, Radius, ETA_dEpsilon_C, BCUTm-1l, SMR_VSA1, VSA_EState6, BCUTp-1h, SMR_VSA3, NdssC, SlogP_VSA4, SlogP_VSA6, ATS8p, ATSC6d, ATSC4d, ATSC3i, EState_VSA3, ATSC2v, GATS2v, AATSC2v, C2SP3, ATSC3m, AATS2v, ATS5p, ATS4i, ATSC4se, ATS7d, fMF, ETA_dBeta, ETA_eta, GATS1v, ATSC5se, ATSC6pe, SlogP_VSA10, AETA_eta_L, ATS5Z, Xc-5d, ATSC3d, ATSC5d, C1SP2, AETA_alpha, ETA_epsilon_2, NtsC, VE2_A, LabuteASA, ATSC7d, AATSC1c, BIC1, EState_VSA5, RPCG, n9Ring, SlogP_VSA2, SM1_Dzv, TopoPSA(NO), CIC0, BIC0, ATSC0v, AMID_h, AATSC0v, MATS2v, NssNH, AMID_O, BCUTv-1l, BCUTare-1h
LASSO	VR1_A, nAromAtom, nSpiro, nBridgehead, ATS5dv, ATS1d, ATS2d, ATS3d, ATS5d, ATS2Z, ATS4Z, ATS5Z, ATS0m, ATS1m, ATS2m, ATS3m, ATS6m, ATS7m, ATS8m, ATS0v, ATS1v, ATS2v, ATS3v, ATS5v, ATS6v, ATS8v, ATS3se, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, AATS2v, ATSC1dv, ATSC3dv, ATSC1d, ATSC2d, ATSC3d, ATSC6d, ATSC2Z, ATSC3m, ATSC4m, ATSC7m, ATSC8m, ATSC0v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC6v, ATSC7v, ATSC8v, ATSC0i, ATSC1i, ATSC2i, ATSC3i, ATSC5i, ATSC6i, BCUTm-1h, BCUTv-1l, SpDiam_Dzv, SpAD_Dzp, SpDiam_Dzi, SpAD_Dzi, BertzCT, C1SP3, C2SP3, SpDiam_Dt, DetourIndex, SpAbs_D, SpDiam_D, SpAD_D, SssCH2, StCH, SaaCH, SdssC, SsNH2, StN, SdsN, SdO, SssO, SaaO, ECIndex, fragCpx, TIC0, TIC3, TIC5, PEOE_VSA1, PEOE_VSA3, PEOE_VSA4, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, SMR_VSA2, SMR_VSA5, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA3, SlogP_VSA4, SlogP_VSA5, SlogP_VSA6, SlogP_VSA11, EState_VSA1, EState_VSA3, EState_VSA4, EState_VSA5, EState_VSA6, EState_VSA8, EState_VSA9, VSA_EState2, VSA_EState4, VSA_EState5, VSA_EState7, MPC5, MPC8, TMPC10, piPC6, n5ARing, n5AHRing, nRot, TopoPSA(NO), Vabc, TMWC10, SRW09

A.1.8 Energia HOMO

Metodos	Atributos seleccionados
Pcorr	SpMAD_Dzi, NaasC, piPC4, AETA_beta_ns_d, SMR_VSA7, SaaNH, ATS1i, SlogP_VSA11, SlogP_VSA10, ETA_epsilon_5, NtN, C2SP2, fMF, GATS1are, n5HRRing, C1SP1, nBondsKD, n5Ring, ATSD4dv, GATS1v, SpMAD_Dzv, VE1_A, AATS2dv, ATS5dv, ETA_dEpsilon_C, ETA_dEpsilon_B, SMR_VSA3, SMR_VSA1, nHBDdon, C1SP2, PEOE_VSA8, ATSC6dv, SlogP_VSA1, nRot, SsNH2, TIC1, SpMAD_A, NdO, ATSC4d, EState_VSA1, ATSC3dv, SMR_VSA5, SaasN, EState_VSA2, GATS2v, ETA_shape_y, piPC8, NssCH2, ATSC4v, ATSC6d, C3SP2, SaaN, AATSC2d, SlogP_VSA3, SaaO, VSA_EState4, JGI1, NssO, ATSC7dv, SlogP_VSA4, ATS6dv, VE2_A, BalabanJ, PEOE_VSA4, SRW09, nHRRing, BCUTdv-1h, FilterItLogS, VSA_EState9, C2SP3, AATSC2m, EState_VSA6, ATS4are, EState_VSA9, ATS4d, GATS1c, SlogP_VSA8, EState_VSA8, BCUTd-1l, SdssC, n8FRing, ATSC1c, AATSC2dv, Xc-4d, ZMIC5, ATSC1d, SlogP_VSA5, TIC5, SsssCH, SsssC, ATSC7c, BCUTse-1h, AATSC2c, ATSC0i, EState_VSA4, C1SP3, BCUTi-1h, ATSC7d, nFRing, Radius, ATS3se, EState_VSA5, n6aRing, nBondsD, ATST7m, n9FRing, BCUTc-1h, BCUTp-1h, PEOE_VSA1, GGI5, GGI6, JGI2, SLogP, AETA_eta_R, ZMIC0, MATS1m, EState_VSA7, VSA_EState3, NdssC, SlogP_VSA2, ZMIC1, PEOE_VSA10, GGI4, SsNH, ATS3dv, VSA_EState7, RPCG, MATS1se, ATS6i, SsssN, PEOE_VSA11, n3HRRing, ATSC3m, NsssCH, BCUTc-1l, SdsCH, ATSC1dv, PEOE_VSA9, ATSC8are, NtCH, ATSC3v, ATSC6c, C4SP3, n5FARing, ATSC5c, n6Ring, RNCG, NaaC, SMR_VSA6, ATSC5dv, ATSHZ, ATSC8c, nAcid, ATSC4c, ATS5i, Xc-6d, ATSC3se, C3SP3, n7FRing, GATS2se, SssCH2, PEOE_VSA12, ATSC3i, ECIndex, PEOE_VSA2, ATS8dv, GGI7, VSA_EState8, ATSC5i, PEOE_VSA13, SsCH3, AMID_X, SMR_VSA4, PEOE_VSA6, ATS3v, VR2_A, SdsN, nBridgehead, PEOE_VSA3, n9Ring, ATSC5v, nBase, PEOE_VSA7, n7AHRing, JGI3, NdNH, n8HRRing, ATSC5d, n6FHRing, ATSC3c, n4HRRing, n9aHRRing, EState_VSA3
Scorr	SpMAD_Dzi, piPC4, BCUTd-1l, NaasC, ATS1i, SMR_VSA7, SlogP_VSA6, SpMAD_Dt, AETA_beta_ns_d, ETA_epsilon_5, ATSD4dv, NtN, GATS1v, C2SP2, SpMAD_Dzv, n5Ring, SaaNH, SpMAD_D, C1SP1, VE1_A, VE2_Dt, ETA_dEpsilon_C, SpMAD_A, ATS2dv, nO, SlogP_VSA11, VE2_A, MID_N, EState_VSA2, nRot, SMR, EState_VSA1, SlogP_VSA10, VSA_EState4, BCUTm-1h, NdO, ETA_shape_y, PEOE_VSA8, ATSC3dv, ETA_dEpsilon_B, piPC8, SdssC, nBondsKD, nHRRing, EState_VSA9, ATSC6dv, GATS1are, ATS6dv, SlogP_VSA3, NssO, ATSC1d, ZMIC5, ATSC4d, NssCH2, JGI1, C3SP2, SsssCH, ATSD4are, SaasN, FCSP3, TIC1, SlogP_VSA5, nHBDdon, ATSC6d, GATS2v, SlogP_VSA1, BCUTp-1h, VSA_EState5, n8FRing, SaaO, ATSC4v, SlogP_VSA4, AETA_eta_RL, BCUTc-1h, Radius, ATSD4, ATSC7dv, MATS1m, AETA_eta_R, SsssC, SsNH2, SMR_VSA5, EState_VSA10, EState_VSA8, AATSC2d, ZMIC0, ATSD4v, GGI5, GGI6, JGT10, ATSD7dv, TIC5, PEOE_VSA10, C1SP3, n9FRing, VSA_EState9, FilterItLogS, VSA_EState7, ATSC3m, AATSC2m, EState_VSA4, BCUTc-1l, AATSC2dv, ATSC3v, Xp-1d, EState_VSA6, SlogP_VSA8, C2SP3, ZMIC1, GGI4, EState_VSA5, ATSC6c, n6aRing, ATSD3dv, C1SP2, PEOE_VSA11, NdssC, SsssN, SlogP_VSA2, BCUTdv-1l, ATSC7v, SsNH, SLogP, ATSC5dv, JGI2, n7FARing, n6Ring, TIC2, n3HRRing, C2SP1, RNCG, ATSD5are, ATSC5d, GATS1c, VSA_EState3, n5AHRing, RPCG, GATS1Z, ATSC1dv, GATS2are, AATSC2c, VE1_Dzp, VE2_Dzp, EState_VSA7, nAcid, PEOE_VSA9, AATSC1se, ATSD8m, ECIndex, nBridgehead, JGI3, ATSC8are, n5FARing, PEOE_VSA1, C3SP3, nBase, SdsN, C4SP3, NsssCH, VR2_A, ATSC5c, GGI7, n4HRRing, ATSD8dv, NdsCH, ETA_dPsi_B, n8FARing, ATSC3se, ATSC4c, GATS1i, n6FRing, SdNH, EState_VSA3
RF	nAromAtom, ETA_beta_ns_d, BCUTd-1l, GATS1se, BCUTZ-1h, SlogP_VSA10, SMR_VSA7, NtN, BCUTi-1h, nBondsS, PEOE_VSA1, nBase, BertzCT, nO, PEOE_VSA4, ATSC0i, BCUTc-1l, nFRing, BCUTdv-1h, NaaN, SpAbs_DzZ, BCUTc-1h, ATSC1se, SLogP, n3HRRing, ETA_dEpsilon_B, ABC, ATSD0v, nRot, GATS2c, GGI2, RNCG, SlogP_VSA11, RPCG, ATSD4dv, VSA_EState6, ATSC2c, EState_VSA9, ATSC1v, BCUTZ-1l, ATSC2se, SpMAD_A, SdssC, SlogP_VSA3, VE1_A, PEOE_VSA3, BCUTdv-1l, IC1, VSA_EState4, SlogP_VSA2, VSA_EState3, ATSC2dv, SlogP_VSA1, BalabanJ, ATSD3dv, ATSC1Z, ZMIC1, SlogP_VSA5, Xch-3d, ATSC3c, ATSD1dv, SpAbs_Dzv, VSA_EState7, VE2_A, PEOE_VSA2, nBondsKD, VSA_EState8, ATSC4d, PEOE_VSA8, ATSC2d, ATSC4c, nN, VSA_EState9, SMR_VSA5, VR1_A, ATSC3se, ATSD3dv, SsssCH, SssCH2, PEOE_VSA6, ATSC1dv, NaaNH, ATS3v, FilterItLogS, ATSD4se, ATSC4se, ATSC3Z, ATSC5c, EState_VSA8, ZMIC2, ATSC2Z, GGI3, ATSC4dv, ATSC5se, ATSC5dv, ATSC3v, PEOE_VSA11, ATSC6c, ATSC5d, ATSD5d, ATSD4d, n5Ring, C2SP2, nBondsD, ATSC2v, IC2, ATSD5dv, nBridgehead, PEOE_VSA9, PEOE_VSA10, ETA_shape_p, JGI4, ATSC6dv, SlogP_VSA4, EState_VSA3, ATSC6d, nC, GGI4, PEOE_VSA7, EState_VSA4, SMR_VSA6, EState_VSA2, Xc-5d, nHRRing, n5HRRing, NsNH2, ETA_shape_y, C1SP2, ECIndex, EState_VSA1, ATSD6d, EState_VSA6, EState_VSA7, ATSC7c, SsssC, SMR_VSA4, EState_VSA5, NsOH, PEOE_VSA12, piPC8
CFS	NaasC, ATS1i, ETA_dEpsilon_B, AETA_beta_ns_d, AMID, nBondsD, GATS1Z, AETA_beta_s, SaaCH, VSA_EState5, EState_VSA10, VE2_Dt, piPC3, AATS1i, GATS2se, NdssC, MATS1p, n5aRing, VSA_EState9, AMID_C, PEOE_VSA1, SMR, BCUTse-1l, ETA_dEpsilon_C, SlogP_VSA11, BIC2, CIC2, TIC2, ATS2i, ATS5i, BCUTc-1l, C2SP2, SlogP_VSA10, Xch-6d, ATS6p, n6AHRing, n5HRRing, nH, VSA_EState2, MID_C, GATS1are, ATSD4d, ETA_beta_ns_d, GGI2, ATS7p, EState_VSA4, NdO, VE2_A, EState_VSA9, SdssC, nHBDdon, GATS2c, BertzCT, SdO, VE2_D, TpiPC10, SpMAD_Dzv, RNCG, ETA_dEpsilon_D, VSA_EState7, GATS1i, ETA_eta, SaasC, Vabc, n4HRRing, nF, EState_VSA3, GATS1m, SRW09, ZMIC2, BIC1, CIC1, TIC1, VE2_Dzp, piPC2, ATSC3i, MATS2i, ATSD3are, ATS6i, MATS1c, EState_VSA2, EState_VSA1, SM1_Dzp, BCUTm-1l, C2SP1, SMR_VSA7, SMR_VSA4, ETA_dPsi_A, VE2_Dzv, piPC1, SlogP_VSA1, AMID_O, SM1_Dzv, AATS2i, ATS3p, SaaNH, SaaO, FilterItLogS, PEOE_VSA8, MATS2Z, NssCH2, NdsCH, PEOE_VSA9, NsCH3, GATS2p, SMR_VSA9, PEOE_VSA10, PEOE_VSA7, SMR_VSA6, PEOE_VSA2, C2SP3, NaaCH, ATSC0c, ATSC0v, Si, C3SP3, SsCH3, GATS2m, MATS2pe, SlogP_VSA3, VSA_EState3, ATSC1c, SpMAD_Dzp, AATSC0v, NsNH2, SssNH, C4SP3, GATS1v, SpMax_Dt, NsOH, SLogP, SIC1, SsssN, nO, AETA_alpha, Mi, AATSC0p, AMID_h, CIC0, BIC0, ATSC0p, SsNH2, NssNH, NaaNH, VE2_Dzare, ZMIC1, PEOE_VSA11, NaasN, JGI5, AETA_eta, VMcGowan, NdsN, nBondsS, AETA_eta_FL, AETA_eta_L, PEOE_VSA12, nAromBond, nC, ETA_epsilon_2, VE2_Dzpe, LabuteASA, VE2_DzZ, MIC2, SlogP_VSA4, PEOE_VSA13, NsssN, MID_O
LAGSO	ATS2dv, ATS2d, ATSD0m, ATS1m, ATS2m, ATS3m, ATS5m, ATS6m, ATS7m, ATS1v, ATS2v, ATS3v, ATSD4v, ATSD5v, ATSD6v, ATSD8v, ATSD2se, ATSD3se, ATSD4se, ATSD0i, ATS1i, ATS2i, ATS3i, ATSD4i, ATSD5i, ATSD6i, ATSD7i, AATS2v, ATSC3dv, ATSC1d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC0v, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC6v, ATSC7v, ATSC1i, ATSC3i, ATSC5i, SpAD_Dzp, SpAbs_Dzi, SpDiam_Dzi, BertzCT, C1SP3, SpAbs_Dt, SpAD_Dt, SaaNH, fragCpx, TIC1, PEOE_VSA1, PEOE_VSA3, PEOE_VSA6, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, SMR_VSA3, SMR_VSA4, SMR_VSA5, SlogP_VSA2, SlogP_VSA4, SlogP_VSA5, SlogP_VSA10, SlogP_VSA11, EState_VSA3, EState_VSA4, EState_VSA8, EState_VSA10, VSA_EState2, MPC4, MPC5, TMPC10, TopoPSA(NO), Vabc, TSRW10

A.1.9 Diferença de energias do HOMO e LUMO

Metódos	Atributos selecionados
Pcorr	nBondsKD, FCSP3, nBondsD, VSA_EState2, C2SP2, ETA_eta_F, TIC1, ETA_eta_R, SMR_VSA7, NdsCH, C1SP2, NdssC, PEOE_VSA2, WPol, ATS4v, SMR_VSA5, NsssCH, SssO, GATS2Z, Xch-4dv, EState_VSA9, C2SP3, TIC2, nAHRing, PEOE_VSA9, GATS1se, GATS2pe, SsssCH, ATSC2d, ETA_dEpsilon_B, FilterItLogS, n4Ring, ATS5pe, PEOE_VSA11, VE2_A, ATSC4d, EState_VSA6, SssCH2, C3SP2, AATS2dv, ZMIC1, ATSC2v, StN, ATSC1m, C1SP1, ATSC2dv, C3SP3, JGI3, nFAHRing, Radius, bpol, n4HRing, NssCH2, nBridgehead, ATSC4p, AATS2v, n6aRing, ATSC3Z, ATSC5dv, ATSC4dv, EState_VSA7, PEOE_VSA4, SlogP_VSA4, BCUTd-1l, VSA_EState3, ETA_beta_ns_d, GGI4, n3HRing, BCUTc-1h, ATSC1v, SdssC, SlogP_VSA2, RNCG, PEOE_VSA12, SaaN, GATS1Z, ECIndex, PEOE_VSA1, SMR_VSA4, BCUTdv-1l, ATSC3pe, ATSC3v, BalabanJ, ATS4d, nRot, VSA_EState6, SpMAD_A, SlogP_VSA3, SsOH, ATSC5d, PEOE_VSA6, SaaNH, n5AHRing, SMR_VSA3, BCUTdv-1h, nSpiro, TopoPSA(NO), n5FARing, VR2_A, BCUTp-1l, ZMIC2, ATS3v, SaaO, BCUTpe-1l, TIC0, SpMAD_Dzp, SlogP_VSA8, ATSC1se, VSA_EState7, AMID_X, SlogP_VSA1, ATSC7p, VSA_EState9, n6Ring, NdsN, JGI2, n9FHRing, n6FHRing, ATSC6d, NaasN, C1SP3, ATSC3i, VSA_EState4, ATSC6p, PEOE_VSA5, ATS4i, ATSC3dv, EState_VSA5, SdNH, GGI5, SlogP_VSA11, SLogP, SMR_VSA6, SlogP_VSA10, AATSC1i, SsNH2, C4SP3, n7FAHRing, ETA_shape_y, VSA_EState5, ATSC8p, SsssN, ATSC6se, n7Ring, GATS1c, PEOE_VSA8, PEOE_VSA7, RPCG, n5ARing, EState_VSA4, PEOE_VSA13, EState_VSA8, ATSC4c, PEOE_VSA3, ZMIC5, n9Ring, piPC8, ATSC5c, JGI4, VSA_EState8, AATSC2c, PEOE_VSA10, n7aRing, NssNH, ATSC7c, BCUTc-1l, n8Ring, ATSC1c, GGI6
SCorr	nBondsKD, nBondsD, C2SP2, VSA_EState2, SMR, NdsCH, TIC1, SlogP_VSA6, C1SP2, SMR_VSA7, ETA_eta_R, NdssC, MIC1, GATS1se, BCUTdv-1h, NsssCH, Xch-4dv, VE2_Dzm, EState_VSA10, EState_VSA9, nH, GATS2Z, WPol, SssO, ATS4v, ATS4Z, SMR_VSA5, PEOE_VSA2, C2SP3, nAHRing, PEOE_VSA9, ATSC2d, GATS2pe, SlogP_VSA3, MIC2, AETA_beta_s, ATS2i, ETA_dEpsilon_B, FilterItLogS, C3SP2, n4ARing, ETA_shape_x, SsssCH, C1SP1, TIC2, nAromAtom, ATSC4d, StN, piPC1, VE3_A, ATSC3dv, PEOE_VSA11, ATSC2dv, EState_VSA6, C3SP3, ATS5pe, ATSC2v, BCUTc-1h, SssCH2, ATSC1m, n4HRing, nFAHRing, ATS0dv, VE1_Dzp, ETA_beta_ns_d, n6aRing, AATS2dv, SdssC, Xc-6dv, ATSC5se, Radius, ATSC4p, VSA_EState3, SlogP_VSA4, ATSC4dv, JGI3, NssCH2, bpol, BCUTc-1l, VR2_A, SlogP_VSA2, EState_VSA7, ZMIC0, VSA_EState6, ATSC1v, AATS2v, n3HRing, nBridgehead, ATSC3Z, BCUTare-1l, ZMIC1, GGI4, SMR_VSA4, NsCH3, nRot, SsOH, ETA_eta_RL, n6Ring, SpMAD_A, ATSC3pe, SaaNH, ATS4d, SlogP_VSA8, PEOE_VSA6, VSA_EState7, SpMAD_Dzp, ZMIC5, SaaO, n5FRing, RNCG, PEOE_VSA10, PEOE_VSA1, nSpiro, PEOE_VSA12, SlogP, ATSC3v, ATS3v, ATSC1se, ATSC6d, n6FAHRing, ATSC6p, JGI2, SpMax_Dzare, AMID_O, StCH, SlogP_VSA10, SMR_VSA6, PEOE_VSA7, TIC3, VSA_EState4, BCUTd-1l, n5AHRing, ATSC7p, GATS1c, VSA_EState9, NaasN, n9FAHRing, GATS1Z, AETA_eta_RL, MID_N, n5Ring, SdNH, C1SP3, GATS2c, ATS4pe, SlogP_VSA11, NdsN, ECIndex, VSA_EState5, GGI5, ATSC6se, RPCG, n7FHRing, C4SP3, nBase, ATSC3i, ETA_shape_y, ATSC3dv, PEOE_VSA8, n7Ring, SlogP_VSA1, EState_VSA4, ATSC4c, ATSC5c, SsNH2, AATSC2c, SsssN, ATSC1c, ATSC6dv, ATSC7dv, EState_VSA5, EState_VSA2, SlogP_VSA5, n9Ring, ATSC6c, PEOE_VSA13, ATSC8v, JGI4, SsF, ETA_dPsi_B, VSA_EState8, piPC8, SsssC, n8ARing, ATSC3c, n7aRing, EState_VSA3, ATSC8dv, EState_VSA8, NaaaC, GGI6, nAcid, PEOE_VSA3, n8FAHRing, n9aHRing, JGI7, NddC, ATSC8d
RF	nBondsKD, C2SP2, nBondsD, SMR_VSA7, C1SP2, IC1, BCUTdv-1h, BCUTZ-1h, NdsCH, ETA_dEpsilon_B, BCUTc-1l, nN, BCUTi-1h, NdO, BCUTZ-1l, PEOE_VSA12, BCUTd-1l, nBondsS, PEOE_VSA2, ATSC2c, BCUTc-1h, nAromAtom, EState_VSA9, BertzCT, ATSC3se, ATS1dv, ATSC2se, ABC, VSA_EState4, EState_VSA8, SdssC, RNCG, VE2_A, SlogP_VSA10, SLogP, RPCG, PEOE_VSA3, PEOE_VSA11, SpAbs_DzZ, ATSC0i, SlogP_VSA4, ATSC2dv, nBase, nBondsT, GATS2c, NdssC, ATSC3dv, ZMIC1, GGI2, PEOE_VSA7, ATS4dv, GGI3, ATSC3dv, GATS1se, SpAbs_Dzv, VE1_A, VR1_A, ATSC4d, BCUTdv-1l, SlogP_VSA1, PEOE_VSA6, SlogP_VSA3, SlogP_VSA2, ATSC5dv, ATSC1dv, PEOE_VSA10, PEOE_VSA1, ATSC4c, ECIndex, ATSC1v, ATSC5d, VSA_EState3, VSA_EState7, ATSC6d, ATSC2d, n4HRing, SpMAD_A, PEOE_VSA9, ATSC6dv, ATS3v, PEOE_VSA8, ATSC4dv, PEOE_VSA13, VSA_EState6, n6aRing, ATSC1Z, ATSC1se, ATSC3Z, FilterItLogS, ATSC3v, ATS0dv, VSA_EState8, SsssCH, BalabanJ, VSA_EState9, ATSC5c, ATS4se, ATSC2Z, ATSC5se, n3HRing, ATSC2v, PEOE_VSA4, ATS5d, SMR_VSA4, ATSC4se, Xch-3d, ATS4d, SMR_VSA5, SssCH2, ZMIC2, ATSC7dv, EState_VSA6, ATSC5dv, GGI4, JGI4, ATSC6c, ETA_shape_p, Xc-5d, n5Ring, NtN, nC, ATS6d, EState_VSA2, NsssN, nBridgehead, EState_VSA3, SlogP_VSA5, NssO, EState_VSA4, IC2, ETA_shape_y, C3SP2, n5HRing, SMR_VSA6, C2SP1, SsssC, NdsN, EState_VSA5, ATSC7c
CFS	FCSP3, GATS1se, CIC1, C1SP1, MATS2c, VSA_EState1, VE2_Dzare, VSA_EState3, ATSC1m, SMR_VSA9, MIC0, AETA_eta_L, SaaCH, SpMAD_Dzi, VE2_A, BCUTd-1h, ETA_eta_FL, NsssCH, ATSC2se, GATS2Z, AETA_eta_R, VE2_Dt, AATS1Z, StsC, VE2_D, AATS1m, SlogP_VSA5, FilterItLogS, SaaC, ATSC4d, ATS5pe, SsssCH, EState_VSA9, VE2_Dzm, AMW, VE2_Dzp, piPC1, Mi, nBondsS, PEOE_VSA5, CIC2, BIC2, PEOE_VSA4, NssO, nH, ZMIC1, ATSC2c, Xc-6dv, ATSC4v, nBondsM, ATSC5dv, ATSC6dv, SpMAD_Dzse, RPCG, PEOE_VSA9, VE2_Dzpe, piPC4, BCUTc-1h, GATS2m, BCUTi-1h, PEOE_VSA10, PEOE_VSA6, JGI3, BalabanJ, SssCH2, ATSC7dv, EState_VSA2, VSA_EState6, VE2_DzZ
LIASSO	VR1_A, ATS1dv, ATS2dv, ATS1d, ATS2d, ATS3d, ATS5d, ATS4Z, ATS1m, ATS2m, ATS3m, ATS5m, ATS6m, ATS7m, ATS8m, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS2se, ATS4se, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATSC2dv, ATSC2d, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC6v, ATSC7v, ATSC8v, ATSC0i, ATSC2i, ATSC3i, SpDiam_Dzv, SpDiam_Dzp, BertzCT, C2SP2, SpAbs_Dt, SpAD_Dt, SdsCH, SaaCH, SsNH2, SssNH, StN, SsOH, SdO, ECIndex, fragCpx, nHBAcc, TIC1, TIC3, MIC1, ZMIC0, PEOE_VSA2, PEOE_VSA3, PEOE_VSA4, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, SMR_VSA3, SMR_VSA4, SMR_VSA6, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA3, SlogP_VSA4, SlogP_VSA5, SlogP_VSA11, EState_VSA3, EState_VSA4, EState_VSA5, EState_VSA6, EState_VSA7, EState_VSA9, VSA_EState1, VSA_EState2, MPC5, MPC6, MPC7, TopoPSA(NO), Vabc, TSRW10

A.2 Subconjunto com moléculas de 18 átomos

A.2.1 Energia interna de atomização para a temperatura de 0 K

Metódos	Atributos selecionados
Pcorr	ATS0v, SMR, PEOE_VSA1, SLogP, SlogP_VSA2, BCUTp-1h, VSA_EState7, ETA_beta, ETA_dEpsilon_D, SMR_VSA7, FCSP3, C2SP2, PEOE_VSA7, NsOH, SMR_VSA6, FilterItLogS, ATSC1se, SlogP_VSA4, ETA_dEpsilon_B, ATS4p, VSA_EState6, GATS1se, SdsCH, ATSC0i, SlogP_VSA5, C3SP2, piPC8, nRot, bpol, Vabc, RNCG, VSA_EState5, nBondsKD, PEOE_VSA6, ATSC2se, nBondsT, ETA_epsilon_5, ATSC3c, NssO, EState_VSA9, EState_VSA1, SaaCH, n5Ring, BCUTZ-1h, ATS5are, EState_VSA8, C1SP3, GATS2are, SdssC, ETA_eta_R, PEOE_VSA9, ETA_shape_p, VSA_EState9, nBondsD, ECIndex, WPol, StN, PEOE_VSA10, TIC5, GATS2i, SMR_VSA4, ATSC2c, EState_VSA2, n6aRing, VR1_A, EState_VSA5, ATS2m, EState_VSA7, SaasC, SlogP_VSA1, nBase, ATS4d, ZMIC5, ATS3i, EState_VSA4, SsCH2, n5FAHring, BalabanJ, ATS5d, n8FRing, ATSC3se, ZMIC0, SRW09, ATSC1dv, n6Ring, SpMAD_Dzi, StCH, SssssC, VE2_A, C3SP3, SlogP_VSA8, n4Hring, NsNH2, BCUTi-1h, NdsC, GATS1c, JGT10, nFARing, ATS7i, SlogP_VSA3, BCUTdv-1h, ATSC3i, C4SP3, NssNH, ATSC3dv, SlogP_VSA10, Xc-6d, JGI2, n7FARing, ATSC4c, AMID_N, BCUTc-1h, ETA_shape_x, RPCG, MIC5, C1SP2, VSA_EState2, SMR_VSA3, n9FARing, TIC1, BCUTc-1l, ATSC7d, NdNH, NdsN, AATSC2d, ATSC5d, PEOE_VSA8, ATSC3dv, ATSC7Z, NaaN, ETA_shape_y, NsCH3, JGI3, ATSC3Z, ATSC8are, JGI4, ATS4are, ATSC6i, ATSC4are, n3Hring, PEOE_VSA13, ATSC8c
Scorr	ATS0v, SMR, VE3_Dzp, PEOE_VSA1, SLogP, BCUTp-1h, ETA_dEpsilon_D, SlogP_VSA2, VSA_EState7, piPC6, ETA_beta, SMR_VSA6, VE1_Dt, NsOH, AATSC2se, PEOE_VSA7, FilterItLogS, ATS4p, GATS1are, RNCG, SpAD_A, ETA_dEpsilon_B, SlogP_VSA4, nH, SMR_VSA7, C2SP2, SlogP_VSA6, AATSC2c, VSA_EState6, VSA_EState5, ETA_eta_RL, SlogP_VSA5, ATSC3c, VSA_EState3, nRot, Vabc, bpol, C3SP2, piPC8, PEOE_VSA6, PEOE_VSA9, n5ARing, StsC, SdsCH, NssO, ATS5are, EState_VSA1, GATS2se, EState_VSA8, EState_VSA9, EState_VSA2, C1SP3, ZMIC5, VSA_EState9, GATS2i, SdssC, nBondsD, GATS1Z, VR1_A, StN, SssCH2, MATS1Z, AATS1Z, PEOE_VSA10, SssCH, ATS3i, EState_VSA5, C2SP1, SMR_VSA4, ATSC3se, n5FHRing, WPol, BCUTv-1l, ATS5d, ETA_shape_x, ATS2m, n6Ring, ATSC1dv, SpMAD_Dzi, BCUTd-1l, ECIndex, ATS4d, n8FARing, ETA_eta_R, n6aRing, TIC5, SlogP_VSA1, VSA_EState2, RPCG, ATS7i, EState_VSA4, BCUTi-1h, ZMIC2, JGI1, NsNH2, NdsC, nBase, EState_VSA7, ATSC3i, ATSC7v, SaaCH, C3SP3, SssssC, n4Hring, JGT10, ATSC3dv, BCUTc-1h, C4SP3, ATS3dv, ATSC4c, ATS4m, BalabanJ, n7Fring, VE2_Dzp, JGI2, NdNH, n4Ring, SlogP_VSA10, ETA_dPsi_B, ATSC8v, ATSC3Z, NssNH, n9FHRing, Xc-6d, SMR_VSA3, ATSC6c, C1SP2, NdsN, BIC2, ATSC6dv, SlogP_VSA8, AATSC2d, SlogP_VSA3, C2SP3, ZMIC1, SMR_VSA5, PEOE_VSA13, JGI4, ZMIC0, ATSC4pe, EState_VSA3, NaaN, VSA_EState4, VE2_D, SssN, PEOE_VSA2, ATSC5c, GGI4, ATS6dv, NsCH3, ATSC6d, ATSC6i, ATSC4i, JGI3, ATS4are, PEOE_VSA12, Radius, PEOE_VSA8, ATSC4d, PEOE_VSA3, n7ARing, nAcid, TIC1, n3AHRing, AETA_beta_ns_d, nX, nSpiro, n6FRing, SaaO, JGI5, EState_VSA10, VSA_EState8, BCUTc-1l, nAHRing, JGI7, GGI6, nBridgehead, n9Ring, SlogP_VSA11, ATS8dv, ETA_shape_y, NssCH, SaasN, ATS7dv, SaaNH, n8AHRing, MID_N, nN
RF	nC, ATSC0i, SMR, SLogP, SlogP_VSA2, ZMIC0, BCUTi-1h, nBondsKD, ETA_eta_R, ATSC2se, nN, BCUTZ-1h, SMR_VSA7, BertzCT, ATSC1se, nBase, ABC, n5Ring, BCUTZ-1l, VR1_A, BCUTdv-1h, SpAbs_A, nBondsD, GATS1se, VE1_A, nO, ATSC2c, C2SP2, RPCG, RNCG, SpAbs_Dzv, ATS0dv, BCUTd-1l, n5FRing, nBridgehead, VE2_A, ATS5se, BCUTc-1l, ATS5d, BCUTc-1h, SssssC, ATS1dv, SpAbs_Dzv, NdO, VSA_EState4, n6Ring, Xc-5d, ATSC1v, SlogP_VSA4, BCUTd-1h, EState_VSA8, NtN, GGI3, SlogP_VSA5, GATS2c, nRot, Xch-3d, SdssC, VSA_EState6, BalabanJ, nSpiro, ATS3dv, ETA_dEpsilon_B, ATSC2dv, PEOE_VSA11, EState_VSA9, NdsCH, nAromAtom, SMR_VSA3, C2SP1, FilterItLogS, ATSC2d, NsOH, ATS4d, VSA_EState7, SlogP_VSA3, GGI4, ATSC4d, NdsC, n7FRing, ATSC1dv, nHring, PEOE_VSA3, NdsN, PEOE_VSA7, BCUTdv-1l, ATSC2Z, IC1, PEOE_VSA12, ATSC3c, ZMIC1, ATSC6d, n6aRing, ATS6d, ATSC3dv, ZMIC2, n3Hring, ATSC2i, SMR_VSA5, SlogP_VSA1, ATSC3se, VSA_EState8, GGI2, PEOE_VSA8, PEOE_VSA2, JGI4, ATSC3Z, PEOE_VSA6, ATS4se, ETA_epsilon_5, ECIndex, ATS4dv, C1SP2, piPC8, ATSC2v, PEOE_VSA9, ATSC4dv, EState_VSA4, NssCH, ATSC5d, n8FRing, IC2, ATSC5dv, ETA_shape_p, nBondsT, ATSC4c, ATSC5se, ATSC4se, ATSC6c, C3SP2, n6FRing, EState_VSA5, SMR_VSA4, ATSC5c, SMR_VSA6, PEOE_VSA10, ATSC6dv, EState_VSA3, NssO, NssCH2
CFS	Mi, FCSP3, GATS1m, SlogP_VSA2, SpMAD_Dzp, nH, piPC3, AATS1se, VE2_Dzi, SpAbs_Dzse, ATSC1d, ATSC3d, ATS1i, VE2_D, nHBacc, NtCH, AATS0i, AETA_eta, ATS5se, VE2_Dzare, AATS1pe, n4Ring, GATS1v, AATS2p, ATS7d, ATS3p, PEOE_VSA1, ATSC1se, nRot, MATS1v, AATS2i, VE2_Dzpe, ATS3dv, VSA_EState5, BCUTm-1l, BCUTp-1h, ETA_dPsi_A, ATSC1c, AATS1i, VE2_Dt, AATS1are, n4ARing, JGI1, GATS2Z, ATS6i, ATSC5d, ATS0i, VE2_DzZ, FilterItLogS, NaaO, NssO, BCUTc-1l, ATS6are, Xc-4dv, AATSC2d, nBase, ATS4are, GATS2v, AATSC1dv, GATS2m, NsCH3, NsOH, SssCH2, AMID_h, AATSC0v, SpMAD_Dzv, VE2_Dzm, AATS1dv, GATS1Z, nAHRing, ETA_dAlpha_B, SM1_Dzp, ETA_epsilon_2, MATS1c, ETA_dEpsilon_D, AATSC1se, ETA_shape_p, nBondsS, AETA_beta_ns_d, ATS5dv, VE2_Dzse, AATS2are, AATS1p, NdsN, SlogP_VSA1, VE2_Dzv
LAGSSO	VR1_A, nBridgehead, ATS2Z, ATS1m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS2se, ATS0i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC8m, ATSC1v, ATSC2v, ATSC4v, ATSC5v, ATSC6v, ATSC8v, ATSC1i, ATSC3i, ATSC5i, SpAD_Dzp, BertzCT, SpDiam_Dt, DetourIndex, StN, SssO, ECIndex, ETA_beta, fragCpx, nHBacc, TIC1, TIC4, PEOE_VSA2, PEOE_VSA3, PEOE_VSA4, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA10, PEOE_VSA12, PEOE_VSA13, SMR_VSA3, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA4, SlogP_VSA6, EState_VSA2, EState_VSA3, EState_VSA8, EState_VSA9, VSA_EState2, VSA_EState4, VSA_EState6, VSA_EState7, MPC3, MPC5, MPC7, MPC8, TMPIC10, TopoPSA(NO)

A.2.2 Energia interna de atomização para a temperatura de 298.15 K

Metódos	Atributos selecionados
Pcorr	ATS0v, SMR, PEOE_VSA1, SLogP, SlogP_VSA2, BCUTp-1h, VSA_EState7, ETA_dEpsilon_D, ETA_beta, SMR_VSA7, FCSP3, PEOE_VSA7, C2SP2, NsOH, SMR_VSA6, ATSC1se, FilterItLogS, SlogP_VSA4, ETA_dEpsilon_B, ATS4p, VSA_EState6, SdsCH, GATS1se, ATSC0i, SlogP_VSA5, piPC8, nRot, C3SP2, RNCG, bpol, Vabc, VSA_EState5, PEOE_VSA6, nBondsKD, ATSC2se, ETA_epsilon_5, EState_VSA9, ATSC3c, nBondsT, NssO, EState_VSA1, ATS5are, n5Ring, SaaCH, BCUTz-1h, C1SP3, EState_VSA8, ETA_shape_p, SdssC, GATS2are, PEOE_VSA9, ETA_eta_R, VSA_EState9, PEOE_VSA10, nBondsD, ECIndex, WPol, StN, TIC5, SMR_VSA4, ATS2m, EState_VSA2, EState_VSA5, GATS2i, VR1_A, ATSC2c, n6aRing, SaasC, ATS3i, EState_VSA7, ATS4d, ZMIC5, SlogP_VSA1, EState_VSA4, SssCH2, BalabanJ, nBase, ATS5d, n5FAHRing, n8FARing, SpMAD_Dzi, SRW09, n6Ring, ATSC3se, ATSC1dv, ZMIC0, SssssC, C3SP3, StCH, nFRing, NsNH2, SlogP_VSA8, GATS1c, ATS7i, VE2_A, n4HRing, JGT10, BCUTi-1h, NdssC, SlogP_VSA3, ATSC3i, BCUTdv-1h, C4SP3, NssNH, ATSC3dv, SlogP_VSA10, n7FARing, JGI2, BCUTc-1h, Xc-6d, AMID_N, ATSC4c, n9FAHRing, RPCG, ETA_shape_x, ATSC7d, SMR_VSA3, MIC5, VSA_EState2, C1SP2, ATSC5d, BCUTc-1l, NdNH, NdsN, PEOE_VSA8, TIC1, ATS3dv, ETA_shape_y, NsCH3, ATSC2d, ATSC7z, ATSC3z, NaaN, ATSC8pe, JGI3, JGI4, PEOE_VSA13, ATSC4are, ATSC6i, ATS4are, ATSC8c, C2SP3, ATSC5i, n3AHRing, nAcid, PEOE_VSA2, GGI4, ATSC4i, ATSC6dv, ATSC5c, ATSC5are, NsF, SaasN, EState_VSA3, JGI5, ATSC7c, AETA_beta_ns_d, SssN, VSA_EState4, SaaNH, n7ARing, Radius, PEOE_VSA3, ATSC4d, ZMIC1, ATSC6d, SMR_VSA5, PEOE_VSA12, n6FRing, nAHRing, nSpiro, ATSC6c, BCUTd-1l, SlogP_VSA11, SaaO, Xch-3dv, JGI7, ATS8dv, n9ARing, GGI6, EState_VSA6
SCorr	ATS0v, SMR, VE3_Dzp, PEOE_VSA1, SLogP, BCUTp-1h, ETA_dEpsilon_D, SlogP_VSA2, VSA_EState7, piPC6, ETA_beta, SMR_VSA6, VE1_Dt, NsOH, AATSC2se, PEOE_VSA7, FilterItLogS, RNCG, SpAD_-A, GATS1are, ATS4p, ETA_dEpsilon_B, SMR_VSA7, SlogP_VSA4, C2SP2, nH, SlogP_VSA6, AATSC2c, VSA_EState6, VSA_EState5, ETA_eta_RL, VSA_EState3, nRot, SlogP_VSA5, ATSC3c, Vabc, piPC8, bpol, C3SP2, PEOE_VSA6, n5ARing, PEOE_VSA9, ATS5are, SdsCH, StsC, NssO, EState_VSA1, GATS2se, EState_VSA8, EState_VSA9, EState_VSA2, C1SP3, ZMIC5, VSA_EState9, SdssC, GATS2i, GATS1Z, MATS1Z, nBondsD, VR1_A, SssCH2, SssCH, ATS3i, PEOE_VSA10, StN, EState_VSA5, AATS1Z, SMR_VSA4, ATS5d, BCUTv-1l, ATS2m, ATSC3se, n5FHRing, C2SP1, SpMAD_Dzi, WPol, n6Ring, ETA_shape_x, ATSC1dv, ATS4d, n8FARing, ECIndex, TIC5, n6aRing, SlogP_VSA1, ATS7i, EState_VSA4, BCUTd-1l, JGI1, RPCG, VSA_EState2, ZMIC2, ETA_eta_R, NsNH2, BCUTi-1h, NdssC, C3SP3, ATSC3i, ATSC7v, nBase, SssssC, EState_VSA7, SaaCH, BCUTc-1h, ATS4m, ATSC3dv, JGT10, n4HRing, C4SP3, ATSC3dv, BalabanJ, ATSC4c, n7FARing, VE2_Dzp, JGI2, ATSC3Z, n9FARing, NdNH, SlogP_VSA10, ATSC8v, ETA_dPsi_B, NssNH, n4ARing, SMR_VSA3, Xc-6d, ATSC6c, BIC2, NdsN, C2SP3, SlogP_VSA8, SlogP_VSA3, SMR_VSA5, C1SP2, ATSC6dv, ZMIC0, PEOE_VSA13, JGI4, ATSC4are, AATSC2d, NaaN, VSA_EState4, SssN, ZMIC1, EState_VSA3, ATS6dv, NsCH3, PEOE_VSA2, ATSC5c, ATSC6d, VE2_D, GGI4, ATSC6i, ATSC4i, ATSC4are, PEOE_VSA8, ATSC4d, JGI3, PEOE_VSA12, PEOE_VSA3, n7ARing, BIC1, nAcid, Radius, AETA_beta_ns_d, n3HRing, nF, n6FRing, SaaO, nSpiro, EState_VSA10, JGI5, VSA_EState8, BCUTc-1l, NssCH, ETA_shape_y, GGI6, ATS7dv, n9ARing, ATS8dv
RF	nC, ATSC0i, SMR, SLogP, ZMIC0, SlogP_VSA2, nBondsKD, BCUTi-1h, ATSC2se, ETA_eta_R, BertzCT, BCUTz-1h, SMR_VSA7, nBase, n5Ring, nN, ATSC1se, ABC, VR1_A, nBondsD, SpAbs_A, BCUTdv-1h, VE1_A, GATS1se, ATS0dv, nO, ATSC2c, BCUTz-1l, C2SP2, RPCG, SpAbs_Dzv, n5FRing, RNCG, BCUTd-1l, ATS5se, nBridgehead, VE2_A, NtN, BCUTc-1l, ATS5d, BCUTc-1h, ATSC1dv, SssssC, SpAbs_DzZ, n6Ring, VSA_EState4, NdO, Xc-5d, SlogP_VSA4, ATSC1v, EState_VSA8, BCUTd-1h, nRot, VSA_EState6, SdssC, Xch-3d, GGI3, EState_VSA9, GATS2c, BalabanJ, ATS3dv, nSpiro, ETA_dEpsilon_B, ATSC2dv, SlogP_VSA5, FilterItLogS, NdsCH, PEOE_VSA11, nAromAtom, C2SP1, SMR_VSA3, NdssC, ATSC2d, NsOH, ATS4d, n7FRing, SlogP_VSA3, PEOE_VSA3, VSA_EState7, nHRing, ATSC1dv, ETA_epsilon_5, GGI4, ATSC4d, IC1, ATSC2Z, BCUTdv-1l, ZMIC1, ATS6d, n3HRing, ATSC3dv, ATSC3c, ATSC6d, ZMIC2, ATSC2i, NdsN, PEOE_VSA6, PEOE_VSA2, SlogP_VSA1, JGI4, n6aRing, ATSC3se, PEOE_VSA12, VSA_EState8, ATS4dv, ATSC3Z, GGI2, PEOE_VSA8, piPC8, SMR_VSA5, ATS4se, ATSC2v, ECIndex, C1SP2, ATSC4dv, EState_VSA4, ATSC5d, n8FRing, n6FRing, NssCH, PEOE_VSA7, ETA_shape_p, ATSC5se, IC2, ATSC4se, ATSC5dv, PEOE_VSA9, ATSC4c, SMR_VSA6, C3SP2, ATSC5c, ATSC6c, nBondsT, SMR_VSA4, PEOE_VSA10, EState_VSA5, ATSC6dv, EState_VSA3, NssNH, NssO
CFS	Mi, FCSP3, GATS1m, SlogP_VSA2, SpMAD_Dzp, nH, piPC3, AATS1se, VE2_Dzi, SpAbs_Dzse, ATSC1d, ATSC3d, ATSC5d, AATS2i, VE2_D, AATS1pe, n4Ring, GATS1Z, nHBAcc, AATSC1se, nBondsS, ETA_dEpsilon_D, GATS1v, BCUTv-1h, BCUTse-1l, PEOE_VSA5, PEOE_VSA1, ATSC1se, ATSC1v, GATS1p, nBase, nRot, AETA_eta, AETA_beta_ns_d, AATS0i, AATSC0p, ETA_epsilon_2, ETA_dEpsilon_C, ETA_dAlpha_B, ATSC1c, nHBDon, MATS1pe, AMID_O, SM1_Dzp, ATS0i, FilterItLogS, CIC5, BIC5, CIC4, BIC4, CIC3, SIC3, VE2_Dzare, ATS5d, GGI1, ATSC3dv, n4ARing, AATS1i, VE2_Dzpe, AATS1are, AATS0p, C1SP3, C2SP3, BalabanJ, NdsN, SRW09, VSA_EState9, GATS2v, VSA_EState8, SlogP_VSA10, nO, ETA_psi_1, TopoPSA(NO), AATSC1c, n4HRing, MATS2c, GATS2pe, TopoPSA, VE2_DzZ, AATSC0m, VSA_EState5, SMR_VSA3, ETA_shape_x, VE2_Dzm, SM1_Dzse, MATS1c, ATS6se, NsOH, nBridgehead, SsNH2, bpol, BCUTc-1l, ATSC3Z, ATS4i, AATSC1pe, AATSC0c, SM1_Dzv, SpMAD_Dzv, JGT10, VE2_-A, AMID_h, AATSC0v, ATS5Z, AATSC1v, GATS1i, VE2_Dzse, VSA_EState3, BCUTdv-1h, Mp, SM1_Dzi, CIC0, AATS2se, AETA_alpha, ETA_dPsi_A, ATS0p, SM1_Dzare, VE2_Dt, piPC1, ATSC1i, ATSC2d, JGI2, SMR_VSA6, SlogP_VSA5, PEOE_VSA2, SlogP_VSA4, AATS2are, VSA_EState7, n3HRing, BCUTi-1l, VE2_Dzv, BIC1, n4AHRing, ATS7p, AATSC1i, NsCH3, BCUTdv-1l, PEOE_VSA10, CIC1, MID_h, VE2_Dzp, SIC1, PEOE_VSA9, PEOE_VSA8, PEOE_VSA7, ETA_shape_p, AETA_eta_RL, ATS4pe, C2SP1, BCUTpe-1l, PEOE_VSA6, AATS2pe, SpDiam_A, ATS6i, ATSC5p, CIC2, BIC2, SMR_VSA9, ATSC6d, ATSC4dv, AATSC2d, Xc-6d, RPCG, PEOE_VSA11, GATS2m, Si, RNCG, BCUTv-1l, SM1_Dzpe, MATS1p, SsCH3, GATS2Z, MATS2Z, NssCH2, SlogP_VSA1, ETA_shape_y, ETA_dBeta
LIASSO	VR1_A, nBridgehead, ATS2Z, ATS1m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS2se, ATS0i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC8m, ATSC1v, ATSC2v, ATSC4v, ATSC5v, ATSC6v, ATSC8v, ATSC1i, ATSC3i, ATSC5i, SpAD_Dzp, BertzCT, SpDiam_Dt, DetourIndex, StN, SssO, ECIndex, ETA_beta, fragCpx, nHBAcc, TIC1, TIC4, PEOE_VSA2, PEOE_VSA3, PEOE_VSA4, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA10, PEOE_VSA12, PEOE_VSA13, SMR_VSA3, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA4, SlogP_VSA6, EState_VSA2, EState_VSA3, EState_VSA5, EState_VSA8, EState_VSA9, VSA_EState2, VSA_EState4, VSA_EState6, VSA_EState7, MPC3, MPC5, MPC7, MPC8, TMPC10, TopoPSA(NO)

A.2.3 Entalpia de atomização para a temperatura de 298.15 K

Metódos	Atributos selecionados
Pcorr	ATS0v, SMR, PEOE_VSA1, SLogP, SlogP_VSA2, BCUTp-1h, VSA_EState7, ETA_dEpsilon_D, ETA_beta, SMR_VSA7, FCSP3, PEOE_VSA7, C2SP2, NsOH, SMR_VSA6, ATSC1se, FilterItLogS, SlogP_VSA4, ETA_dEpsilon_B, ATS4p, VSA_EState6, SdsCH, GATS1se, ATSC0i, SlogP_VSA5, piPC8, nRot, C3SP2, RNCG, bpol, Vabc, VSA_EState5, PEOE_VSA6, nBondsKD, ATSC2se, ETA_epsilon_5, EState_VSA9, ATSC3c, nBondsT, NssO, EState_VSA1, ATS5are, n5Ring, SaaCH, BCUTz-1h, C1SP3, EState_VSA8, ETA_shape_p, SdssC, GATS2are, PEOE_VSA9, ETA_eta_R, VSA_EState9, PEOE_VSA10, nBondsD, ECIndex, WPol, StN, TIC5, SMR_VSA4, ATS2m, EState_VSA2, EState_VSA5, GATS2i, VR1_A, ATSC2c, n6aRing, SaasC, ATS3i, EState_VSA7, ATS4d, ZMIC5, SlogP_VSA1, EState_VSA4, SssCH2, BalabanJ, nBase, ATS5d, n5FAHRing, n8FARing, SpMAD_Dzi, SRW09, n6Ring, ATSC3se, ATSC1dv, ZMIC0, SssssC, C3SP3, StCH, nFRing, NsNH2, SlogP_VSA8, GATS1c, ATS7i, VE2_A, n4HRing, JGT10, BCUTi-1h, NdssC, SlogP_VSA3, ATSC3i, BCUTdv-1h, C4SP3, NssNH, ATSC3dv, SlogP_VSA10, n7FARing, JGI2, BCUTc-1h, Xc-6d, AMID_N, ATSC4c, n9FAHRing, RPCG, ETA_shape_x, ATSC7d, SMR_VSA3, MIC5, VSA_EState2, C1SP2, ATSC5d, BCUTc-1l, NdNH, NdsN, PEOE_VSA8, TIC1, ATS3dv, ETA_shape_y, NsCH3, ATSC2d, ATSC7z, ATSC3z, NaaN, ATSC8pe, JGI3, JGI4, PEOE_VSA13, ATSC4are, ATSC6i, ATS4are, ATSC8c, C2SP3, ATSC5i, n3AHRing, nAcid, PEOE_VSA2, GGI4, ATSC4i, ATSC6dv, ATSC5c, ATSC5are, NsF, SaasN, EState_VSA3, JGI5, ATSC7c, AETA_beta_ns_d, SssN, VSA_EState4, SaaNH, n7ARing, Radius, PEOE_VSA3, ATSC4d, ZMIC1, ATSC6d, SMR_VSA5, PEOE_VSA12, n6FRing, nAHRing, nSpiro, ATSC6c, BCUTd-1l, SlogP_VSA11, SaaO, Xch-3dv, JGI7, ATS8dv, n9ARing, GGI6, EState_VSA6
Scorr	ATS0v, SMR, VE3_Dzp, PEOE_VSA1, SLogP, BCUTp-1h, ETA_dEpsilon_D, SlogP_VSA2, VSA_EState7, piPC6, ETA_beta, SMR_VSA6, VE1_Dt, NsOH, AATSC2se, PEOE_VSA7, FilterItLogS, RNCG, SpAD_A, GATS1are, ATS4p, ETA_dEpsilon_B, SMR_VSA7, SlogP_VSA4, C2SP2, nH, SlogP_VSA6, AATSC2c, VSA_EState6, VSA_EState5, ETA_eta_RL, VSA_EState3, nRot, SlogP_VSA5, ATSC3c, Vabc, piPC8, bpol, C3SP2, PEOE_VSA6, n5ARing, PEOE_VSA9, ATS5are, SdsCH, StsC, NssO, EState_VSA1, GATS2se, EState_VSA8, EState_VSA9, EState_VSA2, C1SP3, ZMIC5, VSA_EState9, SdssC, GATS2i, GATS1Z, MATS1Z, nBondsD, VR1_A, SssCH2, SssssCH, ATS3i, PEOE_VSA10, StN, EState_VSA5, AATS1Z, SMR_VSA4, ATS5d, BCUTv-1l, ATS2m, ATSC3se, n5FHRing, C2SP1, SpMAD_Dzi, WPol, n6Ring, ETA_shape_x, ATSC1dv, ATS4d, n8FARing, ECIndex, TIC5, n6aRing, SlogP_VSA1, ATS7i, EState_VSA4, BCUTd-1l, JGI1, RPCG, VSA_EState2, ZMIC2, ETA_eta_R, NsNH2, BCUTi-1h, NdssC, C3SP3, ATSC3i, ATSC7v, nBase, SssssC, EState_VSA7, SaaCH, BCUTc-1h, ATS4m, ATSC3dv, JGT10, n4HRing, C4SP3, ATS3dv, BalabanJ, ATSC4c, n7FARing, VE2_Dzp, JGI2, ATSC3Z, n9FARing, NdNH, SlogP_VSA10, ATSC8v, ETA_dPsi_B, NssNH, n4ARing, SMR_VSA3, Xc-6d, ATSC6c, BIC2, NdsN, C2SP3, SlogP_VSA8, SlogP_VSA3, SMR_VSA5, C1SP2, ATSC6dv, ZMIC0, PEOE_VSA13, JGI4, ATSC4are, AATSC2d, NaaN, VSA_EState4, SssN, ZMIC1, EState_VSA3, ATS6dv, NsCH3, PEOE_VSA2, ATSC5c, ATSC6d, VE2_D, GGI4, ATSC6i, ATSC4i, ATSC4are, PEOE_VSA8, ATSC4d, JGI3, PEOE_VSA12, PEOE_VSA3, n7ARing, BIC1, nAcid, Radius, AETA_beta_ns_d, n3HRing, nF, n6FRing, SaaO, nSpiro, EState_VSA10, JGI5, VSA_EState8, BCUTc-1l, NssssCH, ETA_shape_y, GGI6, ATS7dv, n9ARing, ATS8dv
RF	nC, ATSC0i, SMR, SLogP, SlogP_VSA2, ZMIC0, nBondsKD, BCUTi-1h, BertzCT, ATSC2se, ETA_eta_R, SMR_VSA7, nN, BCUTz-1h, nBase, ABC, n5Ring, SpAbs_A, VR1_A, BCUTdv-1h, ATSC1se, GATS1se, nBondsD, ATS0dv, VE1_A, BCUTz-1l, ATSC2c, C2SP2, nO, RPCG, RNCG, n5FRing, SpAbs_Dzv, BCUTd-1l, nBridgehead, ATS5se, VE2_A, BCUTc-1l, ATS1dv, NtN, ATS5d, SssssC, BCUTc-1h, n6Ring, SpAbs_DzZ, VSA_EState4, NdO, ATSC1v, Xc-5d, SlogP_VSA4, BCUTd-1h, VSA_EState6, EState_VSA8, nRot, ATS3dv, Xch-3d, SlogP_VSA5, ATSC2dv, GGI3, ETA_dEpsilon_B, GATS2c, BalabanJ, NdsCH, nSpiro, C2SP1, SdssC, FilterItLogS, EState_VSA9, nAromAtom, NsOH, SMR_VSA3, ATSC2d, PEOE_VSA11, ATS4d, NdssC, SlogP_VSA3, VSA_EState7, n7FRing, GGI4, IC1, ATSC4d, nHRing, ETA_epsilon_5, PEOE_VSA3, ATSC1dv, BCUTdv-1l, ATSC2Z, ATSC2i, NdsN, ATS6d, ATSC3dv, ATSC6d, n6aRing, ZMIC1, ATSC3c, n3HRing, PEOE_VSA6, SlogP_VSA1, NssO, PEOE_VSA2, JGI4, ZMIC2, ATSC3se, SMR_VSA5, PEOE_VSA12, ATSC3Z, PEOE_VSA8, VSA_EState8, ATS4se, GGI2, ATS4dv, C1SP2, PEOE_VSA9, ECIndex, ATSC2v, ATSC4dv, piPC8, NssssCH, EState_VSA4, C3SP2, n6FRing, ATSC5d, IC2, n8FRing, ATSC5dv, ATSC4c, ATSC4se, PEOE_VSA10, ETA_shape_p, PEOE_VSA7, EState_VSA5, ATSC5se, SMR_VSA6, ATSC5c, ATSC6c, SMR_VSA4, EState_VSA3, ATSC6dv, NssNH
CFS	Mi, FCSP3, GATS1m, SlogP_VSA2, SpMAD_Dzp, nH, piPC3, AATS1se, VE2_Dzi, SpAbs_Dzse, ATSC1d, ATSC3d, ATSC5d, AATS2i, VE2_D, AATS1pe, n4Ring, GATS1Z, nHBAcc, AATSC1se, nBondsS, ETA_dEpsilon_D, GATS1v, BCUTv-1h, BCUTse-1l, PEOE_VSA5, PEOE_VSA1, ATSC1se, ATSC1v, GATS1p, nBase, nRot, AETA_eta, AETA_beta_ns_d, AATS0i, AATSC0p, ETA_epsilon_2, ETA_dEpsilon_C, ETA_dAlpha_B, ATSC1c, nHBDdon, MATS1pe, AMID_O, SM1_Dzp, ATS0i, FilterItLogS, CIC5, BIC5, CIC4, BIC4, CIC3, SIC3, VE2_Dzare, ATS5d, GGI1, ATSC3dv, n4ARing, AATS1i, VE2_Dzpe, AATS1are, AATS0p, C1SP3, C2SP3, BalabanJ, NdsN, SRW09, VSA_EState9, GATS2v, VSA_EState8, SlogP_VSA10, nO, ETA_psi_1, TopoPSA(NO), AATSC1c, n4HRing, MATS2c, GATS2pe, TopoPSA, VE2_DzZ, AATSC0m, VSA_EState5, SMR_VSA3, ETA_shape_x, VE2_Dzm, SM1_Dzse, MATS1c, ATS6se, NsOH, nBridgehead, SsNH2, bpol, BCUTc-1l, ATSC3Z, ATS4i, AATSC1pe, AATSC0c, SM1_Dzv, SpMAD_Dzv, JGT10, VE2_A, AMID_h, AATSC0v, ATS5Z, AATSC1v, GATS1i, VE2_Dzse, VSA_EState3, BCUTdv-1h, Mp, SM1_Dzi, CIC0, AATS2se, AETA_alpha, ETA_dPsi_A, ATS0p, SM1_Dzare, VE2_Dt, piPC1, ATSC1i, ATSC2d, JGI2, SMR_VSA6, SlogP_VSA5, PEOE_VSA2, SlogP_VSA4, AATS2are, VSA_EState7, n3HRing, BCUTi-1l, VE2_Dzv, BIC1, n4AHRing, ATS7p, AATSC1i, NsCH3, BCUTdv-1l, PEOE_VSA10, CIC1, MID_h, VE2_Dzp, SIC1, PEOE_VSA9, PEOE_VSA8, PEOE_VSA7, ETA_shape_p, AETA_eta_RL, ATS4pe, C2SP1, BCUTpe-1l, PEOE_VSA6, AATS2pe, SpDiam_A, ATS6i, ATSC5p, CIC2, BIC2, SMR_VSA9, ATSC6d, ATSC4dv, AATSC2d, Xc-6d, RPCG, PEOE_VSA11, GATS2m, Si, RNCG, BCUTv-1l, SM1_Dzpe, MATS1p, SsCH3, GATS2Z, MATS2Z, NssCH2, SlogP_VSA1, ETA_shape_y, ETA_dBeta
LIASSO	VR1_A, nBridgehead, ATS2Z, ATS1m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS2se, ATS0i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC8m, ATSC1v, ATSC2v, ATSC4v, ATSC5v, ATSC6v, ATSC8v, ATSC1i, ATSC3i, ATSC5i, SpAD_Dzp, BertzCT, SpDiam_Dt, DetourIndex, StN, SssO, ECIndex, ETA_beta, fragCpx, nHBAcc, TIC1, TIC4, PEOE_VSA2, PEOE_VSA3, PEOE_VSA4, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA10, PEOE_VSA12, PEOE_VSA13, SMR_VSA3, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA4, SlogP_VSA6, EState_VSA2, EState_VSA3, EState_VSA5, EState_VSA8, EState_VSA9, VSA_EState2, VSA_EState4, VSA_EState6, VSA_EState7, MPC3, MPC5, MPC7, MPC8, TMPC10, TopoPSA(NO)

A.2.4 Energia livre de Gibbs de atomização para a temperatura de 298.15 K

Metódos	Atributos selecionados
Pcorr	ATS0v, SMR, SLogP, PEOE_VSA1, SlogP_VSA2, BCUTp-1h, VSA_EState7, ETA_beta, ETA_dEpsilon_D, SMR_VSA7, FCSP3, PEOE_VSA7, C2SP2, NsOH, FilterItLogS, SMR_VSA6, SlogP_VSA4, ATSC1se, ETA_dEpsilon_B, ATS4p, VSA_EState6, GATS1se, SdsCH, SlogP_VSA5, ATSC0i, C3SP2, Vabc, piPC8, bpol, nRot, RNCG, VSA_EState5, PEOE_VSA6, nBondsKD, nBondsT, ATSC2se, ATSC3c, ETA_epsilon_5, NssO, EState_VSA9, EState_VSA1, SaaCH, BCUTZ-1h, EState_VSA8, C1SP3, n5Ring, GATS2are, ATS5are, PEOE_VSA9, SdssC, ETA_eta_R, nBondsD, VSA_EState9, StN, ETA_shape_p, PEOE_VSA10, ECIndex, GATS2i, ATSC2c, WPol, EState_VSA7, n6aRing, SMR_VSA4, EState_VSA2, nBase, SaasC, EState_VSA5, TIC5, SlogP_VSA1, VR1_A, ZMIC5, n5FAHring, SssCH2, ATS2m, EState_VSA4, ATS4d, StCH, ATS3i, ATSC3se, ATSC1dv, n4AHRing, BalabanJ, ATS5d, n8FARing, n6Ring, SRW09, VE2_A, SssssC, ZMIC0, SlogP_VSA8, NdssC, NsNH2, SpMAD_Dzi, GATS1c, JGT10, C3SP3, BCUTdv-1h, SlogP_VSA3, BCUTi-1h, ATSC3i, C4SP3, NssNH, ATS7i, Xc-6d, nFARing, AMID_N, SlogP_VSA10, JGI2, SMR_VSA3, ETA_shape_x, ATSC4c, RPCG, ATSC3dv, C1SP2, VSA_EState2, BCUTc-1h, n7FARing, ATS3dv, BCUTc-1l, AATSC2d, TIC1, NdsN, NdNH, n9FAHring, JGI3, ATSC7d, MIC5, NaaN, ATSC7dv, PEOE_VSA8, ATS4are, n3AHRing, ATSC8are, ATSC5d, ATSC3Z, ATSC4are, NsCH3, ATSC6i, PEOE_VSA13, ATSC8c, ETA_shape_y, nAcid, ATSC6dv, JGI4, AATSC1d, EState_VSA3, GGI4, Radius, ATSC4i, nAHRing, PEOE_VSA2, ATSC5c, JGI5, nX, ZMIC1, SaasN, ATSC5i, ATSC5are, ATSC7c, ATSC6d, C2SP3, SaaNH, VSA_EState4, PEOE_VSA3, PEOE_VSA12, AETA_beta_ns_d, n7Ring, ATSC4d, nSpro, SMR_VSA5, ATSC6c, SssN, n6FRing, SlogP_VSA11, SaaO, ATSC8i, JGI7, EState_VSA6, Xch-3d, nBridgehead, n9AHRing
Scorr	ATS0v, SMR, SLogP, PEOE_VSA1, VE3_Dzp, BCUTp-1h, SlogP_VSA2, ETA_dEpsilon_D, VSA_EState7, piPC6, ETA_beta, SMR_VSA6, NsOH, AATSC2se, FilterItLogS, PEOE_VSA7, ATS4p, VE1_Dt, GATS1are, RNCG, ETA_dEpsilon_B, SlogP_VSA4, C2SP2, SMR_VSA7, SpAD_A, nH, AATSC2c, SlogP_VSA5, SlogP_VSA6, VSA_EState5, ATSC3c, VSA_EState3, VSA_EState6, Vabc, ETA_eta_RL, nRot, bpol, PEOE_VSA6, C3SP2, StsC, PEOE_VSA9, piPC8, SdsCH, NssO, n5ARing, GATS2se, EState_VSA1, ATS5are, EState_VSA8, EState_VSA9, EState_VSA2, C1SP3, GATS2i, AATS1Z, nBondsD, ZMIC5, VSA_EState9, SdssC, StN, C2SP1, SssCH2, PEOE_VSA10, VR1_A, GATS1Z, EState_VSA5, ATSC3se, MATS1Z, n5FHRing, SssCH, BCUTd-1l, SMR_VSA4, ETA_shape_x, BCUTv-1l, ATS3i, ATSC1dv, WPol, n6Ring, ATSC5d, ECIndex, ETA_eta_R, SlogP_VSA1, n6aRing, RPCG, SpMAD_Dzi, VSA_EState2, n8FARing, nBase, ATS2m, ZMIC2, ATS4d, TIC5, NdssC, JGI1, BCUTi-1h, EState_VSA7, NsNH2, EState_VSA4, n4AHRing, ATS7i, SaaCH, ATSC7v, ATSC3i, JGT10, ATS3dv, SssssC, C3SP3, BCUTc-1h, C4SP3, ATSC4c, ATSC3dv, n4ARing, VE2_Dzp, ATS4m, JGI2, ETA_dPsi_B, n7FARing, NssNH, NdNH, SlogP_VSA10, Xc-6d, BalabanJ, SMR_VSA3, ATSC7c, ATSC8v, ATSC6c, C1SP2, AATS1dv, ATSC3Z, BIC2, n9FAHring, ATSC6dv, NdsN, AATSC2d, ZMIC0, SlogP_VSA8, ZMIC1, EState_VSA3, ATSC4pe, PEOE_VSA13, VE2_D, SMR_VSA5, SlogP_VSA3, NaaN, JGI4, VSA_EState4, C2SP3, AATSC1d, JGI3, ATSC5c, ATSC6d, ATS4are, GGI4, Radius, ATSC4i, PEOE_VSA2, ATSC6i, PEOE_VSA12, SssN, PEOE_VSA3, NsCH3, ATSC6dv, ATSC4d, nAcid, n7ARing, n3AHRing, PEOE_VSA8, BIC1, nSpro, JGI5, nF, AETA_beta_ns_d, nAHRing, SaaO, n6FARing, VSA_EState8, nBridgehead, EState_VSA10, AMID_N, JGI7, nN, BCUTc-1l, SlogP_VSA11, n9ARing
RF	nC, ATSC0i, SMR, SLogP, SlogP_VSA2, ZMIC0, nBondsKD, BCUTi-1h, ATSC2se, nN, BCUTZ-1h, ETA_eta_R, BCUTZ-1l, SMR_VSA7, ABC, nBase, n5Ring, ATSC1se, BCUTdv-1h, GATS1se, BertzCT, VR1_A, SpAbs_A, nBondsD, ATS0dv, VE1_A, ATSC2c, nO, RPCG, BCUTd-1l, C2SP2, RNCG, SpAbs_Dzv, n5FRing, nBridgehead, ATS5se, NdO, VE2_A, VSA_EState4, SssssC, BCUTc-1l, ATS5d, ATSC1dv, BCUTc-1h, Xc-5d, SpAbs_DzZ, SlogP_VSA4, ATSC1v, VSA_EState6, n6Ring, BCUTd-1h, VSA_EState7, Xch-3d, EState_VSA8, NtN, GGI3, GATS2c, BalabanJ, nRot, SlogP_VSA5, ETA_dEpsilon_B, ATSC3dv, nAromAtom, nSpro, EState_VSA9, SMR_VSA3, SdssC, ATSC2dv, NdsCH, FilterItLogS, SlogP_VSA3, ATSC2d, ATS4d, PEOE_VSA11, PEOE_VSA3, GGI4, NsOH, nHRing, C2SP1, ATSC4d, NdssC, n7FRing, SlogP_VSA1, ATSC1dv, BCUTdv-1l, ATSC6d, ATSC2Z, NdsN, PEOE_VSA6, n3HRing, IC1, C1SP2, ETA_epsilon_5, ATSC3dv, ZMIC2, PEOE_VSA2, n6aRing, ATSC3c, PEOE_VSA12, GGI2, SMR_VSA5, ATSC2i, ZMIC1, JGI4, ATSC3Z, ATSC3se, ATS6d, PEOE_VSA8, VSA_EState8, ATS4se, ATSC2v, ATS4dv, ECIndex, ATSC4dv, piPC8, EState_VSA4, C3SP2, n6FRing, ATSC5d, PEOE_VSA9, nBondsT, n8FRing, ATSC5dv, NssO, PEOE_VSA7, ETA_shape_p, ATSC4c, IC2, ATSC4se, ATSC5c, ATSC5se, NssCH, ATSC6c, SMR_VSA6, EState_VSA5, SMR_VSA4, EState_VSA3, PEOE_VSA10, ATSC6dv, NssNH
CFS	Mi, FCSP3, GATS1m, SlogP_VSA2, SpMAD_Dzp, VE2_Dt, ATS1dv, nH, nHBAcc, n4Ring, ETA_dEpsilon_B, AATS0i, Xch-3d, VE2_Dzare, AATS1se, ATSC1c, ETA_dEpsilon_D, n4ARing, GATS1v, NtCH, BCUTpe-1l, BCUTp-1h, ETA_epsilon_4, SM1_Dzp, ETA_dAlpha_B, VE2_D, ETA_eta_RL, PEOE_VSA6, PEOE_VSA1, PEOE_VSA7, AATS2i, AATS1p, ETA_dPsi_A, AETA_alpha, ATS0i, MATS1c, AMID_O, GATS1Z, VR1_Dt, ETA_shape_p, SssCH2, AATS1i, AETA_eta_L, NsCH3, GATS2m, AMID_h, AATSC0v, ETA_epsilon_2, ETA_dEpsilon_C, TopoPSA(NO), n4HRing, VE2_Dzpe, SM1_Dzi, VSA_EState5, AATSC0se, ATSC1se, nHBDOn, VE2_Dzi, AATS1pe, SssN, JGI1, GATS2Z, SsCH3, GATS2v, AATS1are, AATSC1se, BalabanJ, nBondsKS, TopoPSA, VSA_EState7, SMR_VSA6, JGI3, GATS1p, MATS1p, ATS5d, VE2_DzZ, MID_h, AMID_C, NdsN, AATSC0m, ETA_psi_1, SlogP_VSA1, AATSC1c, AATSC0Z, VE2_Dzv, ETA_beta_ns_d, FilterItLogS, Xc-4dv, SpMAD_Dzv, nBase, n5aRing, VE2_Dzse, ATSC0Z, SlogP_VSA4, nHRing, SlogP_VSA5, mZagreb2, nBondsS, SMR_VSA7, VSA_EState2, VSA_EState3, AATSC1pe, GATS1pe, BCUTdv-1h, CIC0, SM1_Dzare, ATSC1d, ATSC3d, ATSC5d, ATS7i, n5ARing, ATS4p, ATSC4i, CIC1
LASSO	VR1_A, nBridgehead, ATS2Z, ATS1m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS2se, ATS0i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC8m, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC6v, ATSC8v, ATSC1i, ATSC3i, ATSC5i, SpAD_Dzp, BertzCT, SpDiam_Dt, DetourIndex, StN, SssO, ECIndex, ETA_beta, fragCpx, nHBAcc, TIC1, TIC4, PEOE_VSA2, PEOE_VSA3, PEOE_VSA4, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA10, PEOE_VSA12, PEOE_VSA13, SMR_VSA3, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA4, SlogP_VSA6, EState_VSA2, EState_VSA3, EState_VSA8, EState_VSA9, VSA_EState2, VSA_EState4, VSA_EState6, VSA_EState7, MPC3, MPC5, MPC7, TMPC10, TopoPSA(NO)

A.2.5 Capacidade térmica para a temperatura de 298.15 K

Metódos	Atributos selecionados
Pcorr	ATS2i, ETA_shape_p, AETA_eta_R, AETA_eta_F, ATS4Z, SpMAD_D, SpMAD_A, ATS5pe, ATS3i, BalabanJ, TopoPSA(NO), nFRing, SsssCH2, VE2_A, nAHRing, EState_VSA4, BCUTd-1l, BIC1, NsssCH, RNCG, piPC8, MATS1pe, nRot, n5ARing, C2SP3, SMR_VSA9, AATSC1c, BCUTc-1h, EState_VSA1, GATS2m, nBridgehead, NsCH3, SssscC, VSA_EState3, EState_VSA5, EState_VSA9, n8FRing, n9FRing, TSRW10, ATSC1v, VE1_A, ATS3dv, EState_VSA2, C3SP3, Radius, ATSC2v, PEOE_VSA7, SsssN, ATSC5dv, PEOE_VSA5, ATSC2dv, ATSC5d, SMR_VSA5, n7FRing, NsNH2, ATSC3v, GATS2pe, SMR_VSA2, ETA_dEpsilon_B, ZMIC0, SMR_VSA4, ZMIC5, WPol, n6ARing, ZMIC1, EState_VSA6, JGT10, GATS1pe, PEOE_VSA8, ATS4d, ATSC7d, bpol, PEOE_VSA10, NdO, BCUTi-1h, JGI2, SlogP_VSA1, nN, C1SP2, ATSC4dv, VSA_EState7, EState_VSA7, VR2_A, SMR, BCUTse-1l, GGI4, SLogP, GGI5, nSpiro, SssO, SdsCH, ATSC4d, ETA_beta_ns_d, nBondsKD, AATS2dv, NdssC, BCUTp-1l, CIC5, VSA_EState9, BIC2, PEOE_VSA12, PEOE_VSA13, ATSC8i, ATSC3dv, SdssC, GGI6, SlogP_VSA10, ATSC7i, JGI4, VSA_EState4, EState_VSA3, PEOE_VSA11, JGI3, NaasC, SlogP_VSA11, nBondsD, RPCG, ATSC3pe, ATSC4are, C1SP3, nBase, n3HRing, n6FRing, SMR_VSA7, ATSC2i, ATSC3c, n7ARing, VSA_EState6, SdNH, BCUTc-1l, NssNH, GGI7, NaaNH, NaaO, PEOE_VSA3, SlogP_VSA4, n5FAHRing, SlogP_VSA2, ATSC8c, ATSC6se, PEOE_VSA9, ECIIndex, ATSC4c, ATS8dv, EState_VSA8, C4SP3, NdsN, ETA_epsilon_5, ATSC5c, VSA_EState8, ATSC6d, C2SP2, nF, SMR_VSA3, SlogP_VSA5, AATSC0i, BCUTdv-1h, NaaN, n8Ring, FilterItLogS, SlogP_VSA3, ATSC7c, SlogP_VSA8, ATSC6c, PEOE_VSA2, NaasN, SMR_VSA6, PEOE_VSA6, n6ARing
SCorr	AMID, AETA_eta_R, SpMAD_D, VE2_D, ATS4Z, SpMAD_A, SsssCH, ATS5pe, ETA_dEpsilon_A, BCUTd-1l, SsssCH2, ATS4v, MPC8, nAHRing, EState_VSA4, RNCG, nRot, MATS1pe, BIC1, BCUTc-1h, NsssCH, SMR_VSA9, AATSC1c, C2SP3, MIC1, EState_VSA1, n5ARing, EState_VSA9, GATS2m, NsCH3, JGI1, SssscC, ZMIC5, ETA_shape_y, nBridgehead, n8FRing, VE1_D, VE1_A, EState_VSA2, ATSC6dv, ATSC1v, VSA_EState3, EState_VSA5, C2SP1, ZMIC0, VR2_A, PEOE_VSA1, n7FRing, n9FAHRing, ATS3dv, Radius, ATSC8m, ATSC2v, VSA_EState7, C3SP3, ATSC5dv, Xp-1d, NsOH, ATSC2dv, SsssN, NsNH2, ATSC3m, GATS2pe, EState_VSA6, SMR_VSA2, n4AHRing, BCUTv-1l, ATS4d, SssO, ATSC7i, CIC5, EState_VSA7, GATS1pe, BCUTi-1h, ETA_dEpsilon_B, n6ARing, SlogP_VSA1, nN, SMR_VSA4, C1SP2, bpol, SMR_VSA5, ZMIC1, nH, WPol, PEOE_VSA10, ATS7dv, NdO, JGI2, ATSC4dv, PEOE_VSA7, nSpiro, GGI5, ETA_beta_ns_d, SMR, GGI4, VSA_EState9, PEOE_VSA8, SLogP, PEOE_VSA13, NdssC, JGI4, EState_VSA10, GGI6, VSA_EState5, PEOE_VSA12, ATSC4d, BIC2, SlogP_VSA10, SdssC, nBase, ATSC3dv, SlogP_VSA11, ATSC3pe, JGI3, ATS4are, MATS2c, AMID_N, PEOE_VSA11, n6FRing, nBondsD, ATSC8dv, C1SP3, VSA_EState6, RPCG, NaasC, n3HRing, EState_VSA3, ZMIC2, BCUTse-1l, ATSC7dv, SMR_VSA7, ATSC3c, Xc-6dv, n7Ring, SdsCH, SdNH, ETA_dPsi_B, GGI7, n5FAHRing, NsNH, SlogP_VSA3, SlogP_VSA2, NaaO, ATSC6d, NaaNH, SlogP_VSA4, ATS8dv, ATSC6m, EState_VSA8, ATSC5c, PEOE_VSA3, C4SP3, SlogP_VSA6, VSA_EState4, NdsN, ATSC4c, NaaN, PEOE_VSA9, NsF, SMR_VSA6, n8ARing, FilterItLogS, SlogP_VSA8, C2SP2, VSA_EState8, ECIIndex, ATSC6c, PEOE_VSA6, PEOE_VSA2
RF	ETA_shape_p, ETA_eta_R, nFRing, VE2_A, BalabanJ, BCUTd-1l, ATS5se, GGI4, SpAbs_DzZ, ABC, ATS5d, BCUTZ-1l, ETA_dEpsilon_B, ATS5dv, Xch-3d, GATS1se, C2SP1, GGI5, VE1_A, NsssCH, SpAbs_A, JGI4, IC1, BCUTd-1h, n7FRing, VR1_A, SpAbs_Dzv, ATS4dv, SMR, n6FRing, ZMIC0, EState_VSA9, ATSC1se, ATS4d, BertzCT, n5FRing, nBondsT, ATS6d, nHRing, n5Ring, SssscC, GGI3, ECIIndex, ATS3dv, ATS7d, BCUTi-1h, RPCG, piPC8, ETA_epsilon_5, ATSC1v, ATS0dv, FilterItLogS, BCUTdv-1h, VSA_EState4, ATS4se, RNCG, ATSC2d, BCUTc-1h, nBridgehead, ETA_beta_ns_d, Xc-5d, n9FRing, BCUTc-1l, n8FRing, BCUTZ-1h, EState_VSA5, nRot, GGI2, EState_VSA1, ATSC2Z, GGI6, ATSC4dv, ATSC4d, ATSC3Z, NsOH, EState_VSA8, SdssC, ATSC6c, VSA_EState7, ATS1dv, ATSC4c, GATS2c, NdsN, ATSC1dv, ATSC2v, ATSC2i, SLogP, ATS6dv, BCUTdv-1l, ATSC3dv, ATSC5c, ATSC8Z, SMR_VSA4, ATSC4se, ATSC2se, ATSC7c, ATSC3c, ATSC5d, ZMIC1, VSA_EState6, VSA_EState8, SlogP_VSA3, SlogP_VSA2, SMR_VSA7, ATSC5dv, SMR_VSA3, nSpiro, EState_VSA4, ATSC7d, SlogP_VSA4, ATSC2c, EState_VSA2, ZMIC2, ATSC2dv, ATSC3se, ATSC6d, ATSC5se, EState_VSA3, SMR_VSA5, ATSC6dv, NssN, PEOE_VSA8, PEOE_VSA9, SlogP_VSA10, IC2, n6Ring, n3HRing, NsssCH2, NtCH, SlogP_VSA5, ATSC8c, PEOE_VSA7, nC, NdO, nBondsKD, ATS7dv, NaaNH, SlogP_VSA1, SMR_VSA6, PEOE_VSA10
CFS	ETA_shape_p, SpMax_D, AETA_eta_R, Xch-3d, VMcGowan, NsOH, ATS5p, ETA_shape_x, AETA_eta_F, SpMAD_Dzi, MATS2m, nHBDdon, SMR_VSA9, ATS5are, NdO, EState_VSA9, JGI3, ATS7p, ATS4p, BalabanJ, n3Ring, ETA_dEpsilon_A, BCUTp-1h, Mi, Vabc, n4Ring, ETA_dBeta, SsOH, NsCH3, SpMax_Dzm, SsNH2, Xch-3dv, ATSC2v, ATSC3d, EState_VSA10, StCH, VSA_EState3, SsCH3, LogEE_D, PEOE_VSA1, AMID_C, NdsN, JGI1, ATSC3m, ETA_dEpsilon_B, SdO, nBondsKS, NaaNH, NtN, SMR_VSA1, ETA_dEpsilon_C, AATS0i, AATSC0p, nHBAcc, CIC0, BIC0, ATSC2Z, SpMAD_Dzm, VSA_EState7, ETA_dAlpha_B, SM1_Dzp, ETA_epsilon_2, SM1_Dzv, AATS1i, JGI2, SpMAD_D, ATS2p, C1SP3, ATS4Z, GATS1Z, Xp-5d, ATS6are, ATSC5d, SlogP_VSA5, VSA_EState9, nO, StsC, ETA_eta_R, ATS5m, AATSC1p, GATS1i, AATSC0v, NsNH2, AMID_O, AATSC1se, BCUTc-1l, WPol, ATST7are, VE2_Dt, ATS4d, n5ARing, ATS5v, Xc-4d, mZagreb2, ETA_shape_y, VE2_Dzv, ATSC3m, Diameter, Xc-6d, AETA_dBeta, AATSC0c, Xch-4d, GATS1m, WPath, VSA_EState8, ATS4dv, ATSC2m, C2SP3, NtCH, BCUTse-1l, ATSC0v, ATSC0c, ATSC1c, ETA_dEpsilon_D, ETA_eta_F, ATSC0p, SpMAD_Dzp, nH, BertzCT, ATS4m, AETA_eta, AETA_eta_FL, AETA_eta_L, ETA_eta_FL, VE2_Dzare, MIC4, SlogP_VSA4, AATS2i, GGI1, ATS7m, C2SP2, nBondsS, C1SP2, VE2_Dzse, ATS4v
LISSO	VR1_A, ATS2dv, ATS3dv, ATS0m, ATS1m, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS8m, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS2se, ATS0i, ATS1i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC3dv, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC4m, ATSC5m, ATSC6m, ATSC8m, ATSC1v, ATSC3v, ATSC5v, ATSC6v, ATSC7v, ATSC8v, SpAbs_Dzp, BertzCT, SpAbs_Dt, SpDiam_Dt, SpAd_Dt, StCH, SdNH, StN, SaaN, ECIIndex, fragCpx, TIC1, TIC4, PEOE_VSA1, PEOE_VSA2, PEOE_VSA3, PEOE_VSA5, PEOE_VSA7, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, SMR_VSA3, SMR_VSA4, SMR_VSA5, SMR_VSA7, SMR_VSA9, SlogP_VSA2, SlogP_VSA4, SlogP_VSA5, SlogP_VSA6, EState_VSA2, EState_VSA4, EState_VSA5, EState_VSA8, VSA_EState3, VSA_EState4, VSA_EState7, MPC3, MPC7, TopoPSA(NO), Vabc, TSRW10, Zagreb2

A.2.6 Momento dipolar

Metodos	Atributos seleccionados
Pcorr	StN, ETA_eta_R, Vabc, ATS4Z, WPath, TopoPSA(NO), GATS2are, BCUTc-1h, SdO, ETA_dEpsilon_C, VE2_A, BCUTi-1h, MIC1, VE2_Dt, BCUTd-1l, GATS1se, ATS5dv, SpMAD_A, nHRRing, C1SP1, NdssC, nN, ETA_shape_p, MATS2se, MATS1c, PEOE_VSA12, EState_VSA2, VSA_EState7, ATS4are, ATS3dv, Radius, BCUTp-1l, VSA_EState5, WPol, ATSC6dv, BCUTdv-1h, ETA_dPsi_B, nBondsD, ATSC3dv, GGI4, EState_VSA5, ZMIC0, ATSC5dv, nBridgehead, RNCG, MATS1pe, n9FAHRRing, ATS5are, BCUTv-1h, SlogP_VSA1, nFRing, StCH, SsssN, ATSC1v, nRot, BalabanJ, ATSC7dv, VSA_EState3, NdsN, ATS6dv, NsssCH, n5Ring, TSRW10, EState_VSA4, NssNH, piPC8, ATSC3m, ECIndex, SsssC, ETA_dEpsilon_B, ETA_epsilon_5, SlogP_VSA3, EState_VSA8, GGI5, n8FRing, RPCG, VSA_EState6, SlogP_VSA2, SsCH2, BCUTc-1l, EState_VSA7, ZMIC1, SMR_VSA5, SlogP_VSA11, VSA_EState1, nBase, ZMIC5, C1SP2, C2SP3, n3AHRing, PEOE_VSA6, nBondsKD, NaasC, SlogP_VSA4, nSpiro, SsNH2, JGI3, EState_VSA1, ATSC6d, ATSC4m, C2SP2, ATS7m, ATSC7d, SlogP_VSA6, ATSC8dv, AETA_beta_ns_d, FilterItLogS, PEOE_VSA2, GATS1Z, SaaCH, TIC2, C3SP3, PEOE_VSA8, SLogP, AATS1dv, GATS2p, ATSC3c, GGI6, ATSC5d, ATSC7c, SlogP_VSA5, SaaO, ATS8m, SMR_VSA3, EState_VSA6, BCUTdv-1l, ATSC6i, ATSC6c, PEOE_VSA13, SdsCH, EState_VSA9, VSA_EState4, VR2_A, ATSC4c, PEOE_VSA3, ATSC8c, ATSC0i, PEOE_VSA7, GGI7, ATSC4d, ATS8dv, SlogP_VSA8, SMR_VSA6, VSA_EState8, ATSC3pe, n6FHRing, SaaNH, AATSC2i, n6aRing, JGI2, PEOE_VSA10, PEOE_VSA9, JGI4, SsF, n7FARing, n7HRRing, ATSC5c, ATS4d, SMR_VSA4, SdNH, n9ARing, SdssC, NaaN, SMR, n8ARing, SlogP_VSA10, n6Ring, C3SP2, C4SP3, EState_VSA3, nAcid, NaasN, VSA_EState9, n5FHRing, PEOE_VSA11
SCorr	BCUTd-1l, NtN, SpMax_Dzm, Vabc, GATS2are, ETA_eta_R, NdO, BCUTdv-1h, TopoPSA(NO), BCUTc-1h, ATS4dv, VE2_D, ETA_dEpsilon_C, GATS1se, ATS4v, nN, BCUTi-1h, nH, MIC1, MATS2se, PEOE_VSA12, SpMAD_A, NdssC, ATS4are, C1SP1, EState_VSA2, MATS1c, ATSC6dv, SsssCH, nHRRing, nBondsD, ATSC3dv, Radius, SlogP_VSA3, ATSC5dv, piPC4, GGI4, EState_VSA10, WPol, VSA_EState5, AATSC1pe, ATSC3dv, SlogP_VSA1, VE3_Dzp, LabuteASA, ZMIC5, ETA_dPsi_B, n9FAHRRing, EState_VSA5, VSA_EState7, ATSC7dv, BCUTv-1h, RNCG, StCH, nBridgehead, ZMIC0, nRot, ATS5are, NdsN, SsssN, ATSC1v, MATS2c, C1SP2, RPCG, nBase, ATSC3m, NssNH, MPC8, SRW09, VSA_EState1, SsssC, ETA_epsilon_5, VSA_EState3, n4AHRing, ETA_shape_y, GGI5, EState_VSA7, ATS6dv, n5Ring, VSA_EState6, ATSC6c, EState_VSA4, BIC1, ECIndex, ETA_dEpsilon_B, NsssCH, n8FRing, SlogP_VSA6, JGI1, SMR_VSA3, SlogP_VSA11, FilterItLogS, Xc-6dv, SlogP_VSA2, BCUTdv-1l, EState_VSA1, SsOH, ZMIC1, SssCH2, n3HRRing, C2SP3, GATS2p, VR2_A, SlogP_VSA5, nSpiro, SsNH2, BCUTc-1l, ATSC8dv, NaasC, SMR_VSA5, ATSC4m, AETA_beta_ns_d, SLogP, ATSC6d, EState_VSA8, C2SP2, GGI6, BIC2, ATS7dv, PEOE_VSA8, PEOE_VSA1, PEOE_VSA6, ATSC6i, ATSC3c, PEOE_VSA3, EState_VSA9, ATSC7d, EState_VSA6, ZMIC2, C3SP3, JGI3, SaaO, ATSC3pe, GATS1Z, PEOE_VSA10, PEOE_VSA13, SMR_VSA7, JGI4, SlogP_VSA4, ATS3v, SIC5, ATS8dv, bpol, EState_VSA3, VE1_A, ATSC4d, GGI7, VSA_EState4, SlogP_VSA8, NdsCH, VSA_EState8, ATSC4c, n6aRing, NaaN, SMR_VSA6, SsF, PEOE_VSA9, n6FAHRRing, PEOE_VSA11, n7HRRing, SdNH, ATSC8d, JGI2, ATS4d, n6AHRing, n9ARing, n7FRing, C4SP3, n5FARing, SdssC, VSA_EState9, nAcid, PEOE_VSA2, C3SP2, SMR, PEOE_VSA7, n8ARing, ETA_eta_RL, C1SP3, NaasN, SlogP_VSA10, SaaN, AMID_N, SMR_VSA4, ATSC5c
RF	BCUTd-1l, NtN, BCUTdv-1h, BCUTi-1h, NdO, SMR_VSA3, BCUTZ-1h, NtCH, BCUTc-1h, BCUTc-1l, RPCG, nBase, ATSC2c, RNCG, PEOE_VSA3, BCUTZ-1l, PEOE_VSA2, VSA_EState4, ATSC1se, GATS2c, ATS4dv, ATSC2se, VSA_EState6, ATSC4c, ETA_eta_R, ATSC3c, SlogP_VSA3, SlogP_VSA2, PEOE_VSA12, ATSC5c, SpAbs_Dzv, ATSC3se, ATSC2dv, PEOE_VSA11, ATSC4se, ATSC4dv, ATSC5se, ATS4se, BertzCT, ATSC6c, SpAbs_A, SLogP, ATS3dv, ATS5se, ATSC6dv, ATSC2i, SMR_VSA7, PEOE_VSA9, ATSC3dv, ATSC5dv, SpAbs_Dzv, SMR, ATSC3Z, SdssC, VSA_EState7, PEOE_VSA8, BCUTdv-1l, SlogP_VSA1, ATSC4d, ATS4d, GATS1se, VSA_EState8, ATSC2d, VR1_A, ATSC1dv, PEOE_VSA10, ZMIC1, VE1_A, FilterItLogS, ATSC5d, EState_VSA2, ZMIC2, BCUTd-1h, ATS0dv, SMR_VSA5, ATSC6d, ATSC2v, ATS6d, EState_VSA3, ATSC7dv, ATSC2Z, ATSC7c, ATS5d, BalabanJ, VE2_A, IC1, PEOE_VSA13, ATS5dv, GGI3, ATS1dv, ETA_dEpsilon_B, JGI4, ABC, EState_VSA4, GGI4, SsssC, ECIndex, Xc-5d, IC2, ETA_epsilon_5, nBondsT, ATSC1v, GGI2, EState_VSA8, ATSC7d, EState_VSA9, PEOE_VSA7, EState_VSA6, PEOE_VSA6, NaaN, NdsN, SMR_VSA6, EState_VSA1, Xch-3d, ATS7d, nHRRing, EState_VSA5, nBondsD, NdssC, NssNH, ATS6dv, ATSC0i, ETA_shape_p, n3HRRing, SlogP_VSA5, C2SP2, SlogP_VSA4, VSA_EState9, nN, C1SP3, EState_VSA7, NsssCH, GGI5, C1SP2, nRot, SMR_VSA4, C2SP3, piPC8, ZMIC0, NssCH2
CFS	StN, SdO, BCUTi-1h, NtN, MATS2c, ATSC3dv, SlogP_VSA3, ATS4i, SMR_VSA2, SMR_VSA3, NdssC, SlogP_VSA5, Diameter, nFRing, BCUTc-1h, SlogP_VSA4, PEOE_VSA1, AATSC1c, GATS1c, ATSC4are, ATSC5dv, NaaN, NdO, EState_VSA7, AETA_eta_L, EState_VSA6, EState_VSA9, ATSC3are, fMF, Vabc, VE2_Dt, MIC4, ATSC4dv, ATSC2dv, PEOE_VSA8, PEOE_VSA9, MATS2se, nH, VE3_Dzp, VE2_Dzare, WPath, ATSC5d, VSA_EState9, GATS2v, VSA_EState8, GGI4, ATSC4d, ATSC2d, NsssC, ATSC3m, ATSC4m, WPol, PEOE_VSA7, GATS1i, ATSC5se, ATSC6se, ATSC4pe, AATSC2dv, ATSC3c, ATSC4c, ATSC5c, BCUTdv-1h, ATSC1p, EState_VSA8, GATS1p, C3SP2, Xch-3dv, TopoPSA(NO), SM1_Dzv, PEOE_VSA12, SaasN, SpMAD_Dzi, ATS3pe, MATS2v, SpMAD_D, ATS3i, SaaNH, CIC1, MIC1, CIC2, SdsCH, TopoPSA, VSA_EState5, NsssN, ATSC6i, ATSC7i, ATSC5i, SpMAD_Dzv, VSA_A, VR2_A, AETA_eta_R, RNCG, SlogP_VSA1, EState_VSA2, CIC0, ETA_epsilon_5, AETA_beta, MATS1v, GATS1v, SIC3, CIC3, IC3, CIC4, SIC5, CIC5, IC5, BCUTv-1h, BCUTi-1l, SdssC, ZMIC5, VE2_Dzp, piPC1, ETA_dPsi_A, ETA_dEpsilon_C, EState_VSA1, GATS2m, ATSC1dv, ATSC4i, ATS5i, ATSC3se, ATS8Z, SsNH2, ATSC4Z, ATSC3Z, GGI3, SMR_VSA7, FCSP3, VMcGowan, ATS2i, ATS4v, VE2_Dzv, MIC2, VE2_D, VE3_Dt, ATSC8i, VE2_Dzse, VR1_Dzm, ATS5m, ATS4se, ATSC2c, GATS2c, VSA_EState3, SM1_Dzp, AATSC0dv, VE2_Dzi, VE1_Dzp, RPCG, SssNH, ATSC4se, PEOE_VSA3
LASSO	nBase, VR1_A, nAromAtom, ATS2dv, ATS3d, ATS1m, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS8m, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS3se, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, AATS2v, ATSC0dv, ATSC3dv, ATSC3d, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC8m, ATSC0v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC6v, ATSC7v, ATSC8v, ATSC1i, ATSC2i, ATSC3i, ATSC6i, BCUTv-1l, SpAbs_Dzv, SpDiam_Dzare, BertzCT, nBondsA, SpAbs_Dt, SpDiam_Dt, SpAD_Dt, SaaCH, StN, SdsN, SsOH, SdO, SaaO, ECIndex, fragCpx, TIC2, TIC5, ZMIC1, PEOE_VSA1, PEOE_VSA3, PEOE_VSA4, PEOE_VSA6, PEOE_VSA7, PEOE_VSA9, PEOE_VSA11, PEOE_VSA12, SMR_VSA3, SMR_VSA6, SMR_VSA7, SlogP_VSA2, SlogP_VSA4, SlogP_VSA6, SlogP_VSA11, EState_VSA1, EState_VSA2, EState_VSA3, EState_VSA4, EState_VSA6, EState_VSA8, EState_VSA9, VSA_EState5, VSA_EState8, MPC7, TMPC10, TopoPSA(NO), Vabc, TSRW10

A.2.7 Polarizabilidade isotrópica

Metódos	Atributos selecionados
Pcorr	SMR, apol, AMID_O, BCUTse-1l, ETA_dEpsilon_B, FCSP3, ETA_beta, VSA_EState5, StsC, EState_VSA8, GATS1p, ETA_epsilon_5, PEOE_VSA1, VSA_EState7, SMR_VSA7, BCUTse-1h, ATSP4, ECIndex, ATSC0i, ATSC1se, WPath, bpol, ATSP3are, SLogP, SlogP_VSA2, SlogP_VSA4, NssO, FilterItLogS, PEOE_VSA5, BCUTdv-1h, VE2_A, GATS2p, C2SP2, SlogP_VSA3, SsOH, VR1_A, ATSP4are, PEOE_VSA6, VSA_EState6, EState_VSA2, EState_VSA1, TIC1, C3SP2, SaaCH, MID_N, piPC8, SdssC, SaasC, PEOE_VSA9, nBondsKD, ZMIC0, EState_VSA7, EState_VSA10, ETA_eta_R, AETA_eta_RL, PEOE_VSA7, SdsCH, BCUTi-1h, SssssC, ATSC7d, Radius, GGI5, ATSC4d, ATSC1dv, n6aRing, ATSP3Z, PEOE_VSA8, ZMIC1, ATSC3c, SMR_VSA3, SMR_VSA6, GGI4, ATSC8i, GGI6, C1SP3, TIC2, EState_VSA9, NssCH, SMR_VSA5, ATSC2d, PEOE_VSA10, ATSC2se, Xc-3d, nRot, SaaNH, RNCG, SaasN, ATSC2dv, BCUTc-1h, NssCH2, C2SP3, ZMIC2, ATSC5d, ATSP4d, ATSC7p, JGT10, ATSC2c, AATS1dv, ETA_beta_ns_d, StN, BalabanJ, EState_VSA5, EState_VSA6, SlogP_VSA8, SssN, Xc-5d, GGI7, EState_VSA3, C1SP2, TIC4, SaaN, ATSC4p, EState_VSA4, ATSC3m, JGI2, ATSC4se, VSA_EState4, SlogP_VSA5, NsF, MATS1Z, nBase, ATSP4m, n5AHRing, SssCH, JGI3, SsCH3, nAHRing, SlogP_VSA11, n6AHRing, NdsN, ATSC6d, PEOE_VSA2, ATSC4c, nBondsD, TSRW10, SaaO, C4SP3, ATSP5pe, ATSC6Z, n7FAHRing, BCUTd-1l, PEOE_VSA11, nFHRing, n5Ring, VSA_EState8, SMR_VSA4, n5FARing, ATSC6i, PEOE_VSA13, SlogP_VSA10, nBridgehead, RPCG, ATSC3pe, GATS1se, JGI4, ATSP8dv, n9FRRing, SlogP_VSA1, VSA_EState9, Xch-3dv, GATS2pe, ATSC3dv, n9AHRing, n3AHRing, n4AHRing, C3SP3, NdNH, BCUTc-1l, NdsC, PEOE_VSA12, ATSC7c, SssNH, ATSC6c, ATSC8c, n8AHRing, n7HRing, n6FHRing, ATSC5c, nSpiro, nAcid, PEOE_VSA3, NsNH2, n8FAHRing
Scorr	SMR, ATSC0i, BCUTse-1l, ETA_dEpsilon_B, ETA_beta, FCSP3, EState_VSA8, GATS1p, VSA_EState7, VSA_EState5, ATSP4, AATSC1se, SMR_VSA9, nH, C2SP1, PEOE_VSA1, SMR_VSA7, ETA_eta_RL, SlogP_VSA6, SlogP_VSA2, bpol, NssO, SlogP_VSA4, ECIndex, BCUTi-1h, ATSP3are, VE1_Dzp, BCUTm-1h, VE2_Dzare, FilterItLogS, WPath, nN, EState_VSA2, GATS2p, SsOH, C2SP2, VSA_EState6, PEOE_VSA6, PEOE_VSA9, VR1_A, SpAD_A, EState_VSA1, MID_N, SlogP_VSA3, TIC1, C3SP2, ATSP4se, SaasC, piPC8, VSA_EState4, ATSC2se, EState_VSA10, VE3_A, SdssC, ZMIC1, SpAbs_Dzare, ATSC3c, PEOE_VSA7, ATSC4d, SMR_VSA6, ATSC2c, AETA_eta_RL, ATSC1dv, BCUTd-1l, GGI5, RNCG, ETA_shape_x, SssssC, EState_VSA7, Radius, SdsCH, ATSP3Z, PEOE_VSA8, ETA_eta_R, BIC5, GGI4, C1SP3, VSA_EState3, NssCH, ATSP4d, ETA_beta_ns_d, n6aRing, JGI1, ZMIC2, ATSP8p, SaaN, ATSC2d, SaaNH, PEOE_VSA4, EState_VSA3, PEOE_VSA10, NdO, SMR_VSA5, TIC2, SaasN, SMR_VSA3, BCUTc-1h, C2SP3, GGI6, JGT10, ATSC6d, SssN, ATSP6dv, EState_VSA5, NssCH2, ZMIC4, ZMIC0, ATSC7i, EState_VSA6, JGI2, ATSC4se, SssCH, ETA_shape_y, BalabanJ, ATSC4p, nRot, SlogP_VSA11, EState_VSA4, PEOE_VSA2, SsCH3, nAHRing, C1SP2, ATSP7dv, Xc-5d, MIC1, ATSC4c, ATSC3m, n5AHRing, ATSC7dv, MATS1Z, SaaO, GATS1are, Xc-6d, SlogP_VSA5, BCUTc-1l, ATSP4m, ETA_dPsi_B, nBase, GGI7, ATSC8dv, SlogP_VSA8, n5Ring, GATS2se, n6AHRing, JGI3, EState_VSA9, NdsN, ATSC6se, C4SP3, ATSC6Z, SlogP_VSA10, n7FAHRing, nBridgehead, nFAHRing, ATSP5pe, ATSP8dv, RPCG, n4AHRing, PEOE_VSA11, NsF, ATSC6c, VSA_EState8, n9FAHRing, SMR_VSA4, nBondsD, PEOE_VSA13, ATSC6i, JGI4, n5FAHRing, C3SP3, VSA_EState9, n6FAHRing, nSpiro, PEOE_VSA12, n9ARing, NdNH, ATSC3se, n3AHRing, SlogP_VSA1, n7ARing, n8HRing, PEOE_VSA3, NdsC, SsNH2, nAcid, ATSC5c, SssNH, ATSC3dv, n8FARing
RF	SMR, ATSC0i, ETA_dEpsilon_B, BCUTZ-1l, nC, BertzCT, nO, C2SP1, ZMIC0, ABC, BCUTZ-1h, BCUTi-1h, ATSP6d, SdssC, SMR_VSA7, BCUTd-1l, nBondsKD, VR1_A, nN, VE2_A, ATSP0dv, SpAbs_DzZ, ATSC1se, RNCG, ATSP7d, BCUTd-1h, SLogP, VSA_EState4, RPCG, SpAbs_Dzv, VSA_EState7, SlogP_VSA2, SlogP_VSA4, GATS1se, ATSP5se, BCUTc-1h, ATSC6dv, EState_VSA8, VE1_A, ATSC6d, ETA_epsilon_5, ATSP5d, SpAbs_A, ECIndex, BCUTdv-1h, BCUTc-1l, ATSP4se, ETA_eta_R, PEOE_VSA6, PEOE_VSA7, FilterItLogS, BalabanJ, ATSC2d, ATSC7d, PEOE_VSA11, ZMIC1, ATSC1v, ATSC2i, n5Ring, ATSC5d, BCUTdv-1l, ATSP3dv, SssssC, ATSC2c, IC1, ATSC1dv, EState_VSA9, VSA_EState6, ATSP1dv, ATSC2dv, nSpiro, ATSC4d, ATSP5dv, ATSP4dv, ATSC2se, GATS2c, ATSP8d, NdsN, JGI4, GGI3, SMR_VSA3, PEOE_VSA9, nBondsD, ATSP4d, GGI4, Xch-3d, NssCH, EState_VSA4, GGI2, NtCH, nBase, EState_VSA5, ATSC5c, nFRRing, ATSC3c, PEOE_VSA12, ATSC3dv, ATSC3Z, ATSC5se, ATSC5dv, ATSC6c, ATSC4dv, ATSC2Z, ATSC4c, SlogP_VSA10, ATSC2v, PEOE_VSA8, ATSC4se, PEOE_VSA2, NdsC, ETA_shape_p, ATSC7dv, ATSC3se, ZMIC2, VSA_EState8, NdsCH, EState_VSA7, EState_VSA3, ATSP8Z, EState_VSA2, SMR_VSA5, Xc-5d, nHRing, ATSC7c, n6FRRing, SlogP_VSA1, SlogP_VSA5, nF, PEOE_VSA3, VSA_EState9, SMR_VSA4, IC2, piPC8, EState_VSA6, PEOE_VSA10, ATSP6dv, SMR_VSA6, C2SP2, SlogP_VSA3, ETA_beta_ns_d
CFS	SMR, Sp, SpMax_D, ETA_dEpsilon_C, LabuteASA, SMR_VSA7, C2SP1, BertzCT, SdssC, ATSP6p, apol, VMcGowan, Xch-4dv, nN, PEOE_VSA7, nBondsD, ETA_dEpsilon_B, PEOE_VSA8, StsC, n6aRing, nSpiro, AATSC1dv, ATSC3dv, AATS2v, ATSP8p, ATSP4p, ATSC7d, Mp, NdsN, VSA_EState4, ATSP0p, Diameter, piPC3, Vabc, BCUTc-1l, SaasC, nBridgehead, NdsCH, VSA_EState5, BCUTi-1l, ATSP4v, SssCH2, ATSC2v, NssNH, PEOE_VSA9, VE2_Dt, nO, Sv, ATSP6se, AETA_eta, SaasN, NdNH, NtsC, piPC6, NdsC, GATS1m, ETA_alpha, BCUTse-1l, nH, VE3_Dt, BCUTp-1h, GATS1v, ATSP0v, PEOE_VSA1
LASSO	VR1_A, nAromAtom, nSpiro, nBridgehead, ATSP1dv, ATSP4dv, ATSP7dv, ATSP1d, ATSP2d, ATSP3d, ATSP5d, ATSP6d, ATSP2Z, ATSP0m, ATSP1m, ATSP2m, ATSP3m, ATSP5m, ATSP6m, ATSP8m, ATSP0v, ATSP1v, ATSP2v, ATSP3v, ATSP5v, ATSP6v, ATSP7v, ATSP8v, ATSP2se, ATSP4se, ATSP0i, ATSP1i, ATSP2i, ATSP3i, ATSP4i, ATSP5i, ATSP6i, ATSP7i, ATSP8i, AATS2v, ATSC1dv, ATSC4dv, ATSC2d, ATSC3d, ATSC6d, ATSC0m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC0v, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC6v, ATSC7v, ATSC8v, ATSC1i, ATSC2i, ATSC5i, ATSC6i, BCUTv-1l, SpDiam_Dzv, SpDiam_Dzp, SpAD_D, SpDiam_Dzi, SpAD_Dzi, BertzCT, nBondsA, DetourIndex, SpDiam_D, SssCH2, StCH, SaaCH, SdssC, SsNH2, SssNH, StN, SdsN, SdO, SaaO, ECIndex, fragCpx, TIC0, TIC1, TIC2, TIC3, TIC4, PEOE_VSA2, PEOE_VSA3, PEOE_VSA4, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA12, SMR_VSA4, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA3, SlogP_VSA5, SlogP_VSA6, EState_VSA1, EState_VSA2, EState_VSA3, EState_VSA4, EState_VSA5, EState_VSA8, EState_VSA9, EState_VSA10, VSA_EState2, VSA_EState4, VSA_EState5, VSA_EState7, MPC5, MPC8, TMPC10, piPC6, n5ARing, n5AHRing, nRot, TopoPSA(NO), TMWC10, SRW09

A.2.8 Energia HOMO

Metodos	Atributos seleccionados
Pcorr	SpMAD_Dzi, piPC4, NaasC, AETA_beta_ns_d, NtN, ATS1i, SMR_VSA7, C1SP1, C2SP2, SaaNH, SlogP_VSA11, SMR, PEOE_VSA8, nBondsKD, ETA_epsilon_5, SlogP_VSA10, ATS4dv, SMR_VSA3, ATS2dv, n5HRRing, EState_VSA9, GATS1v, nHRRing, ATS5dv, VE1_A, SpMAD_Dzv, SlogP_VSA4, ETA_dEpsilon_C, SpMAD_A, SMR_VSA1, ETA_shape_y, C1SP2, MID_N, GATS1are, ATSC6dv, ETA_dEpsilon_B, GATS2v, piPC8, ATSC4d, nRot, ATSC3dv, SsssN, EState_VSA2, BalabanJ, EState_VSA8, EState_VSA1, NssO, SRW09, SaasN, C3SP2, ETA_shape_p, SlogP_VSA5, nFHRing, ATSC7dv, BCUTdv-1h, TIC1, ATSC1d, ATSC4v, JGI1, VSA_EState3, SsssC, ATS6dv, ATSC6d, ATS4d, VSA_EState9, AATSC2d, Ns-CH2, VE2_A, SaaO, BCUTpe-1h, BCUTd-1l, AATS1i, SlogP_VSA1, nBase, SsNH2, PEOE_VSA9, nHBDon, GATS1c, ATS4are, n5FRRing, AETA_eta, AETA_eta_R, SaaN, Xc-4d, SMR_VSA5, SlogP_VSA3, AATSC2m, ATS7m, n9FAHRRing, ATSC1i, Radius, TIC5, EState_VSA5, SdssC, BCUTp-1h, EState_VSA6, ZMIC5, GGI5, JGT10, SssNH, TIC2, C1SP3, GGI6, AATSC2dv, EState_VSA4, n4AHRing, ATSC1dv, PEOE_VSA2, ATSC7d, NdO, GATS2are, SlogP_VSA8, SdsCH, VSA_EState7, ATSC7c, ATSC3m, ATSC0i, ATSC1c, VSA_EState4, ATSC6c, SMR_VSA6, ATS3dv, SMR_VSA4, n6aRing, ZMIC1, C2SP3, FilterItLogS, RNCG, C4SP3, PEOE_VSA10, JGI2, ATSC5c, n7FAHRRing, GGI4, RPCG, nF, ATSC5d, BCUTc-1h, AATSC2c, ATSC3i, EState_VSA7, nBridgehead, ATSC3se, JGI3, ATSC4c, ATSC5dv, PEOE_VSA3, ATSC8dv, SlogP_VSA2, n6Ring, C3SP3, BCUTc-1l, n5FRRing, PEOE_VSA5, ATS5are, GGI7, ATSC3c, ATS8dv, ATSC8c, PEOE_VSA11, JGI4, VR1_A, SsCH3, n6FAHRRing, ZMIC0, PEOE_VSA6, ATSC8i, PEOE_VSA12, BCUTi-1h, ATSC5se, nBondsD, nAcid, nSpiro, ECIndex, NdNH, NdsN, EState_VSA3
SCorr	SpMAD_Dzi, piPC4, ATS1i, MID_N, NtN, BCUTd-1l, Vabc, EState_VSA9, C1SP1, ATS4dv, PEOE_VSA8, NaasC, GATS1v, AETA_beta_ns_d, SlogP_VSA6, ATS2dv, SpMAD_A, SpMAD_Dzv, SsssN, C2SP2, VE1_A, SMR_VSA7, ETA_dEpsilon_C, ETA_shape_y, ETA_epsilon_5, nBase, n5Ring, nO, SMR, nFHRing, EState_VSA2, MPC8, TopoPSA(NO), nRot, ATSC1d, nBondsKD, SlogP_VSA5, SaaNH, NssO, EState_VSA1, VE2_A, SsssCH, SlogP_VSA11, ATSC3dv, SMR_VSA3, SlogP_VSA4, ATS6dv, PEOE_VSA2, GATS2p, ATS4d, ATSC6dv, ATSC4d, AETA_eta_R, EState_VSA8, BCUTp-1h, SlogP_VSA10, JGI1, C3SP2, SsssC, ZMIC5, PEOE_VSA9, ETA_dEpsilon_B, SaaN, ATSC6d, n8FAHRRing, Radius, EState_VSA5, Xp-1d, GGI5, NssCH2, n4AHRing, SdssC, VSA_EState4, ATS4v, BCUTm-1h, n9FAHRRing, VSA_EState5, ATSC4v, SaasN, ATS4are, VSA_EState3, nN, BCUTc-1l, SssNH, ZMIC2, TIC5, ATSC3v, TIC1, GATS1are, GATS2c, C1SP2, ATS7dv, SaaO, AATSC1pe, GGI6, EState_VSA4, ATSC7dv, VSA_EState9, SlogP_VSA3, ATSC6c, BCUTdv-1l, ATS5are, PEOE_VSA10, VR2_A, C1SP3, VSA_EState7, TIC2, AATSC2d, n7FAHRRing, BCUTc-1h, SlogP_VSA2, PEOE_VSA3, JGI3, JGT10, VSA_EState2, SsNH2, nBridgehead, ATSC1dv, RNCG, ATSC5dv, GGI2, SMR_VSA5, ATS3dv, ATS8m, PEOE_VSA11, ZMIC1, AATSC2m, GATS1c, n6aRing, n6Ring, GGI4, SlogP_VSA1, NsOH, n6FAHRRing, C2SP1, SlogP_VSA8, JGI4, MATS2se, C4SP3, bpol, ATSC3se, ATS8dv, AATSC1c, ATSC5c, AATSC2dv, ATSC3c, GGI7, SMR_VSA4, RPCG, PEOE_VSA12, EState_VSA3, n3AHRing, SsF, nH, SdsCH, Xc-6dv, n5FRRing, nSpiro, EState_VSA6, SlogP, BCUTi-1h, MATS2c, NsssCH, EState_VSA10, ATSC5se, ATSC4c, SMR_VSA6, ATSC8are, nAcid, C3SP3, C2SP3, ETA_dPsi_B, VSA_EState6, NdNH, NdsN, ECIndex, PEOE_VSA7, PEOE_VSA1, VE2_Dzp, EState_VSA7, n7ARing, PEOE_VSA6, VSA_EState8, NdssC, FilterItLogS, PEOE_VSA13, nBondsD, ZMIC0
RF	ETA_beta_ns_d, nAromAtom, NtN, BCUTi-1h, SMR_VSA7, BertzCT, GATS1se, SMR, nBase, SMR_VSA3, ATSC0i, SlogP_VSA10, BCUTd-1l, BCUTZ-1h, EState_VSA9, BCUTdv-1h, VSA_EState6, BCUTc-1h, BCUTc-1l, GATS2c, ATSC1se, SpAbs_DzZ, PEOE_VSA3, nN, n3HRRing, RNCG, RPCG, nO, NaaN, SlogP, ABC, BCUTZ-1l, ETA_dEpsilon_B, VSA_EState4, nFRRing, ATSC2dv, ATSC2c, PEOE_VSA8, SlogP_VSA2, SlogP_VSA1, ETA_epsilon_5, ATS4dv, ATSC2se, ATSC1v, nBondsKD, ATSC3dv, Xch-3d, SpAbs_A, VSA_EState9, BCUTdv-1l, SdssC, nRot, VE1_A, GGI2, VE2_A, VSA_EState8, BCUTd-1h, ATSC2i, ATS1dv, VR1_A, PEOE_VSA2, nBridgehead, ATSC1dv, ATSC2d, ATSC3dv, ATS0dv, SpAbs_Dzv, SlogP_VSA11, VSA_EState7, ATSC4d, ATSC3c, FilterItLogS, ATS5se, IC1, ATSC4c, GGI3, ETA_eta_R, ATSC3se, PEOE_VSA11, ZMIC1, nHRRing, ATSC3Z, SMR_VSA5, ATS4se, nBondsT, PEOE_VSA9, ATSC4se, EState_VSA8, BalabanJ, ATSC5dv, ATSC5c, ATSC4dv, C2SP2, nC, C1SP2, ATS4d, ZMIC2, SlogP_VSA5, ATSC5se, ATSC5d, ATSC2v, ATSC2Z, ATSC6c, n5Ring, ATS5d, JGI4, nBondsD, PEOE_VSA7, PEOE_VSA6, PEOE_VSA10, EState_VSA3, SlogP_VSA4, SlogP_VSA3, ATSC6d, ZMIC0, EState_VSA4, ATSC6dv, NaaNH, Xc-5d, IC2, GGI4, SMR_VSA6, SsssC, EState_VSA2, n6aRing, EState_VSA7, ATSC6d, ECIndex, EState_VSA1, NdssC, ATSC5dv, EState_VSA6, EState_VSA5, ETA_shape_p, Nss-ns, SMR_VSA4, ATSC7c, NdsCH, PEOE_VSA12, piPC8
CFS	piPC4, AETA_beta_ns_d, PEOE_VSA8, SMR_VSA7, ATS1i, GATS1m, VSA_EState2, GATS2are, NaasC, NdssC, nBondsKS, ETA_dEpsilon_B, nHBDon, ATS2i, nBondsKD, nH, ATSC1d, SMR, PEOE_VSA1, ATSC3i, AMID_C, AATS2i, ATS2v, ETA_beta_ns_d, EState_VSA10, ETA_dEpsilon_C, SpMAD_Dzp, MATS1v, GATS1v, TpiPC10, GATS1Z, ATS4d, VSA_EState9, SpMAD_Dt, SaaNH, nBondsD, C2SP1, BCUTm-1l, BCUTp-1h, BCUTZ-1l, GATS1pe, RNCG, NdO, SaaO, PEOE_VSA9, CIC0, BIC1, VSA_EState8, FilterItLogS, SssCH2, SlogP_VSA11, SlogP_VSA1, ATSC1c, ATSC2m, NdsN, SMR_VSA1, ETA_psi_1, SpMAD_Dzv, AATSC1v, SaasN, SMR_VSA6, PEOE_VSA7, SdO, NaaNH, C2SP3, ATSC2v, GATS1are, ATSC0p, Si, AETA_eta_RL, ETA_shape_p, AMID, C2SP2, PEOE_VSA10, VSA_EState5, BCUTdv-1h, BCUTc-1l, GATS1i, MATS1p, NaaO, n6AHRing, C1SP1, VE2_Dzp
LASSO	ATS2dv, ATS2d, ATS3d, ATS1m, ATS2m, ATS3m, ATS4m, ATS6m, ATS7m, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS7v, ATS8v, ATS2se, ATS3se, ATS4se, ATS0i, ATS1i, ATS2i, ATS3i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC1d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC5m, ATSC6m, ATSC8m, ATSC0v, ATSC1v, ATSC3v, ATSC4v, ATSC5v, ATSC6v, ATSC7v, ATSC8v, ATSC1i, ATSC3i, ATSC4i, SpAD_Dzp, SpDiam_Dzi, SpAD_Dzi, BertzCT, DetourIndex, SaaN, TIC1, PEOE_VSA1, PEOE_VSA3, PEOE_VSA5, PEOE_VSA6, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA12, SMR_VSA3, SMR_VSA4, SMR_VSA5, SMR_VSA7, SlogP_VSA2, SlogP_VSA4, SlogP_VSA5, SlogP_VSA10, SlogP_VSA11, EState_VSA2, EState_VSA3, EState_VSA9, VSA_EState2, TMPC10, TopoPSA(NO), TSRW10

A.2.9 Diferença de energias do HOMO e LUMO

Metódos	Atributos selecionados
Pcorr	nBondsKD, nBondsD, FCSP3, VSA_EState2, SMR, C2SP2, TIC1, ETA_eta_FL, PEOE_VSA2, C1SP2, NdsCH, NdsCH, SMR_VSA7, SssO, ETA_eta_R, PEOE_VSA9, EState_VSA9, ZMIC1, VE2_Dzi, NssCH, ATSC4v, SMR_VSA5, WPol, TIC2, Xch-4dv, GATS2Z, C3SP3, GATS1se, C2SP3, ATSC2d, MID_N, SMR_VSA3, SlogP_VSA4, ETA_shape_x, ATS5pe, SdssC, ATSC4d, C3SP2, ETA_dEpsilon_B, ATSC3dv, GATS2v, GATS2pe, EState_VSA7, ZMIC0, ATSC1v, n4Ring, SMR_VSA4, VE2_A, PEOE_VSA11, BCUTi-1h, C1SP1, ATSC3i, AATS2dv, ATSC2dv, SssCH2, AATSC1i, EState_VSA6, VSA_EState3, BCUTc-1h, ATSC5dv, FilterItLogS, StN, ATSC4p, PEOE_VSA8, Radius, RNCG, ATSC3Z, JGI3, AETA_beta_ns_d, ATSC4dv, nFARing, nBridgehead, AATS1i, ZMIC2, ATSC3pe, BCUTd-1l, PEOE_VSA1, GGI4, NssN, BCUTdv-1l, SlogP_VSA1, EState_VSA8, PEOE_VSA12, nBase, ATSC3i, PEOE_VSA6, n6aRing, TSRW10, AMID_O, nAHRing, ETA_shape_p, nSpiro, nRot, n5FRing, VSA_EState9, C4SP3, PEOE_VSA10, GATS1Z, SaaNH, SaaN, SlogP_VSA3, nX, SpMAD_Dzi, JGT10, VSA_EState5, ATSC1se, NssNH, PEOE_VSA5, AATS2v, BalabanJ, SaaO, ECIndex, SlogP_VSA11, NdsN, ATSC4d, VR2_A, ATSC5d, EState_VSA4, ATSC3dv, PEOE_VSA13, ATSC7d, NaasN, ATSC7p, VSA_EState7, ATSC6p, BCUTdv-1h, JGI2, VSA_EState6, n3HRing, SlogP_VSA8, ATSC7dv, VSA_EState4, n6FARing, SsNH2, ATSC6d, n7ARing, ATSC4i, SlogP_VSA5, ATSC0i, n5AHRing, SdNH, ATSC8p, GGI5, n6Ring, EState_VSA1, ATSC5c, PEOE_VSA7, ETA_dPsi_B, piPC8, SlogP_VSA10, ATSC6se, n9FRing, BCUTpe-1l, GATS1c, EState_VSA5, RPCG, BCUTp-1l, EState_VSA3, VSA_EState8, n8ARing, nAcid, n7FARing, ATSC4c, JGI4, AATSC2c, SlogP_VSA2, SssssC, BCUTc-1l, ATSC7c, SMR_VSA6, ATSC8c, ATSC3c, n9ARing, ATSC8dv, C1SP3, SlogP, GGI6, GGI7
Scorr	nBondsKD, nBondsD, VSA_EState2, C2SP2, SMR, C1SP2, NdsCH, TIC1, SssO, NdsCH, SlogP_VSA6, SMR_VSA7, PEOE_VSA2, MIC1, PEOE_VSA9, ETA_eta_R, EState_VSA9, BCUTdv-1h, ZMIC1, NssNH, GATS1se, SlogP_VSA3, ATSC4v, MID_N, nH, EState_VSA10, VE2_Dzi, Xch-4dv, SMR_VSA5, GATS2Z, WPol, ATSC2d, SdssC, C3SP3, ETA_shape_x, TIC2, C2SP3, ATSC4Z, GATS2se, BCUTc-1h, C3SP2, nN, SlogP_VSA4, ATSC4d, ETA_dEpsilon_B, C1SP1, ATSC2dv, n4Ring, NaasC, ATSC5pe, MATS1v, piPC1, ATSC2v, StN, SssCH, ATSC3dv, VE3_Dzp, SMR_VSA3, ATSC5se, AMID_O, EState_VSA7, SMR_VSA4, AATSC1i, VE3_A, LabuteASA, ZMIC0, SsOH, ETA_beta_ns_d, VE3_Dt, PEOE_VSA11, AATS2dv, MW, EState_VSA6, FilterItLogS, ATSC4p, Xc-6dv, VSA_EState3, PEOE_VSA8, SssCH2, PEOE_VSA10, BCUTc-1l, Radius, ATSC4dv, nBridgehead, ATSC3Z, NsCH3, RNCG, ATSC3pe, BCUTd-1l, VR2_A, JGI3, NssN, GGI4, nBase, PEOE_VSA6, n6aRing, Xp-1d, NaaN, nRot, SpMAD_Dzi, ZMIC2, ATSC3i, JGT10, SlogP_VSA1, BCUTi-1h, SaaNH, C4SP3, nSpiro, PEOE_VSA12, n5FARing, NssNH, SaaO, SlogP_VSA11, VSA_EState9, nAHRing, StCH, ETA_shape_p, n4HRing, SpMAD_A, ATSC6p, ATSC4d, nHBAcc, VSA_EState6, bpol, VSA_EState4, EState_VSA4, NdsN, PEOE_VSA1, BCUTpe-1l, ATSC6d, VSA_EState5, PEOE_VSA13, EState_VSA8, SaasN, ZMIC5, VSA_EState7, n6Ring, MATS1se, GATS1Z, AATS2v, RPCG, JGI2, n7ARing, ATSC7p, SpMAD_Dzp, ATSC5c, ATSC3dv, SlogP_VSA8, SlogP_VSA10, SsNH2, n6FRing, TIC3, n5Ring, ECIndex, ATSC6se, ATSC4i, GGI5, GATS1c, SdNH, SsF, n9FRing, SlogP_VSA5, n3HRing, ETA_dPsi_B, EState_VSA1, ATSC4c, ETA_epsilon_5, VSA_EState8, ATSC7dv, piPC8, ATSC8v, EState_VSA3, ATSC6dv, nAcid, JGI4, n8ARing, SssssC, SlogP_VSA2, n7FRing, MATS2c, SMR_VSA6, ATSC6c, ETA_shape_y, EState_VSA2, PEOE_VSA3, ATSC8dv, GGI7, n9HRing, ATSC3c, EState_VSA5, ATSC7dv, PEOE_VSA7, C1SP3, SlogP, n8FARing, GGI6
RF	nBondsKD, nBondsD, SMR_VSA7, SMR, C2SP2, BCUTZ-1h, ETA_dEpsilon_B, C1SP2, BCUTi-1h, nN, BCUTdv-1h, nBase, BCUTc-1h, NdsCH, BCUTc-1l, NdO, BCUTd-1l, ATSC2c, PEOE_VSA2, ABC, BCUTZ-1l, ATSC2se, PEOE_VSA12, RNCG, SlogP, RPCG, SlogP_VSA10, nBondsT, GATS1se, SMR_VSA3, EState_VSA8, ATSC2dv, GATS2c, ATSC1dv, ATSC3dv, IC1, EState_VSA9, n3HRing, VSA_EState6, SdssC, nAromAtom, ETA_eta_R, PEOE_VSA3, ATSC1se, ATSC3se, BCUTd-1h, SlogP_VSA2, ATSC0i, BCUTdv-1l, PEOE_VSA6, SlogP_VSA3, NdsCH, VE2_A, BertzCT, PEOE_VSA11, NssN, ATSC3c, PEOE_VSA10, ZMIC1, SpAbs_A, VSA_EState4, ATSC1v, ATSC1dv, VE1_A, nBridgehead, PEOE_VSA13, ATSC4d, GGI3, SlogP_VSA1, NssO, JGI4, PEOE_VSA8, GGI2, ATSC3dv, ATSC4dv, PEOE_VSA9, SpAbs_DzZ, VR1_A, ZMIC0, Xch-3d, SpAbs_Dzv, VSA_EState7, FilterItLogS, VSA_EState9, BalabanJ, ATSC3Z, ATSC2i, ATSC4c, ATSC4dv, VSA_EState8, ATSC0dv, ATSC5dv, ETA_shape_p, ATSC6dv, ATSC2d, ATSC5se, ATSC4d, SlogP_VSA4, ATSC4se, ATSC5d, PEOE_VSA7, ATSC5c, ATSC4se, C2SP1, ATSC2Z, ZMIC2, ATSC5se, Xc-5d, ATSC2v, n5Ring, ATSC5d, ATSC6d, EState_VSA4, ATSC6c, ETA_epsilon_5, NaaN, nHRing, SMR_VSA5, GGI4, ATSC6d, EState_VSA3, ECIndex, ATSC5dv, SlogP_VSA5, SssssC, nC, n7FRing, IC2, SMR_VSA6, EState_VSA7, EState_VSA2, n6aRing, ETA_beta_ns_d, ATSC7dv, NdsN, EState_VSA6, SMR_VSA4, EState_VSA5, ATSC7c, piPC8
CFS	FCSP3, EState_VSA9, CIC1, GATS1pe, VSA_EState1, C3SP3, PEOE_VSA5, AATSC2c, PEOE_VSA9, VE2_Dzse, VSA_EState3, ATSC1Z, BCUTc-1h, C2SP3, SaaCH, NtsC, ATSC4d, Mi, SssCH, NssO, SsOH, ATSC2se, GATS2Z, ETA_shape_x, SaasC, nBondsT, VE2_A, PEOE_VSA12, NssCH, GATS1v, BCUTd-1h, SpDiam_Dzare, ATSC6dv, ATSC5pe, RPCG, ZMIC1, SssO
LISSO	VR1_A, ATSC2dv, ATSC5dv, ATSC1d, ATSC2d, ATSC3d, ATSC4d, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC6m, ATSC7m, ATSC8m, ATSC9m, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC6v, ATSC7v, ATSC8v, ATSC9v, ATSC1i, ATSC2i, ATSC3i, ATSC4i, ATSC5i, ATSC6i, ATSC7i, ATSC8i, ATSC9i, ATSC2dv, ATSC4d, ATSC1m, ATSC2m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC1v, ATSC2v, ATSC3v, ATSC5v, ATSC6v, ATSC7v, ATSC8v, ATSC9v, ATSC0i, ATSC2i, ATSC3i, ATSC4i, SpDiam_Dzi, BertzCT, SpAbs_Dt, SpAD_Dt, SaaCH, StN, SdO, SssO, ECIndex, fragCpx, nHBAcc, TIC0, TIC1, TIC2, TIC3, PEOE_VSA1, PEOE_VSA3, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, SMR_VSA4, SMR_VSA6, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA3, SlogP_VSA6, SlogP_VSA11, EState_VSA1, EState_VSA3, EState_VSA4, EState_VSA5, EState_VSA6, EState_VSA9, VSA_EState2, VSA_EState3, VSA_EState4, MPC5, MPC7, MPC8, TMPC10, TopoPSA(NO), Vabc, TSRW10

A.3 Subconjunto com moléculas de 19 átomos

A.3.1 Energia interna de atomização para a temperatura de 0 K

Metódos	Atributos selecionados
Pcorr	ATS0v, SlogP_VSA2, VSA_EState7, TpiPC10, BCUTv-1h, SMR, PEOE_VSA7, SMR_VSA6, SlogP_VSA4, BCUTi-1h, C2SP2, VSA_EState5, SMR_VSA7, SdsCH, AATSC1se, ETA_dEpsilon_B, GATS1se, PEOE_VSA6, piPC8, GATS1c, SMR_VSA4, C1SP3, SlogP_VSA5, ATS5are, nRot, FilterItLogS, RNCG, nBase, ZMIC0, SlogP_VSA1, n5Ring, C3SP3, ATS4p, ETA_beta, nBondsKD, EState_VSA9, ATS5d, nBondsD, ETA_dPsi_B, GATS1m, VSA_EState9, NssNH, AATSC2se, C3SP2, ATSC3c, NsssN, SdssC, Vabc, ATS4Z, EState_VSA4, EState_VSA1, n3HRRing, PEOE_VSA9, ATSC3i, SMR_VSA9, StCH, nFARing, EState_VSA5, TIC5, n4AHRing, SssCH2, C4SP3, nAHRing, GATS2are, NdsN, EState_VSA2, n9FARing, AATSC0i, GATS2i, ZMIC5, WPol, NsCH3, BalabanJ, n8FARing, SssssC, SRW09, ATSC3se, BCUTc-1h, ATS4d, n6Ring, PEOE_VSA8, ETA_eta_R, VR1_A, VSA_EState8, NsNH2, PEOE_VSA12, n5FHRing, ATSC5d, RPCG, NdssC, NsssCH, EState_VSA8, ATSC3dv, ECIndex, NssO, PEOE_VSA3, BCUTd-1l, nBridgehead, ATSC7d, n7FARing, ATS3dv, n6aRing, JGI2, PEOE_VSA2, SaaCH, ATSC4d, C2SP3, n5FRing, PEOE_VSA10, VSA_EState4, EState_VSA6, ATSC2p, ATSC7Z, SdO, BCUTdv-1h, ATSC4c, SlogP_VSA10, PEOE_VSA11, n7ARing, NdNH, PEOE_VSA4, EState_VSA7, ATSC5c, SaasC, BIC2, MIC1, ATSC6d, ATSC8c, GGI5, n5AHRing, n6FAHRing, VSA_EState6, ATSC8pe, Xc-6d, MID_X, ATSC1dv, PEOE_VSA13, AATSC2d, ATS4are, JGI4, ZMIC1, ATSC4are, ATS7dv, ATSC6i, GGI4, ATSC7c, NaaN, Xch-3d, nAcid, GGI6, n6FARing, MIC5, EState_VSA10, ATSC6dv, ATSC6c, n9ARing, SMR_VSA5, ETA_beta_ns_d, SaaO, JGI3, BCUTc-1l, nSpiro, ATSC4i
Scorr	ATS0v, SlogP, LabuteASA, VSA_EState7, TpiPC10, VE1_Dzv, SMR, RNCG, VSA_EState5, SMR_VSA6, PEOE_VSA7, SlogP_VSA4, VE1_Dt, C2SP2, GATS1c, GATS1se, ETA_dEpsilon_B, nN, FilterItLogS, SMR_VSA7, ATS5are, nRot, nBase, PEOE_VSA6, C1SP3, nH, SMR_VSA4, nBondsM, SlogP_VSA5, SlogP_VSA1, piPC8, n5ARing, SdsCH, AATSC2se, ETA_eta_RL, AATSC2c, ZMIC5, ATSC3c, NsNH, ATS5d, C3SP3, SssCH, ATS4p, EState_VSA9, VSA_EState9, SssN, nBondsKD, ATS4m, BCUTc-1h, C3SP2, nBondsD, Vabc, C2SP1, SdssC, ATSC3i, AATS1Z, n3AHRing, SssCH2, MIC4, EState_VSA1, PEOE_VSA9, MATS1Z, C4SP3, EState_VSA4, ETA_dPsi_B, TIC5, GATS2are, nHRRing, VE2_Dzv, PEOE_VSA8, n4HRRing, EState_VSA2, AATSC2dv, NdsN, GATS2i, EState_VSA5, ATSC3se, VSA_EState6, n9FARing, AATS1dv, GATS1m, ATS3dv, n8FRing, VR1_A, n6Ring, PEOE_VSA12, NsCH3, NsNH2, RPCG, SssssC, ATS4d, PEOE_VSA3, ETA_shape_x, ATSC3dv, EState_VSA8, n5FAHRing, JGI1, SdO, EState_VSA6, PEOE_VSA2, NdssC, nBridgehead, BalabanJ, ATSC7v, WPol, n7FRing, NssO, VSA_EState8, NsssCH, C2SP3, ATS6dv, JGI2, ATSC4c, ATSC4d, MIC1, n6aRing, PEOE_VSA11, BCUTv-1l, n5FRing, VE2_A, ATSC2p, ZMIC2, BIC2, VSA_EState4, ATSC8i, SlogP_VSA10, NdNH, ECIndex, SMR_VSA5, ATSC6d, PEOE_VSA10, ATSC6c, ATSC5c, SMR_VSA2, GGI5, n7ARing, BIC1, n6FAHRing, n5AHRing, PEOE_VSA13, SaaCH, EState_VSA7, EState_VSA3, ATSC8Z, ATSC4are, ATSC6dv, ZMIC0, GGI4, ATSC1dv, ATS4i, GGI6, JGI4, AATSC2d, nAcid, ATSC6i, SsF, ATS7dv, ZMIC1, n6FARing, ETA_beta_ns_d, GGI3, n9Ring, nSpiro, Radius, ATSC4i, C1SP2, EState_VSA10, NaaN, Xc-6d
RF	nC, SlogP_VSA2, ZMIC0, VSA_EState7, nBondsKD, ATSC0i, ETA_eta_R, BCUTZ-1h, C2SP2, ATS0dv, SMR_VSA7, BCUTi-1h, BertzCT, n5Ring, ATSC2se, GATS1se, BCUTdv-1h, SpAbs_Dzv, ATSC1se, nBondsD, BCUTZ-1l, SpAbs_A, RNCG, nO, ABC, nN, VE1_A, VR1_A, nBase, nBridgehead, ATSC2c, VE2_A, BCUTd-1l, n4HRRing, BCUTc-1h, RPCG, n6Ring, nRot, SpAbs_DzZ, SlogP_VSA4, Xc-5d, BCUTc-1l, n5FRing, ATS4v, nSpiro, ATS5d, ATSC1dv, BCUTd-1h, ATS5se, SlogP_VSA1, NdsCH, Xch-3d, GGI3, ATSC2dv, MIC2, PEOE_VSA9, SssssC, n5HRRing, ATS3dv, VSA_EState4, BalabanJ, GATS2c, NdO, VSA_EState5, EState_VSA9, n8FRing, VSA_EState6, ETA_dEpsilon_B, ATSC2d, SlogP_VSA5, n5FHRing, FilterItLogS, n7FRing, SdssC, ZMIC1, ATS3Z, NtN, nAromAtom, GGI4, IC1, JGI4, nHRRing, EState_VSA8, NdssC, NssNH, n6aRing, ATS4d, C1SP2, ATSC2Z, PEOE_VSA2, EState_VSA10, BCUTdv-1l, n6FRing, ATS6d, ATSC4d, VSA_EState8, ATSC3dv, ATSC1dv, PEOE_VSA12, PEOE_VSA8, PEOE_VSA7, piPC8, ATSC3se, NsssN, ATS4dv, ZMIC2, ATSC3Z, ATS4se, ATSC3c, n3HRRing, NsssCH, SMR_VSA5, ATSC5d, IC3, PEOE_VSA3, PEOE_VSA6, ATSC6d, ETA_shape_p, SMR_VSA4, NssO, ECIndex, VSA_EState9, SMR_VSA6, PEOE_VSA10, C3SP2, ATSC4c, ATSC4dv, EState_VSA4, ATSC4se, ATSC5dv, SlogP_VSA10, ATSC5se, ATSC5c, ATSC6c, IC2, NssCH2, EState_VSA5, nBondsT, C2SP1, NsNH2
CFS	Mi, SlogP_VSA2, nH, AATS1se, n4Ring, GATS1m, Xch-3d, SpMAD_Dzp, Si, FCSP3, ETA_dEpsilon_B, n4ARing, GATS1v, nHBacc, VE2_Dzare, ABCGG, AATS1i, BCUTv-1h, BCUTm-1l, VSA_EState7, ETA_dAlpha_B, ATSC1c, GATS1Z, AATS1pe, AATS1p, AATS2i, AATSC1i, GATS1p, ATS0p, AATS0i, ATSC0p, ETA_dPsi_A, SM1_Dzv, ATS0i, AATSC0p, AMID_h, VSA_EState5, ETA_epsilon_4, VE2_D, BIC0, ATSC0v, ETA_dEpsilon_D, MATS1are, PEOE_VSA1, VSA_EState8, SpMAD_Dzare, ATS3p, VSA_EState9, GATS2p, ATSC2p, GATS2v, ATSC1dv, CIC0, ATSC0c, nBase, VE2_DzZ, AATS1are, n4HRRing, nRot, BCUTc-1l, AATS2se, SM1_Dzp, ETA_epsilon_2, AETA_alpha, SM1_Dzare, VE2_Dzse, ETA_alpha, ATSC5pe, AETA_eta, GATS1i, VSA_EState3, AATSC1se, AATSC0c, AATSC0v, ATSC4i, ATSC3i, C4SP3, SsNH2, AMID_O, VE2_Dt, SMR_VSA6, nHBDOn, AATSC1c, MIC0, nC, PEOE_VSA3, SMR_VSA5, mZagreb2, C3SP3, PEOE_VSA2, AMID_C, ETA_beta_ns_d, FilterItLogS
LISSO	VR1_A, nBridgehead, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS8m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC8m, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC6v, ATSC7v, ATSC1i, ATSC2i, SpAD_Dzp, BertzCT, SpDiam_Dt, DetourIndex, SaaCH, SssO, ECIndex, ETA_beta, fragCpx, TIC1, PEOE_VSA2, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, SMR_VSA2, SMR_VSA5, SMR_VSA7, SlogP_VSA1, SlogP_VSA2, SlogP_VSA6, EState_VSA2, EState_VSA4, EState_VSA5, EState_VSA6, EState_VSA8, VSA_EState4, VSA_EState6, VSA_EState7, MPC3, MPC4, MPC5, MPC8, TopoPSA(NO), SRW09

A.3.2 Energia interna de atomização para a temperatura de 298.15 K

Metódos	Atributos selecionados
Pcorr	ATS0v, SlogP_VSA2, VSA_EState7, TpiPC10, BCUTv-1h, SMR, PEOE_VSA7, SMR_VSA6, SlogP_VSA4, BCUTi-1h, C2SP2, VSA_EState5, SMR_VSA7, SdsCH, AATSC1se, ETA_dEpsilon_B, GATS1se, PEOE_VSA6, piPC8, SMR_VSA4, GATS1c, C1SP3, ATS5are, nRot, SlogP_VSA5, FilterItLogS, RNCG, nBase, SlogP_VSA1, ZMIC0, n5ARing, C3SP3, ATS4p, ATS5d, EState_VSA9, nBondsKD, ETA_beta, nBondsD, GATS1m, ETA_dPsi_B, VSA_EState9, NssNH, AATSC2se, C3SP2, ATSC3c, SdssC, ATS4Z, NssN, EState_VSA4, EState_VSA1, nFRing, ATSC3i, Vabc, n3HRing, PEOE_VSA9, EState_VSA5, StCH, C2SP1, TIC5, SssCH2, C4SP3, n4HRing, GATS2are, nAHRing, NdsN, EState_VSA2, n9FRing, AATSC0i, BalabanJ, ZMIC5, NsCH3, GATS2i, WPol, n8FAHRing, TSRW10, SssssC, BCUTc-1h, ATS4d, ATSC3se, AATSC2c, n6Ring, PEOE_VSA8, VR1_A, VSA_EState8, NsNH2, PEOE_VSA12, ATSC5d, ETA_eta_R, NssCH, n5FAHRing, ATSC3dv, BCUTd-1l, RPCG, nBridgehead, EState_VSA8, NdssC, PEOE_VSA3, EIndex, ATSC7d, NssO, n7FRing, ATS3dv, n6aRing, JGI2, C2SP3, PEOE_VSA2, ATSC4d, PEOE_VSA10, SaaCH, n5FRing, EState_VSA6, VSA_EState4, ATSC2p, ATSC7Z, BCUTdv-1h, SdO, ATSC4c, PEOE_VSA11, SlogP_VSA10, n7Ring, PEOE_VSA4, NdNH, MIC1, ATSC5c, BIC2, GGI5, EState_VSA7, n5AHRing, SaaCH, ATSC8c, ATSC6d, VSA_EState6, n6FAHRing, ATSC8pe, MID_X, PEOE_VSA13, Xc-6d, ZMIC1, JGI4, ATSC1dv, AATSC2d, ATS7dv, ATS4are, GGI4, ATSC4are, ATSC6i, NaaN, ATSC7c, GGI6, n6FRing, nAcid, Xch-3d, SMR_VSA5, MIC2, ATSC6c, n9Ring, ETA_beta_ns_d, ATSC6dv, EState_VSA10, nSpiro, SaaO, JGI3
Scorr	ATS0v, SLogP, LabuteASA, VSA_EState7, TpiPC10, VE1_Dzv, SMR, RNCG, VSA_EState5, SMR_VSA6, PEOE_VSA7, SlogP_VSA4, VE1_Dt, C2SP2, GATS1c, GATS1se, ETA_dEpsilon_B, nN, nRot, ATS5are, SMR_VSA7, SMR_VSA4, C1SP3, FilterItLogS, nBase, nH, PEOE_VSA6, SlogP_VSA5, nBondsM, SlogP_VSA1, piPC8, n5ARing, SdsCH, AATSC2se, ETA_eta_RL, AATSC2c, ZMIC5, ATS5d, SssCH, ATSC3c, C3SP3, NssNH, EState_VSA9, VSA_EState9, ATS4p, ATS4m, BCUTc-1h, SssN, nBondsKD, C3SP2, nBondsD, ATSC3i, SdssC, SssCH2, C2SP1, Vabc, n3AHRing, MIC4, EState_VSA1, AATS1Z, MATS1Z, PEOE_VSA9, EState_VSA4, TIC5, C4SP3, VE2_Dzv, ETA_dPsi_B, GATS2are, EState_VSA2, PEOE_VSA8, nHRing, NdsN, EState_VSA5, n4HRing, AATSC2dv, GATS2i, GATS1m, n9FRing, ATSC3se, VSA_EState6, AATS1dv, ATS3dv, n8FAHRing, VR1_A, NsCH3, n6Ring, PEOE_VSA12, NsNH2, ATS4d, SssssC, RPCG, ATSC3dv, PEOE_VSA3, ETA_shape_x, EState_VSA8, BalabanJ, JGI1, EState_VSA6, nBridgehead, n5FHRing, PEOE_VSA2, ATSC7v, SdO, n7FRing, NdssC, NssCH, WPol, C2SP3, VSA_EState8, ATS6dv, NssO, JGI2, ATSC4d, ATSC4c, MIC1, BCUTv-1l, PEOE_VSA11, n6aRing, VE2_A, n5FAHRing, ATSC2p, ATSC8i, BIC2, VSA_EState4, ZMIC2, SlogP_VSA10, NdNH, SMR_VSA5, PEOE_VSA10, EIndex, SMR_VSA2, ATSC6d, GGI5, ATSC5c, BIC1, ATSC6c, n7ARing, n5AHRing, n6FAHRing, PEOE_VSA13, SaaCH, ATSC8Z, EState_VSA3, ATSC4are, GGI4, ATS4i, EState_VSA7, ZMIC0, GGI6, ATSC6se, JGI4, ATSC1dv, AATSC2d, nAcid, ZMIC1, ATS7dv, ATSC6i, n6FRing, ETA_beta_ns_d, SsF, GGI3, Radius, nSpiro, C1SP2, n9Ring, VSA_EState4, NaaN, nFAHRing
RF	nC, SlogP_VSA2, ZMIC0, VSA_EState7, nBondsKD, ATSC0i, ETA_eta_R, BCUTi-1h, BertzCT, SMR_VSA7, BCUTZ-1h, n5Ring, SpAbs_Dzv, ATSC2se, ATS0dv, ATSC1se, C2SP2, GATS1se, BCUTdv-1h, RNCG, SpAbs_A, BCUTZ-1l, nBondsD, ABC, VE1_A, VR1_A, nBase, ATSC2c, nN, nBridgehead, BCUTd-1l, n4HRing, VE2_A, RPCG, nO, BCUTc-1h, n6Ring, nRot, SpAbs_DzZ, n5FRing, BCUTc-1l, Xc-5d, ATS4v, ATS5d, Xch-3d, BCUTd-1h, ATS1dv, nSpiro, ATS5se, MIC2, GGI3, NdsCH, n5HRing, SlogP_VSA4, SlogP_VSA1, SssssC, PEOE_VSA9, VSA_EState4, BalabanJ, NdO, ATSC2dv, VSA_EState5, ATS3dv, GATS2c, ETA_dEpsilon_B, SlogP_VSA5, VSA_EState6, EState_VSA9, ATS3Z, NtN, n7FRing, ATSC2d, n8FRing, nAromAtom, SdssC, FilterItLogS, n5FHRing, nHRing, ZMIC1, IC1, GGI4, NdssC, JGI4, NssNH, ATS4d, EState_VSA8, PEOE_VSA2, BCUTdv-1l, ATSC2Z, ATS6d, EState_VSA10, ATSC4d, NssN, PEOE_VSA7, n6FRing, n6aRing, C1SP2, PEOE_VSA8, VSA_EState8, ATSC3dv, ATS4dv, ATSC3se, ATSC1dv, NssCH, piPC8, SMR_VSA5, ZMIC2, PEOE_VSA12, ETA_shape_p, ATSC6d, C3SP2, ATSC3Z, ATSC3c, NssO, ATS4se, PEOE_VSA6, ATSC5d, PEOE_VSA3, IC3, n3HRing, EIndex, C2SP1, SlogP_VSA10, ATSC4dv, ATSC4c, SMR_VSA4, ATSC6c, VSA_EState9, PEOE_VSA10, ATSC4se, EState_VSA4, SMR_VSA6, ATSC5c, ATSC5dv, NssCH2, ATSC5se, IC2, PEOE_VSA11, ATSC6dv, NsNH2
CFS	Si, SlogP_VSA2, nH, AATS1se, n4Ring, GATS1m, SpMAD_Dzp, n3Ring, Mi, ATSC0p, nHBAcc, VE2_Dzi, n4HRing, AATS1are, SlogP_VSA4, ATS0i, AATSC0p, AATS0i, CIC0, ETA_dPsi_A, ETA_dEpsilon_C, ETA_dAlpha_B, AATSC1se, ETA_dEpsilon_D, ATSC1c, ATSC0c, VE2_Dzp, AATS1pe, NssssC, SpMAD_Dzse, GATS1i, AATS1p, AATS1i, AMID_C, ETA_epsilon_2, SM1_Dzv, BIC0, AATSC0v, GATS1v, MATS1v, AATS2i, AATSC1i, GATS1p, AATS0p, AMID_h, ATSC0v, FilterItLogS, ATS5se, VE2_D, MID_h, nC, SM1_Dzi, MID_C, SlogP_VSA5, PEOE_VSA1, VSA_EState9, VE2_Dzv, LabuteASA, BCUTc-1l, C1SP3, ATSC4i, ATSC3v, Xpc-4dv, SM1_Dzare, VE2_Dzare, piPC1, SM1_Dzpe, VE2_Dzpe, ETA_alpha, nRot, VE2_Dt, VE1_Dt, NdsN, BCUTv-1h, MIC0, VE2_DzZ, ETA_epsilon_4, BCUTp-1l, BCUTdv-1h, GATS1pe, AATSC1c, TopoPSA(NO), n4ARing, NsCH3, GATS2m, TIC0, JGI1, AATSC2d, AETA_eta_L, PEOE_VSA5, BCUTse-1l, ETA_dEpsilon_B, FCSP3, nBondsM, n3HRing, NsOH, CIC5, BIC4, CIC4, BIC5, ATSC1se, TopoPSA, VE2_Dzm, AATS2se, SssN, ATSC1i, SpMAD_Dzv, MATS2v, ETA_psi_1, SM1_Dzm, VE2_Dzse, Xp-1d, JGT10, ATSC3m, nBridgehead, SlogP_VSA10, ATSC4dv, ATSC3c, AATSC2c
LASSO	VR1_A, nBridgehead, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS8m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC8m, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC6v, ATSC7v, ATSC1i, ATSC2i, SpAD_Dzp, BertzCT, SpDiam_Dt, DetourIndex, SaaCH, SssO, EIndex, ETA_beta, fragCpx, TIC1, PEOE_VSA2, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, SMR_VSA2, SMR_VSA5, SMR_VSA7, SlogP_VSA1, SlogP_VSA2, SlogP_VSA6, EState_VSA2, EState_VSA4, EState_VSA5, EState_VSA6, EState_VSA8, VSA_EState2, VSA_EState4, VSA_EState6, VSA_EState7, MPC3, MPC4, MPC5, MPC8, TopoPSA(NO), SRW09

A.3.3 Entalpia de atomização para a temperatura de 298.15 K

Metódos	Atributos selecionados
Pcorr	ATS0v, SlogP_VSA2, VSA_EState7, TpiPC10, BCUTv-1h, SMR, PEOE_VSA7, SMR_VSA6, SlogP_VSA4, BCUTi-1h, C2SP2, VSA_EState5, SMR_VSA7, SdsCH, AATSC1se, ETA_dEpsilon_B, GATS1se, PEOE_VSA6, piPC8, SMR_VSA4, GATS1c, C1SP3, ATS5are, nRot, SlogP_VSA5, FilterItLogS, RNCG, nBase, SlogP_VSA1, ZMIC0, n5ARing, C3SP3, ATS4p, ATS5d, EState_VSA9, nBondsKD, ETA_beta, nBondsD, GATS1m, ETA_dPsi_B, VSA_EState9, NssNH, AATSC2se, C3SP2, ATSC3c, SdssC, ATS4Z, NssN, EState_VSA4, EState_VSA1, nFRing, ATSC3i, Vabc, n3HRing, PEOE_VSA9, EState_VSA5, StCH, C2SP1, TIC5, SssCH2, C4SP3, n4HRing, GATS2are, nAHRing, NdsN, EState_VSA2, n9FRing, AATSC0i, BalabanJ, ZMIC5, NsCH3, GATS2i, WPol, n8FARing, TSRW10, SssssC, BCUTc-1h, ATS4d, ATSC3se, AATSC2c, n6Ring, PEOE_VSA8, VR1_A, VSA_EState8, NsNH2, PEOE_VSA12, ATSC5d, ETA_eta_R, NssCH, n5FAHRing, ATSC3dv, BCUTd-1l, RPCG, nBridgehead, EState_VSA8, NdssC, PEOE_VSA3, ECIIndex, ATSC7d, NssO, n7FRing, ATS3dv, n6aRing, JGI2, C2SP3, PEOE_VSA2, ATSC4d, PEOE_VSA10, SaaCH, n5FRing, EState_VSA6, VSA_EState4, ATSC2p, ATSC7Z, BCUTdv-1h, SdO, ATSC4c, PEOE_VSA11, SlogP_VSA10, n7Ring, PEOE_VSA4, NdNH, MIC1, ATSC5c, BIC2, GGI5, EState_VSA7, n5AHRing, SaaC, ATSC8c, ATSC6d, VSA_EState6, n6FHRing, ATSC8pe, MID_X, PEOE_VSA13, Xc-6d, ZMIC1, JGI4, ATSC1dv, AATSC2d, ATS7dv, ATS4are, GGI4, ATSC4are, ATSC6i, NaaN, ATSC7c, GGI6, n6FRing, nAcid, Xch-3d, SMR_VSA5, MIC2, ATSC6c, n9Ring, ETA_beta_ns_d, ATSC6dv, EState_VSA10, nSpiro, SaaO, JGI3
Scorr	ATS0v, SLogP, LabuteASA, VSA_EState7, TpiPC10, VE1_Dzv, SMR, RNCG, VSA_EState5, SMR_VSA6, PEOE_VSA7, SlogP_VSA4, VE1_Dt, C2SP2, GATS1c, GATS1se, ETA_dEpsilon_B, nN, nRot, ATS5are, SMR_VSA7, SMR_VSA4, C1SP3, FilterItLogS, nBase, nH, PEOE_VSA6, SlogP_VSA5, nBondsM, SlogP_VSA1, piPC8, n5ARing, SdsCH, AATSC2se, ETA_eta_RL, AATSC2c, ZMIC5, ATS5d, SssCH, ATSC3c, C3SP3, NssNH, EState_VSA9, VSA_EState9, ATS4p, ATS4m, BCUTc-1h, SssN, nBondsD, C3SP2, nBondsD, ATSC3i, SdssC, SssCH2, C2SP1, Vabc, n3AHRing, MIC4, EState_VSA1, AATS1Z, MATS1Z, PEOE_VSA9, EState_VSA4, TIC5, C4SP3, VE2_Dzv, ETA_dPsi_B, GATS2are, EState_VSA2, PEOE_VSA8, nHRing, NdsN, EState_VSA5, n4HRing, AATSC2dv, GATS2i, GATS1m, n9FRing, ATSC3se, VSA_EState6, AATS1dv, ATS3dv, n8FARing, VR1_A, NsCH3, n6Ring, PEOE_VSA12, NsNH2, ATS4d, SssssC, RPCG, ATSC3dv, PEOE_VSA3, ETA_shape_x, EState_VSA8, BalabanJ, JGI1, EState_VSA6, nBridgehead, n5FHRing, PEOE_VSA2, ATSC7v, SdO, n7FRing, NdssC, NssCH, WPol, C2SP3, VSA_EState8, ATS6dv, NssO, JGI2, ATSC4d, ATSC4c, MIC1, BCUTv-1l, PEOE_VSA11, n6aRing, VE2_A, n5FARing, ATSC2p, ATSC8i, BIC2, VSA_EState4, ZMIC2, SlogP_VSA10, NdNH, SMR_VSA5, PEOE_VSA10, ECIIndex, SMR_VSA2, ATSC6d, GGI5, ATSC5c, BIC1, ATSC6c, n7ARing, n5AHRing, n6FAHRing, PEOE_VSA13, SaaCH, ATSC8Z, EState_VSA3, ATSC4are, GGI4, ATS4i, EState_VSA7, ZMIC0, GGI6, ATSC6se, JGI4, ATSC1dv, AATSC2d, nAcid, ZMIC1, ATS7dv, ATSC6i, n6FRing, ETA_beta_ns_d, SsF, GGI3, Radius, nSpiro, C1SP2, n9Ring
RF	nC, SlogP_VSA2, ZMIC0, VSA_EState7, nBondsKD, ATSC0i, ETA_eta_R, C2SP2, BCUTZ-1h, BCUTi-1h, SMR_VSA7, BertzCT, n5Ring, ATS0dv, ATSC2se, SpAbs_Dzv, ATSC1se, BCUTZ-1l, GATS1se, BCUTdv-1h, RNCG, nBondsD, VE1_A, SpAbs_A, ABC, nBase, VR1_A, nBridgehead, ATSC2c, VE2_A, n4HRing, BCUTd-1l, BCUTc-1h, RPCG, nO, nN, n6Ring, BCUTc-1l, SpAbs_DzZ, nRot, Xc-5d, n5FRing, MIC2, ATS4v, SlogP_VSA4, ATS1dv, ATS5d, ATS5se, BCUTd-1h, nSpiro, Xch-3d, ETA_dEpsilon_B, SlogP_VSA1, NdsCH, PEOE_VSA9, GGI3, BalabanJ, VSA_EState5, n5HRing, GATS2c, VSA_EState4, SssssC, ATS3dv, NtN, ATSC2dv, ATS3Z, EState_VSA9, SdssC, VSA_EState6, n5FHRing, NdO, n7FRing, ATSC2d, n8FRing, NssN, ZMIC1, SlogP_VSA5, FilterItLogS, nAromAtom, JGI4, GGI4, IC1, nHRing, EState_VSA8, SlogP_VSA10, ATS4d, PEOE_VSA2, n6FRing, ATSC2Z, NdssC, n6aRing, ATSC6d, BCUTdv-1l, ATSC4d, ATSC1dv, VSA_EState8, PEOE_VSA8, EState_VSA10, PEOE_VSA7, ATSC3dv, ATSC3se, PEOE_VSA6, piPC8, C1SP2, ATS4dv, SMR_VSA5, NssCH, NssNH, ATSC3Z, SMR_VSA4, ZMIC2, PEOE_VSA12, ATSC6d, ATSC3c, ATS4se, ATSC5d, C3SP2, VSA_EState9, PEOE_VSA10, IC3, PEOE_VSA3, ETA_shape_p, ECIIndex, ATSC4dv, EState_VSA4, ATSC4c, n3HRing, ATSC5dv, ATSC6c, SMR_VSA6, ATSC4se, ATSC5c, NssO, C2SP1, NssCH2, NsNH2, ATSC5se, IC2, PEOE_VSA11, EState_VSA2
CFS	Si, SlogP_VSA2, nH, AATS1se, n4Ring, GATS1m, SpMAD_Dzp, n3Ring, Mi, ATSC0p, nHBAcc, VE2_Dzi, n4HRing, AATS1are, SlogP_VSA4, ATS0i, AATSC0p, AATS0i, CIC0, ETA_dPsi_A, ETA_dEpsilon_C, ETA_dAlpha_B, AATSC1se, ETA_dEpsilon_D, ATSC1c, ATSC0c, VE2_Dzp, AATS1pe, NssssC, SpMAD_Dzse, GATS1i, AATS1p, AATS1i, AMID_C, ETA_epsilon_2, SM1_Dzv, BIC0, ATSC0v, GATS1v, MATS1v, AATS2i, AATSC1i, GATS1p, AATS0p, AMID_h, AATSC0v, FilterItLogS, ATSC5se, VE2_D, MID_h, nC, SM1_Dzi, MID_C, SlogP_VSA5, PEOE_VSA1, VSA_EState9, VE2_Dzv, LabuteASA, BCUTc-1l, C1SP3, ATSC4i, ATSC3v, Xpc-4dv, SM1_Dzare, VE2_Dzare, piPC1, SM1_Dzpe, VE2_Dzpe, ETA_alpha, nRot, VE2_Dt, VE1_Dt, NdsN, BCUTv-1h, MIC0, VE2_DzZ, ETA_epsilon_4, BCUTp-1l, BCUTdv-1h, GATS1pe, AATSC1c, TopoPSA(NO), n4ARing, NsCH3, GATS2m, TIC0, JGI1, AATSC2d, AETA_eta_L, PEOE_VSA5, BCUTse-1l, ETA_dEpsilon_B, FCSP3, nBondsM, n3HRing, NsOH, CIC5, BIC4, CIC4, BIC5, ATSC1se, TopoPSA, VE2_Dzm, AATS2se, SssN, ATSC1i, SpMAD_Dzv, MATS2v, ETA_psi_1, SM1_Dzm, VE2_Dzse, ETA_eta_RL, JGT10, ATSC3m, nBridgehead, SlogP_VSA10, ATSC4dv, ATSC3c, AATSC2c
LASSO	VR1_A, nBridgehead, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS8m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC8m, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC6v, ATSC7v, ATSC1i, ATSC2i, SpAD_Dzp, BertzCT, SpDiam_Dt, DetourIndex, SaaCH, SssO, ECIIndex, ETA_beta, fragCpx, TIC1, PEOE_VSA2, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, SMR_VSA2, SMR_VSA5, SMR_VSA7, SlogP_VSA1, SlogP_VSA2, SlogP_VSA6, EState_VSA2, EState_VSA4, EState_VSA5, EState_VSA6, EState_VSA8, VSA_EState2, VSA_EState4, VSA_EState6, VSA_EState7, MPC3, MPC4, MPC5, MPC8, TopoPSA(NO), SRW09

A.3.4 Energia livre de Gibbs de atomização para a temperatura de 298.15 K

Metódos	Atributos selecionados
Pcorr	AATS0v, SlogP_VSA2, VSA_EState7, BCUTv-1h, TpiPC10, SMR, PEOE_VSA7, SlogP_VSA4, SMR_VSA6, BCUTi-1h, C2SP2, VSA_EState5, SMR_VSA7, ETA_dEpsilon_B, SdsCH, AATSC1se, GATS1se, PEOE_VSA6, SlogP_VSA5, GATS1c, FilterItLogS, C1SP3, SMR_VSA4, piPC8, ATS5are, RNCG, nRot, nBase, SlogP_VSA1, ZMIC0, n5Ring, C3SP3, ATS4p, nBondsKD, nBondsD, EState_VSA9, ETA_dPsi_B, AATSC2se, NssNH, ATS5d, C3SP2, VSA_EState9, Vabc, ATSC3c, NsssN, SdssC, GAT51m, C2SP1, PEOE_VSA9, n3HRRing, StCH, ATS4Z, EState_VSA1, EState_VSA4, ATSC3i, nAHRing, EState_VSA5, n4HRRing, TIC5, nFARing, GATS2are, C4SP3, SssCH2, NdsN, EState_VSA2, GATS2i, ZMIC5, AATSC0i, n9FARing, SssssC, WPol, NsCH3, ATSC3se, n8FRRing, SRW09, BalabanJ, BCUTc-1h, PEOE_VSA8, ETA_eta_R, n6Ring, NsNH2, VR1_A, ATS4d, n5FAHRing, RPCG, VSA_EState8, NdssC, PEOE_VSA12, EState_VSA8, NssO, ECIndex, PEOE_VSA3, ATSC5d, ATSC3dv, ATSC3dv, NsssCH, BCUTd-1l, ATSC7d, n6aRing, nBridgehead, n7FRRing, JGI2, ATSC5m, SaaCH, ATSC4d, n5FRRing, PEOE_VSA2, PEOE_VSA10, VSA_EState4, SdO, C2SP3, EState_VSA6, ATSC2p, ATSC7Z, BCUTdv-1h, ATSC4c, SlogP_VSA10, PEOE_VSA11, EState_VSA7, NdNH, n7ARing, ATSC5c, PEOE_VSA4, SaasC, ATSC6d, BIC2, n6FAHRing, MIC1, ATSC8c, VSA_EState6, Xc-6d, ATS4are, ATSC8pe, AATSC2d, GGI5, n5HRRing, SsF, ATSC1dv, PEOE_VSA13, AATS1dv, ATSC4are, Xch-3d, ATSC7c, JGI4, NaaN, GGI4, ATSC6dv, nAcid, ATSC6i, EState_VSA10, ATSC7dv, ATSC6c, MIC2, n6FRRing, JGI3, ZMIC1, n9Ring, ETA_beta_ns_d, SaaO, GGI6, BCUTc-1l, SMR_VSA5, ATSC4i, nSpiro, VE2_A, SlogP_VSA11
Scorr	ATS0v, SLogP, LabuteASA, VSA_EState7, SMR, TpiPC10, VE1_Dzv, RNCG, VSA_EState5, SMR_VSA6, SlogP_VSA4, PEOE_VSA7, C2SP2, VE1_Dt, GATS1se, ETA_dEpsilon_B, GATS1c, nN, FilterItLogS, nBondsM, PEOE_VSA6, nBase, SMR_VSA7, SlogP_VSA5, C1SP3, nH, SMR_VSA4, ATS5are, nRot, SlogP_VSA1, AATSC2se, SdsCH, piPC8, n5ARing, AATSC2c, ETA_eta_RL, ZMIC5, ATSC3c, NssNH, ATS4p, C3SP3, ATS5d, SsssCH, EState_VSA9, VSA_EState9, nBondsKD, SsssN, Vabc, C2SP1, nBondsD, C3SP2, BCUTc-1h, ATS4m, SdssC, n3AHRing, ATSC3i, MIC4, PEOE_VSA9, SssCH2, EState_VSA1, nH-Ring, ETA_dPsi_B, GATS2are, C4SP3, n4HRRing, MATS1Z, EState_VSA4, TIC5, PEOE_VSA8, VE2_Dzv, AATSC2dv, GATS2i, EState_VSA2, NdsN, AATS1dv, EState_VSA5, ATSC3se, VSA_EState6, ATSC3dv, n9FARing, GATS1m, n8FARing, VR1_A, RPCG, PEOE_VSA12, n6Ring, NsNH2, ETA_shape_x, NsCH3, SssssC, PEOE_VSA3, EState_VSA8, SdO, ATSC3dv, ATS4d, n5FRRing, NdssC, JGI1, PEOE_VSA2, NssO, EState_VSA6, ATSC7v, nBridgehead, BalabanJ, WPol, VSA_EState8, n7FARing, JGI2, ATSC4c, C2SP3, ATSC4d, NsssCH, VE2_Dt, n6aRing, PEOE_VSA11, ATS6dv, BCUTv-1l, MIC1, n5FARing, ZMIC2, ATSC2p, VSA_EState4, ECIndex, BIC2, VE2_A, SlogP_VSA10, NdNH, ATSC6d, ATSC6c, ATSC8i, SMR_VSA5, ATSC5c, PEOE_VSA10, SMR_VSA2, n6FAHRing, n7Ring, BIC1, SaaCH, GGI5, EState_VSA7, PEOE_VSA13, n5AHRing, EState_VSA3, ATSC6dv, ATSC4are, AATSC2d, ATSC1dv, ATSC8Z, ZMIC0, GGI4, ETA_eta_R, ATS4are, nAcid, JGI4, SsF, BCUTd-1l, GGI6, ATSC6i, AETA_beta_ns_d, Xc-6d, n6FARing, n9Ring, EState_VSA10, ZMIC1, ATSC7dv, ATSC4i, SaaO, GGI3, nSpiro, NaaN
RF	nC, SlogP_VSA2, ZMIC0, nBondsKD, VSA_EState7, ATSC0i, ETA_eta_R, BCUTZ-1h, BCUTi-1h, SMR_VSA7, ATS0dv, n5Ring, C2SP2, ATSC2se, BCUTZ-1l, GATS1se, BCUTdv-1h, ATSC1se, nBondsD, ABC, BertzCT, RNCG, VR1_A, SpAbs_A, nN, nBase, SpAbs_Dzv, BCUTd-1l, VE1_A, nBridgehead, n4HRRing, ATSC2c, BCUTc-1h, ETA_dEpsilon_B, VE2_A, RPCG, nO, n5FRRing, BCUTc-1l, n6Ring, SpAbs_DzZ, SlogP_VSA1, VSA_EState5, ATS4v, Xc-5d, Xch-3d, nSpiro, nRot, BCUTd-1h, ATS5se, VSA_EState6, VSA_EState4, NdO, GGI3, PEOE_VSA9, MIC2, n5HRRing, ATS5d, NdsCH, SssssC, ATS1dv, EState_VSA9, BalabanJ, NtN, GATS2c, ATSC2d, ATSC2dv, ATS3dv, n8FRRing, n7FRRing, SdssC, SlogP_VSA4, SlogP_VSA5, FilterItLogS, n5FHRing, ATSC3Z, ZMIC1, GGI4, nHRRing, NssNH, JGI4, nAromAtom, ATS4d, EState_VSA8, IC1, ATSC2Z, BCUTdv-1l, NdssC, n6FRRing, EState_VSA10, PEOE_VSA2, n6aRing, NsNH2, PEOE_VSA6, PEOE_VSA7, ATSC4d, SlogP_VSA10, VSA_EState8, PEOE_VSA8, ATSC3dv, ATS4dv, ATSC1dv, piPC8, C1SP2, ATSC3se, PEOE_VSA12, SMR_VSA5, ATSC3Z, ATS4se, ZMIC2, ATSC3c, ATS6d, PEOE_VSA3, IC3, ATSC6d, PEOE_VSA10, ECIndex, SMR_VSA4, SMR_VSA6, NsssN, VSA_EState9, ATSC5d, EState_VSA4, ATSC4dv, ATSC4c, n3HRRing, C2SP1, NssO, ATSC5c, ATSC4se, ETA_shape_p, ATSC5dv, NsssCH, PEOE_VSA11, ATSC5se, ATSC6c, C3SP2, IC2, NssCH2, EState_VSA5
CFS	Si, SlogP_VSA2, nH, AATS1se, n4Ring, GATS1m, Xch-3d, SpMAD_Dzp, Mi, FCSP3, ETA_dEpsilon_B, n4ARing, GATS1v, nHBacc, VE2_Dt, ABCGG, AATS1i, ATSC0p, ATS0i, AATSC1i, GATS1p, VSA_EState7, AATS0v, AATS0i, AATSC0p, ETA_dAlpha_B, VE2_Dzare, SM1_Dzi, VSA_EState5, MATS1c, TopoPSA(NO), ATSC1d, GATS1i, GATS2v, ATSC2v, GATS2m, JGI1, AATSC2d, MATS1p, AATS2i, AETA_eta, ATSC3d, Xc-5dv, SpMAD_Dzv, VE2_D, AATS1are, Sp, ETA_dPsi_A, SM1_Dzp, PEOE_VSA1, ATSC1se, nHBDOn, n4HRRing, BalabanJ, VE2_Dzpe, ETA_epsilon_4, SMR_VSA6, n3Ring, GATS1Z, AATS1pe, AATSC1c, MATS2se, FilterItLogS, AATSC1se, ETA_epsilon_2, AATS0p, ATS5pe, nBridgehead, ATSC4i, ATSC5i, C2SP3, C1SP3, ETA_dEpsilon_D, VE2_DzZ, AATS2se, apol, VSA_EState9, GATS2p, VSA_EState8, ATSC3se, ATSC4pe, AMID_h, ETA_shape_x, ATS7pe, VE2_Dzm, MID, SM1_Dzare, VE2_Dzse, StCH, BIC0, Xp-3dv, ATSC3m, ATSC5se, NsCH3, BCUTdv-1l, ATSC4c, MID_h, CIC0, MIC0, SM1_Dzv, TopoPSA, AMID_C, ATSC6i, ATS4i, NdsN, ATSC3c, bpol, BCUTc-1l, ETA_beta_ns_d, SpMAD_Dzi, SpDiam_A, VE2_Dzv, SM1_Dzpe, SsssN, ATS8d, ATSC1i, NaasN, AATS2are, VE2_Dzi, ATS4Z, SpMax_A, ATSC3i, MATS2i, GATS2i, ATSC2i, SlogP_VSA5, SssNH, EState_VSA9, C3SP3, NdsCH, ATS7i, ATS2i, ETA_shape_p, GATS2Z, NaaO, NaaNH, ETA_eta, Vabc, BCUTare-1l, ATSC0v, BCUTi-1h, BCUTv-1h, SM1_Dzse, BCUTp-1l, EState_VSA10, ATSC1are, NsOH, SssssC, SsNH2, nAHRing, VMcGowan, AETA_eta_L, SM1_Dzm, VE2_Dzp, VSA_EState3, AATSC1pe, ETA_dEpsilon_A, ETA_dEpsilon_C, SM1_DzZ, AETA_alpha, AATSC0Z, ETA_psi_1, TIC0, nC, ATSC6m, ATSC5m, ATS4se, ATSC7m, CIC1, BIC2, CIC3, BIC3, CIC2, BIC1, AATSC0v, BCUTm-1l, ATS8i, ATSC5dv, GGI4, VE2_A, NtsC
LAGSO	VR1_A, nBridgehead, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS8m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC8m, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC6v, ATSC7v, ATSC1i, ATSC2i, SpAD_Dzp, BertzCT, SpDiam_Dt, DetourIndex, SaaCH, SssO, ECIndex, ETA_beta, fragCpx, TIC1, PEOE_VSA2, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, SMR_VSA2, SMR_VSA5, SMR_VSA7, SlogP_VSA1, SlogP_VSA2, SlogP_VSA4, SlogP_VSA6, EState_VSA2, EState_VSA4, EState_VSA5, EState_VSA6, EState_VSA8, VSA_EState2, VSA_EState4, VSA_EState6, VSA_EState7, MPC3, MPC4, MPC5, MPC8, TopoPSA(NO), SRW09

A.3.5 Capacidade térmica para a temperatura de 298.15 K

Metódos	Atributos selecionados
Pcorr	ATS2i, AETA_eta_R, SpMAD_A, LogEE_D, ATS4Z, BalabanJ, ATS5pe, piPC8, SssCH2, nFRing, VE2_-A, ATS4v, BIC1, EState_VSA4, BCUTd-1l, nFHRing, nRot, MATS1pe, NsssCH, nAHRing, n5ARing, SMR_VSA9, NsCH3, SssssC, VE1_A, RNCG, EState_VSA1, AATSC1c, nBridgehead, n9FRing, BCUTc-1h, C2SP3, EState_VSA5, GATS2m, ATS4d, TSRW10, n8FARing, PEOE_VSA7, VE2_Dt, n5AHRing, SMR_VSA4, MIC2, ZMIC1, EState_VSA9, C3SP3, ETA_dEpsilon_B, ATSC5dv, EState_VSA2, n6ARing, SLogP, EState_VSA6, NdO, SssO, ATSC5d, Radius, PEOE_VSA10, JGI2, ATSC3m, GATS2pe, EState_VSA10, ATS3dv, EState_VSA7, n7FARing, GATS1pe, ATSC2dv, ATSC7d, BCUTse-1l, NsNH2, nSpiro, BIC2, ZMIC5, C1SP2, SIC5, VR2_A, VSA_EState7, GGI5, n4AHRing, SdsCH, SMR_VSA2, ATSC8i, RPCG, GGI4, SMR_VSA5, PEOE_VSA11, nBondsKD, BCUTp-1l, ATSC3dv, JGI4, NdssC, C1SP3, ATSC4dv, ATS7dv, SlogP_VSA1, GGI6, VSA_EState9, PEOE_VSA12, ATSC7p, ATS4i, BCUTi-1h, ATSC3pe, WPol, SMR_VSA7, ETA_beta_ns_d, PEOE_VSA9, nBondsD, PEOE_VSA2, SssN, ATSC3c, SlogP_VSA10, n7Ring, EState_VSA3, VSA_EState8, SdssC, NaasC, PEOE_VSA8, ATSC4d, SlogP_VSA11, JGI3, n6FHRing, ZMIC0, n6FRing, ATSC5c, AATSC2d, AATSC0i, BCUTc-1l, VSA_EState4, PEOE_VSA13, GGI7, n3HRing, PEOE_VSA3, NdNH, NdsN, ATSC8c, ATSC6i, C4SP3, ATSC6se, NaaNH, Xc-6d, n5FARing, ATS8dv, NaaO, FilterItLogS, NssNH, n8Ring, NaaN, ATSC4c, ATSC7c, n5FHRing, SsF, SlogP_VSA5, PEOE_VSA6, EState_VSA8, NaasN
Scorr	ATS2i, VE2_D, AETA_eta_R, VE2_Dt, SpMAD_A, SssCH, BCUTd-1l, MPC8, ATS4Z, VE2_A, Ss-sCH2, ATS5pe, AETA_eta_RL, SpMax_Dzv, BIC1, nFHRing, ATSC1pe, MIC1, ATS4v, nRot, EState_VSA4, VE1_A, GATS1m, nAHRing, BCUTc-1h, EState_VSA1, n5ARing, AATSC1c, NsssCH, SMR_VSA9, NsCH3, RNCG, nBridgehead, C2SP3, ATS4d, SssssC, GATS2m, EState_VSA5, JGI1, EState_VSA9, n8FARing, n9FARing, ETA_shape_y, VR2_A, ATS6dv, ZMIC5, SssO, MIC2, EState_VSA2, n5AHRing, ATS8m, ZMIC1, EState_VSA6, ATSC5dv, VSA_EState7, n6ARing, SMR_VSA1, EState_VSA7, SMR_VSA4, C3SP3, ATSC2v, CIC5, n7FRing, SLogP, GATS2pe, PEOE_VSA7, ATSC3m, Radius, NdO, ETA_dEpsilon_B, PEOE_VSA10, BCUTi-1h, C1SP2, GATS1pe, JGI2, NsNH2, ATSC2dv, nSpiro, BCUTm-1h, ATSC3dv, ATSC7i, ATS7dv, BIC2, bpol, ATSC1dv, EState_VSA10, GGI5, SMR_VSA2, n4HRing, JGI4, VSA_EState4, nBondsKD, SlogP_VSA1, RPCG, C1SP3, GGI6, GGI4, VSA_EState9, AATSC1dv, ATSC3dv, NdssC, ATSC3pe, SdssC, PEOE_VSA12, PEOE_VSA11, AETA_beta_ns_d, VSA_EState5, ATSC4dv, nH, SMR_VSA7, ATS4i, ATSC3c, SlogP_VSA10, nBondsD, SMR_VSA5, PEOE_VSA9, n7ARing, WPol, PEOE_VSA2, VSA_EState8, BCUTse-1l, SlogP_VSA11, n6FAHRing, n6FRing, ATSC5c, NaasC, SssN, GGI7, PEOE_VSA13, AATSC2c, ATSC4d, ATSC6Z, NdNH, JGI3, PEOE_VSA3, ATS8dv, EState_VSA3, ZMIC0, Xc-6dv, n5FRing, ETA_dPsi_B, NaaNH, n3HRing, NdsN, C4SP3, SdsCH, EState_VSA8, NaaO, Ns-sNH, NaaN, SlogP_VSA5, n8ARing, C2SP2, n5FHRing, BCUTc-1l, ATSC6d, ATSC4c, FilterItLogS, SlogP_VSA3, PEOE_VSA8, SsF, PEOE_VSA6, ATSC6c, ZMIC2, NaasN, SMR_VSA6, n9ARing
RF	ETA_shape_p, ETA_eta_R, BCUTd-1l, BalabanJ, VE2_A, nFRing, ETA_dEpsilon_B, n9FRing, SpAbs_-Dzv, ATS4v, ABC, BCUTZ-1l, NsssCH, VE1_A, GGI5, C2SP1, Xch-3d, ATS5dv, GGI4, SpAbs_A, SpAbs_-Dzv, GATS1se, n7FRing, ATS5se, piPC8, BCUTd-1h, BCUTi-1h, VR1_A, n6FRing, VSA_EState5, nRot, ATSC1se, ZMIC0, n5Ring, IC1, ATS5d, ATS0dv, n8FRing, ATS7d, ATS6d, RNCG, nBridgehead, ATS3dv, SssssC, n5FRing, JGI4, GGI3, BertzCT, EState_VSA9, ECIndex, VSA_EState4, BCUTc-1h, Xc-5d, ATSC2d, BCUTdv-1h, RPCG, FilterItLogS, ATSC3Z, BCUTc-1l, BCUTZ-1h, ATS3Z, ATS4dv, VSA_EState7, EState_VSA1, ATSC2Z, ATSC4d, SlogP_VSA4, ATS1dv, GATS2c, ATS4d, MIC2, ATSC2se, ATSC6c, ZMIC1, ATS8Z, SMR_VSA7, ATS4se, SlogP_VSA2, nHRing, ATSC3dv, VSA_EState8, BCUTdv-1l, SMR_VSA4, n4HRing, ATSC7d, ATSC7c, NdO, nSpiro, ATSC4dv, EState_VSA2, GGI6, ATSC1dv, ATS6dv, EState_VSA4, ATSC5d, ATSC2c, ATSC2dv, SdssC, ATSC5c, ATSC5dv, ZMIC2, ATSC4c, SMR_VSA5, n6FHRing, ATSC3c, ATSC3se, NsssN, ATSC6d, ATSC6dv, nBondsT, nFHRing, EState_VSA3, PEOE_VSA2, ATSC5se, ATSC4se, SlogP_VSA5, NdsN, NssCH2, PEOE_VSA9, ATS8d, ETA_beta_ns_d, nC, IC2, PEOE_VSA8, NtCH, IC3, PEOE_VSA7, EState_VSA10, EState_VSA8, VSA_EState6, n5HRing, PEOE_VSA3, GGI7, ATSC0i, EState_VSA5, PEOE_VSA10, SMR_VSA6, ATSC7dv, n5FHRing, n6Ring
CFS	ETA_shape_p, SpMax_D, AETA_eta_R, Xch-3d, VMcGowan, SsOH, ATS5p, NsssC, SpMAD_D, ETA_eta_R, VE2_Dt, AMID, ATS4Z, C2SP1, JGI1, LogEE_D, ETA_dEpsilon_D, NsCH3, NtCH, nRot, WPol, JGI3, ATS6Z, ATS5se, ATSC1i, AETA_eta_F, n3Ring, BalabanJ, ATS7p, ATS4se, ATS6i, AETA_eta, ETA_eta_F, NtsC, AATSC0c, SM1_Dzv, ETA_dPsi_A, SM1_Dzp, PEOE_VSA10, JGT10, AETA_eta_FL, SpMAD_Dzare, ATSC2Z, NsOH, ATS4v, ATS7v, SsNH2, AETA_eta_L, Vabc, ATS5pe
LASSO	VR1_A, ATS2dv, ATS3dv, ATS0m, ATS2m, ATS3m, ATS4m, ATS5m, ATS7m, ATS8m, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS1i, ATS2i, ATS3i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC3dv, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC4m, ATSC5m, ATSC7m, ATSC8m, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC6v, ATSC8v, ATSC2i, ATSC4i, SpAD_Dzp, BertzCT, SpAbs_Dt, SpDiam_Dt, SpAD_Dt, DetourIndex, StCH, SdNH, StN, fragCpx, TIC1, TIC4, PEOE_VSA1, PEOE_VSA2, PEOE_VSA3, PEOE_VSA5, PEOE_VSA7, PEOE_VSA10, PEOE_VSA11, SMR_VSA3, SMR_VSA4, SMR_VSA5, SMR_VSA7, SMR_VSA9, SlogP_VSA2, SlogP_VSA4, SlogP_VSA5, SlogP_VSA6, EState_VSA4, EState_VSA5, EState_VSA10, VSA_EState3, VSA_EState4, VSA_EState7, MPC3, MPC7, TMPC10, TopoPSA(NO), Vabc, TSRW10, Zagreb2

A.3.6 Momento dipolar

Metodos	Atributos seleccionados
Pcorr	NtN, IC0, BCUTi-1h, VSA_EState2, BCUTc-1h, MIC1, AETA_eta_F, nN, ATS4Z, BCUTv-1h, PEOE_VSA2, GATS2se, SpMax_D, VSA_EState7, VSA_EState5, SpMAD_A, AETA_eta_R, ZMIC1, ETA_dPsi_B, ATS5are, MATS1c, nRot, SlogP_VSA1, piPC8, GATS1se, nFARing, BCUTd-1l, NdssC, VSA_EState6, MIC2, NssNH, ATS4v, C3SP3, nBondsD, MATS2se, VE2_A, ETA_dEpsilon_B, NsssCH, StCH, TSRW10, ATSC6dv, VSA_EState1, ATSC5dv, EState_VSA10, Radius, BalabanJ, SlogP, C1SP2, PEOE_VSA12, ATSC3dv, n9FARing, nBridgehead, nBondsKD, n5Ring, PEOE_VSA3, BCUTc-1l, SMR_VSA4, GGI4, C2SP1, EState_VSA5, nBase, BCUTdv-1h, nFHRing, PEOE_VSA7, ATSC3m, n8FARing, VE1_A, EState_VSA7, GGI5, NsCH3, RPCG, EState_VSA2, ATS4are, PEOE_VSA6, SlogP_VSA4, SlogP_VSA6, AATSC0i, BIC2, ATSC7dv, NsssN, C1SP3, SssCH2, EState_VSA4, VSA_EState3, RNCG, ATSC5d, AATSC1dv, SdsCH, C2SP3, nSpiro, C1SP1, NsNH2, IC5, nHRing, PEOE_VSA13, VSA_EState4, ZMIC5, SssssC, VSA_EState8, ZMIC2, ATS3dv, n5AHRing, JGI3, VR2_A, GGI6, PEOE_VSA11, n7FARing, ATSC6d, SMR_VSA5, ATSC4Z, C4SP3, ATSC8se, ATS4d, PEOE_VSA10, ATSC7c, PEOE_VSA9, ATSC7dv, PEOE_VSA8, ATSC2p, ZMIC0, FilterItLogS, ECIndex, n6Ring, VSA_EState9, NaaN, ATSC3pe, WPol, EState_VSA9, SlogP_VSA11, SMR_VSA6, GATS2i, ATSC8c, EState_VSA6, NdNH, EState_VSA8, ATSC3c, ETA_beta_ns_d, n4AHRing, NaasC, SaaCH, n6FARing, EState_VSA1, C3SP2, SaaO, n9ARing, ATSC6c, ATSC4c, SlogP_VSA5, n6aRing, SaasN, n5FAHRing, EState_VSA3, SlogP_VSA10, NdN, ATSC4d, ATSC7d, n5FRing, nAcid, SdssC, JGI4, n8AHRing, ATS8dv, NddC, n7ARing, ATSC5c, n6FAHRing, SsF, SaaNH, GGI7, n3AHRing
SCorr	BCUTi-1h, VSA_EState2, AMID_C, BCUTd-1l, NtN, BCUTc-1h, BCUTdv-1h, ATS4Z, MIC1, PEOE_VSA2, AETA_eta_F, ETA_dEpsilon_C, GATS2se, AMID, SpMAD_D, AETA_beta, SpMAD_A, VSA_EState5, AETA_eta_R, SpAbs_Dzv, VSA_EState7, ETA_dPsi_B, MPC8, GATS1se, ATS5are, nRot, SlogP_VSA1, NdssC, nBondsD, VE2_D, VSA_EState1, ZMIC1, C1SP2, ATS4v, SsssCH, StCH, MIC2, VSA_EState6, nBondsKD, NssNH, MATS1c, ATSC5dv, VE2_A, ATSC6dv, nBase, ETA_dEpsilon_B, NsssCH, MATS2se, n9FARing, C3SP3, SlogP_VSA3, SlogP, ATSC3dv, BIC1, BCUTdv-1l, ZMIC5, Radius, GGI4, AETA_eta_RL, GATS2Z, PEOE_VSA12, EState_VSA5, PEOE_VSA3, ATSC5se, VE1_A, nBridgehead, EState_VSA2, VR2_A, EState_VSA10, n5Ring, ATSC1are, ATSC3m, nFHRing, ATS4are, n8FARing, ATS6dv, EState_VSA7, IC5, RPCG, SMR_VSA4, PEOE_VSA10, GGI5, ATSC7dv, C1SP3, Xc-6dv, NsssN, SlogP_VSA4, BIC2, SssCH2, SlogP_VSA6, ATS1dv, ATSC2dv, AATSC1dv, ATS3dv, PEOE_VSA8, ATSC6c, SsOH, PEOE_VSA11, RNCG, PEOE_VSA7, nSpiro, MATS2c, C2SP3, PEOE_VSA6, NsNH2, SssssC, EState_VSA4, VSA_EState8, PEOE_VSA13, ATSC3pe, n5AHRing, PEOE_VSA9, nHRing, n7Fring, JGI1, ETA_shape_y, ATSC8se, FilterItLogS, ATS4d, NaaN, ATSC4Z, EState_VSA8, GGI6, C1SP1, GATS2i, SlogP_VSA5, C4SP3, ATSC6d, EState_VSA1, EState_VSA3, WPol, ATSC7dv, JGI3, ATSC2p, VSA_EState9, C2SP2, SlogP_VSA11, EState_VSA6, n6Ring, LabuteASA, GATS1m, nH, ECIndex, C3SP2, SdssC, ZMIC2, SlogP_VSA10, NdNH, SMR_VSA5, BCUTc-1l, n9ARing, SMR_VSA6, SaaO, NdsCH, ZMIC0, n6FARing, SaasN, n5FAHRing, EState_VSA9, n6aRing, n4HRing, ATSC4d, n5Fring, AETA_beta_ns_d, SaaCH, ATSC4c, NdsN, ATSC3c, n3HRing, ATSC5c, n8HRing, JGI4, nAcid, n7HRing, SddC, ATS8d, n6FAHRing, ATS8dv, nF, VSA_EState4, JGI7, SaaNH
RF	BCUTi-1h, BCUTd-1l, NtN, NdO, BCUTZ-1h, BCUTdv-1h, PEOE_VSA3, BCUTc-1l, SlogP_VSA2, RPCG, BCUTc-1h, BCUTZ-1l, NtCH, ATSC2c, GATS2c, ATSC8dv, RNCG, nN, ATSC3c, ATSC1se, ATSC4c, ATS4dv, ATSC6c, VSA_EState4, VSA_EState6, ATSC5c, ATSC2se, ATSC3se, PEOE_VSA10, SpAbs_Dzv, ATS4d, ATSC4dv, ATS4se, VSA_EState5, ATSC5se, SpAbs_DzZ, ATSC2dv, ATS4v, ATSC6dv, ATS3dv, ATSC4se, VSA_EState7, SdssC, ATSC5dv, ATSC5se, ATSC3dv, ATSC7dv, ATSC4d, VSA_EState8, SpAbs_A, PEOE_VSA9, ATSC3Z, GATS1se, C2SP1, ATSC2d, ETA_dEpsilon_B, EState_VSA10, PEOE_VSA2, ATSC1dv, BCUTd-1h, BCUTdv-1l, ATSC5d, PEOE_VSA8, EState_VSA2, ATS3Z, FilterItLogS, SMR_VSA7, nBase, VE1_A, ATS6d, VR1_A, EState_VSA3, ATS5dv, BertzCT, MIC2, VE2_A, SMR_VSA5, ATSC6d, ETA_eta_R, PEOE_VSA12, ATSC7c, BalabanJ, ATSC0i, ATSC2Z, GGI3, EState_VSA9, ZMIC2, ZMIC1, JGI4, ATS5d, IC1, Xc-5d, ATS1dv, NaaN, ABC, GGI4, SssssC, PEOE_VSA7, ATS0dv, nHRing, EState_VSA4, nO, nC, IC2, ECIndex, IC3, PEOE_VSA11, SlogP_VSA4, ATS7d, Xch-3d, EState_VSA8, EState_VSA5, ZMIC0, EState_VSA6, SMR_VSA6, SlogP_VSA1, ATS6dv, NdsN, ATSC7d, ATSC8c, n4HRing, SlogP_VSA5, PEOE_VSA13, EState_VSA1, VSA_EState9, SMR_VSA4, EState_VSA7, nBondsT, C2SP3, ETA_shape_p, n3HRing, C1SP3, GGI5, nBondsD, NdssC, nBondsKD, nRot, NsssCH, PEOE_VSA6, C2SP2
CFS	NtN, VSA_EState2, TIC0, StN, SdO, BCUTi-1h, BCUTc-1h, ATSC4dv, AATS2p, PEOE_VSA3, ATSC3m, ATS4i, Diameter, SMR_VSA5, SdssC, SMR_VSA2, SlogP_VSA3, RPCG, BCUTdv-1h, PEOE_VSA4, NdO, VSA_EState3, SssssC, ATS4p, BCUTpe-1l, VSA_EState5, TopoPSA(NO), CIC0, GATS1c, ETA_epsilon_5, AMID_C, TopoPSA, AATSC1c, Si, BCUTv-1h, fMF, GATS1v, ATSC4d, GGI4, SMR_VSA7, PEOE_VSA2, FCSP3, ETA_dEpsilon_B, PEOE_VSA1, SlogP_VSA4, VE2_Dt, AATS1dv, ATS6dv, SM1_Dzv, AMID_h, ATSC6i, ATSC7i, ATSC5i, ATS8are, ATS3i, ATS5p, ATS3v, ATS6i, VE2_Dzv, ZMIC0, AMID_N, BCUTpe-1h, AATS1p, GATS1p, MATS1p, BCUTv-1l, GATS1i, BCUTp-1h, ETA_dPsi_A, RNCG, BCUTp-1l, BCUTare-1h, SaasN, ATS3se, SpMAD_D, nH, VE1_Dzv, JGI2, ATSC2d, ATS4se, GGI5, nN, AETA_eta, nBondsD, VSA_EState6, AMID_O, NtsC, VE2_A, VR2_A, ATS5are, GGI3, BCUTd-1l, ETA_dBeta, BCUTse-1l, ATSC4v, WPol, ETA_dEpsilon_C, ETA_dAlpha_B, SM1_Dzv, ETA_epsilon_2, AETA_alpha, SM1_Dzse, ATSC2c, ATSC1c, ATSC3c, ATSC4se, MATS2se, SMR_VSA6, AATSC1i, bpol, StsC, BIC4, ATSC2dv, VE2_Dzare, VR1_Dzv, ATS5dv, CIC4, IC4, CIC5, SIC5, CIC3, TIC3, MATS2v, GATS2v, VSA_EState9, SlogP_VSA5, C1SP3, SdsCH, NsssCH, SaaNH, ZMIC5, BIC5, NdssC, ETA_shape_x, ATSC3dv, GATS2c, AATSC2c, ATSC5dv, AETA_dBeta, nBondsKS, ATS4m, VE2_Dzv, VE3_Dt, VE2_Dzse, ETA_eta_R
LASSO	VR1_A, nAromAtom, ATS3dv, ATS4dv, ATS5dv, ATS2d, ATS1m, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC0dv, ATSC3dv, ATSC3d, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC1v, ATSC3v, ATSC4v, ATSC5v, ATSC6v, ATSC8v, ATSC3i, ATSC5i, ATSC6i, BCUTv-1l, SpDiam_Dzare, BertzCT, nBondsA, SpDiam_Dt, DetourIndex, StCH, StN, SdsN, SsOH, SdO, SaaO, ECIndex, fragCpx, TIC4, TIC5, MIC1, ZMIC0, PEOE_VSA2, PEOE_VSA3, PEOE_VSA4, PEOE_VSA5, PEOE_VSA6, PEOE_VSA8, PEOE_VSA10, PEOE_VSA11, SMR_VSA3, SMR_VSA6, SMR_VSA7, SlogP_VSA2, SlogP_VSA3, SlogP_VSA4, SlogP_VSA6, SlogP_VSA11, EState_VSA2, EState_VSA3, EState_VSA4, EState_VSA6, EState_VSA7, EState_VSA10, VSA_EState3, VSA_EState7, MPC5, MPC8, TMPC10, TopoPSA(NO), Vabc

A.3.7 Polarizabilidade isotrópica

Metódos	Atributos selecionados
Pcorr	SMR, Mp, ETA_epsilon_4, ETA_dEpsilon_B, BCUTse-1l, VSA_EState5, SlogP_VSA2, StsC, ETA_epsilon_5, VSA_EState7, EState_VSA8, SlogP_VSA4, AATSC0i, FilterItLogS, ATSC1se, SMR_VSA7, ATS3are, BCUTm-1h, PEOE_VSA6, NssO, ATS4are, ATS4p, EState_VSA2, ECIndex, NsOH, BCUTdv-1h, WPath, ATS3Z, PEOE_VSA9, PEOE_VSA7, ATS7p, GATS2p, ATS3dv, EState_VSA1, SIC1, C3SP2, ATSC2se, SdsCH, SdssC, nAHRing, nBondsKD, EState_VSA10, SssssC, SMR_VSA6, VE2_A, ZMIC1, piPC8, SaasC, VR1_A, AETA_eta_RL, ATSC3c, EState_VSA9, C1SP3, BCUTc-1h, PEOE_VSA10, ATSC8i, Radius, GGI5, EState_VSA7, EState_VSA5, ATSC7d, RNCG, SMR_VSA5, SlogP_VSA5, GGI6, ZMIC2, nFHRing, ATSC5d, Xc-3d, VSA_EState6, EState_VSA3, n5AHRing, Xc-5d, AATSC1dv, ZMIC5, ETA_beta_ns_d, NssCH, VSA_EState4, SMR_VSA4, n6aRing, SIC2, SssCH, BCUTi-1h, GGI7, nRot, ATSC2d, JGI2, C2SP3, NdO, EState_VSA4, SaaN, ATSC5dv, ATSC4d, C4SP3, n7FAHRing, SaasN, SaaNH, BIC5, ATSC3m, ETA_eta_R, n6AHRing, ATSC7p, nBondsD, PEOE_VSA4, SlogP_VSA10, NssCH2, GGI4, ATSC4se, RPCG, ATSC6v, n4HRing, MIC4, JGI3, n3AHRing, PEOE_VSA11, SlogP_VSA11, ATSC7dv, SaaO, n6FAHRing, SsCH3, n8FAHRing, ATSSi, GATS1se, ATS4dv, ATSC7dv, C1SP2, SsF, EState_VSA6, GATS2c, ATSC3dv, BalabanJ, n9Ring, NdsN, nFRing, nBridgehead, MATS1Z, ATSC4p, ATSC5c, SlogP_VSA1, ATSC8dv, n5ARing, n7HRing, C3SP3, NdssC, ATSC4c, PEOE_VSA8, ATSC6c, ATSC3pe, SsNH2, SssNH, ATSC5se, PEOE_VSA13, n9FHRing, SRW09, BCUTd-1l, nBase, Xch-3dv, nSpiro, PEOE_VSA12, n8HRing, PEOE_VSA2, SddC, VSA_EState9, n6FARing, PEOE_VSA3, JGI4, n5FAHRing, ATSC7c, NssN, ATSC8c, BCUTc-1l, VSA_EState8, n5FARing, NdNH, ATS4d, ATSC6d
SCorr	SMR, piPC1, LabuteASA, AMID_O, ETA_epsilon_4, ETA_dEpsilon_B, BCUTse-1l, ETA_beta, SlogP_VSA2, ETA_epsilon_5, EState_VSA8, VSA_EState5, VSA_EState7, SMR_VSA9, GATS1p, FilterItLogS, SlogP_VSA4, SMR_VSA7, TpiPC10, ATS4p, NssO, BCUTm-1h, MIC5, EState_VSA2, ATS3are, ATSC2se, PEOE_VSA9, ATS3dv, ATS3Z, NsOH, ZMIC0, PEOE_VSA6, VSA_EState4, GATS2p, SlogP_VSA3, ATS4are, SIC1, C2SP2, nH, VE3_Dzp, PEOE_VSA7, ECIndex, RNCG, ETA_eta_RL, WPath, ZMIC1, nAHRing, EState_VSA1, C3SP2, AATS2Z, EState_VSA10, BCUTc-1h, SaasC, SMR_VSA6, nBondsKD, ZMIC2, ZMIC4, VE2_Dzare, ETA_shape_x, SssssC, ATSC3c, BCUTd-1l, GATS1Z, SdsCH, ETA_dPsi_B, C1SP3, EState_VSA3, ATSC2dv, SdssC, BIC5, SpAbs_Dzare, ATS8p, AETA_beta_ns_d, piPC8, GGI5, SaaN, EState_VSA7, nFHRing, ETA_shape_y, PEOE_VSA10, VR1_A, VE2_A, n5AHRing, NssCH, AMID_N, Radius, AATSC1dv, SMR_VSA5, JGI1, SlogP_VSA5, EState_VSA5, EState_VSA9, IC2, VSA_EState6, SssCH, PEOE_VSA4, GGI6, SaaNH, ATSC5dv, ATS6dv, Xc-5d, SaasN, ATSC4d, JGI2, EState_VSA4, n7FHRing, ATSC2d, NdO, ATS7dv, MIC1, SpAD_A, SsCH3, C4SP3, SlogP_VSA10, ATSC4se, n6AHRing, SMR_VSA4, n4AHRing, SlogP_VSA11, NssCH2, SaaO, n6aRing, GGI4, nRot, ATSC3m, PEOE_VSA11, ATSC7i, ATSC6Z, VE3_A, ATSC2Z, SlogP_VSA1, n3HRing, GGI7, GATS2se, GATS1se, n8FAHRing, ATSC7dv, JGI3, nBondsD, RPCG, ATS8dv, BalabanJ, n6FHRing, BCUTc-1l, NdsN, ATSC4c, MATS1Z, nBridgehead, ATSC4p, EState_VSA6, C1SP2, SssNH, ATS4dv, ATSC6d, n5Ring, ETA_eta_R, ATSC6c, ATS4d, ATS4Z, ATSSi, nBase, n9Ring, SsF, SsNH2, NssN, ATSC5c, n7AHRing, PEOE_VSA3, ATSC3dv, nSpiro, ATSC3se, VSA_EState8, PEOE_VSA8, VSA_EState9, PEOE_VSA13, n6FRing, n5FRing, PEOE_VSA2, n9FRing, n8AHRing, NdC, JGI4, NdNH, NdssC
RF	ATSC0i, ETA_dEpsilon_B, BCUTZ-1l, nC, ZMIC0, VSA_EState5, nO, C2SP1, SlogP_VSA2, ATS0dv, SMR_VSA7, nBondsKD, ABC, nFRing, BertzCT, BCUTi-1h, nN, ATSC1se, RNCG, BCUTZ-1h, VE2_A, GATS1se, BCUTd-1l, SpAbs_DzZ, ATS4v, SdssC, ATS6d, ATS7d, nBondsD, SpAbs_Dzv, BCUTdv-1h, VE1_A, SpAbs_A, ATS5se, BCUTd-1h, IC1, NdsN, EState_VSA8, VSA_EState7, VSA_EState4, FilterItLogS, BCUTc-1h, NdsCH, RPCG, VR1_A, EState_VSA9, BCUTc-1l, C2SP2, ATSC4se, ETA_eta_R, ZMIC1, n5Ring, ATS5d, SlogP_VSA4, ATSC2d, ATS3Z, C3SP2, BalabanJ, ATS1dv, PEOE_VSA7, ATSC7d, ATS3dv, ATSC8dv, ECIndex, nRot, nBridgehead, SlogP_VSA1, ATSC6d, ATSC5d, NssCH, PEOE_VSA6, ATSC2dv, nSpiro, ATSC6dv, SMR_VSA4, ATS8d, ATS4d, nFHRing, JGI4, GGI3, ATSC4d, MIC2, ATSC2se, EState_VSA4, ATS4dv, GATS2c, Xch-3d, VSA_EState6, NdO, ATS8Z, BCUTdv-1l, SlogP_VSA5, nBondsT, ATSC2c, SssssC, EState_VSA10, ATSC6c, EState_VSA7, SMR_VSA5, ATSC3Z, ATS6dv, n5FRing, ATSC5se, ATSC5dv, ATSC3dv, ATS5dv, NdssC, GGI4, ATSC3se, n6FRing, ATSC5c, ATSC4dv, PEOE_VSA9, nHRing, nBase, ATSC4se, ATSC2Z, ATSC3c, Xc-5d, ATSC1dv, PEOE_VSA3, PEOE_VSA8, VSA_EState8, ATSC4c, nAromAtom, EState_VSA3, n6aRing, ETA_shape_p, PEOE_VSA11, ZMIC2, C1SP2, EState_VSA5, PEOE_VSA2, ATSC7dv, ATSC7c, IC3, piPC8, PEOE_VSA10, NtCH, n5HRing
CFS	SMR, apol, ATS6v, Mp, ETA_dEpsilon_B, BertzCT, VMcGowan, Sp, BCUTi-1l, C2SP1, SdssC, nN, ATS0v, ATS8v, PEOE_VSA7, ATS4v, ATSC7d, ATS0p, SMR_VSA7, StsC, nSpiro, VSA_EState2, ATSC1d, AATS1i, AATS0p, Radius, ETA_dEpsilon_C, SM1_Dzse, ETA_psi_1, ETA_dAlpha_B, AETA_alpha, ETA_epsilon_2, SM1_Dzp, Sse, PEOE_VSA6, NdsN, ETA_alpha, VSA_EState5, SMR_VSA3, n6aRing, BCUTp-1h, SssCH2, BIC1, CIC5, AATS0v, ATS6are, ATSC5d, ATS8p, NdsCH, PEOE_VSA8, C3SP2, SlogP_VSA4, ATS6i, fMF, NtsC, BCUTm-1l, SlogP_VSA5, IC1, ATSC1c, SlogP_VSA10, LabuteASA, VE2_Dt, FilterItLogS, ATSC2p, GATS2p, AATSC1dv, ATSC3dv, Sv, NdNH, AATSC2p, GATS2i, CIC1, ATS7are, ATS4p, BIC5, CIC4, BIC4, CIC3, BIC3, CIC2, TIC1, SM1_Dzv, SdsN, ATSC4se, ATSC5se, MATS2v, ECIndex, ATSC6are, RPCG, SlogP_VSA1, nC, ATSC3d, Xc-5dv, SlogP_VSA3, GATS1Z, Mv, VSA_EState9, n7Ring, BCUTc-1l, BCUTc-1h
LASSO	VR1_A, nAromAtom, nSpiro, nBridgehead, ATS4dv, ATS1d, ATS2d, ATS3d, ATS4d, ATS5d, ATS6d, ATS0m, ATS1m, ATS2m, ATS3m, ATS5m, ATS6m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS3se, ATS4se, ATS0i, ATS1i, ATS2i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATSC1dv, ATSC2d, ATSC3d, ATSC6d, ATSC0m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC0v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC6v, ATSC7v, ATSC8v, ATSC3i, BCUTv-1l, SpDiam_Dzp, SpDiam_Dzi, BertzCT, nBondsA, DetourIndex, SpDiam_D, SssCH2, StCH, SaaCH, SdssC, StN, SdsN, SdO, SssO, SaaO, ECIndex, fragCpx, TIC1, TIC4, TIC5, PEOE_VSA1, PEOE_VSA3, PEOE_VSA4, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA12, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA3, SlogP_VSA5, SlogP_VSA6, SlogP_VSA11, EState_VSA1, EState_VSA3, EState_VSA4, EState_VSA5, EState_VSA8, EState_VSA9, VSA_EState2, VSA_EState4, VSA_EState7, MPC3, MPC5, MPC6, piPC6, nARing, n5ARing, n5AHRing, nRot, TopoPSA(NO), TMWC10, SRW09

A.3.8 Energia HOMO

Metodos	Atributos seleccionados
Pcorr	piPC4, SpMAD_Dzi, NaasC, ETA_beta_ns_d, SMR_VSA7, SlogP_VSA10, ATS1i, SaaNH, ETA_epsilon_5, nO, fMF, SlogP_VSA11, ATS4dv, SMR, AATSC0i, ATS5dv, SMR_VSA3, ATS2dv, C1SP1, nRot, GATS1v, SaaN, EState_VSA9, PEOE_VSA8, ATSC3dv, n5HRRing, GATS1are, NtN, EState_VSA2, nBondsKD, SpMAD_A, n5Ring, ATSC6dv, GATS2v, SRW09, ATSC4d, NdO, VE1_A, BalabanJ, EState_VSA8, ETA_shape_y, VSA_EState4, BCUTdv-1h, SlogP_VSA1, ATSC1d, VSA_EState9, C3SP2, ATS6dv, ATSC7dv, ATSC6d, BCUTc-1h, ATSC4v, piPC8, SaaN, SaaO, NssO, ATSC4dv, AATSC2m, ETA_dEpsilon_B, BCUTse-1h, EState_VSA1, AATSC2dv, EState_VSA10, FilterItLogS, nHRRing, SsssN, ATSC1dv, SsNH2, ATS4d, SMR_VSA5, nFRing, VE2_A, C1SP2, NssCH2, SIC1, PEOE_VSA10, SdssC, SlogP_VSA5, ZMIC5, ATS3dv, EState_VSA4, ATS4are, JGI1, SsssC, ATS7m, ZMIC0, AATSC2d, SssNH, AETA_eta_R, AATSC2c, EState_VSA7, GATS1c, BCUTc-1l, Radius, C2SP3, BCUTd-1l, GGI5, VSA_EState3, GGI6, nFAHRRing, EState_VSA5, ATS7dv, n8FARing, IC5, SLogP, ATSC7d, BCUTi-1h, IC2, ATSC7c, ATSC3m, EState_VSA3, RNCG, AATSC1se, EState_VSA6, PEOE_VSA7, GATS2se, C1SP3, SdsCH, PEOE_VSA9, BCUTp-1h, Xc-4d, ATS4v, ATSC4c, SlogP_VSA4, ATSC6c, ZMIC1, VSA_EState6, JGI3, nBondsD, n3HRRing, nBase, SsCH3, n9FRing, n7FARing, n4AHRing, ATSC5dv, n6aRing, RPCG, nAcid, ATSC5d, PEOE_VSA3, JGI4, SMR_VSA6, ATS5are, GGI7, GGI4, PEOE_VSA11, nBridgehead, ATSC8are, VSA_EState7, ATS8dv, n6FARing, PEOE_VSA13, n6Ring, VSA_EState5, Xch-3dv, SdsN, n6FHRing, ECIndex, PEOE_VSA12, ATSC3pe, SMR_VSA4, n5FARing, NdssC, PEOE_VSA2, NdNH, ATSC8i, VR1_A, ATSC8c, nSpiro, ATSC5c, PEOE_VSA6, C3SP3, n7ARing, n9Ring, MIC2, ATSC5se, ATSC3c, ATSC5i, NddC, VSA_EState8, n8Ring, n5FAHRRing
Scorr	piPC4, SpMAD_Dzi, ATS1i, BCUTd-1l, SMR_VSA7, ATS4dv, AATSC0i, NaasC, VE2_Dt, ETA_epsilon_5, EState_VSA9, GATS1v, nRot, AETA_beta_ns_d, BCUTm-1h, SpMAD_Dzv, SpMAD_A, ATS2dv, ATSC3dv, C2SP2, EState_VSA2, n5Ring, C1SP1, ATS1dv, VSA_EState4, ZMIC0, SMR_VSA3, BCUTc-1h, NdO, SlogP_VSA3, SMR, SsssCH, EState_VSA8, VE2_D, VE1_A, n5HRRing, ATS6dv, SaaNH, PEOE_VSA10, EState_VSA1, piPC8, ZMIC5, VE2_A, SlogP_VSA10, SaaN, SMR_VSA2, ETA_shape_y, PEOE_VSA8, SdssC, ATSC4d, ATSC6dv, ATS4d, BalabanJ, ATS3dv, C3SP2, SlogP_VSA11, GATS2v, SaaN, AATSC1se, nFHRing, SsssN, ATSC6d, NssO, nHRRing, EState_VSA4, SsssC, ATSC1dv, ATSC4v, RNCG, ATSC3m, Radius, BCUTp-1h, AATSC2m, GGI5, ATS7dv, NtCH, AATSC2dv, ETA_dEpsilon_B, VSA_EState9, GGI6, SlogP_VSA1, ATS4are, ZMIC2, ATS4v, nBondsKD, GATS1are, SlogP_VSA5, SaaO, n8FARing, MIC5, EState_VSA10, ATS5are, JGI1, FilterItLogS, JGI3, ATSC7dv, SLogP, AETA_eta_RL, NssCH2, EState_VSA5, VSA_EState5, VR2_A, EState_VSA3, PEOE_VSA7, ATS8m, SssNH, n7FARing, C1SP3, ATSC6c, ATSC5dv, AATSC2pe, n9FARing, SMR_VSA5, n3HRRing, GATS2se, n4HRRing, NsOH, CIC5, AATSC2c, SsCH3, PEOE_VSA9, PEOE_VSA11, VSA_EState6, nBase, VSA_dPsi_B, n6FRing, BCUTc-1l, nBridgehead, JGI4, ATSC3pe, ZMIC1, nBondsD, SsNH2, GGI4, CIC2, IC1, GGI7, n6Ring, PEOE_VSA3, n6FHRing, AATSC2d, SlogP_VSA4, SMR_VSA4, C2SP3, EState_VSA7, nSpiro, ATS8dv, PEOE_VSA13, Xc-6dv, n6aRing, RPCG, SdsN, MIC1, nAcid, C3SP3, ECIndex, ATSC4c, NssCH, NdNH, C4SP3, EState_VSA6, ATSC5se, n7ARing, n5FAHRRing, ATSC3c, C1SP2, n9ARing, VSA_EState8, VE1_Dzp, VE2_Dzp, ATSC8c, PEOE_VSA6, NdsCH, LabuteASA, GATS1Z, PEOE_VSA12, NdssC, n8Ring, NddC, SMR_VSA6, PEOE_VSA2, n5FARing, ATSC5c, SsF, nH
RF	ETA_beta_ns_d, BCUTd-1l, nAromAtom, SMR_VSA7, SlogP_VSA10, BCUTZ-1h, ATSC0i, BCUTi-1h, GATS1se, BertzCT, BCUTc-1l, n3HRRing, ATSC1se, nFRing, SdssC, NtN, BCUTdv-1h, nO, BCUTc-1h, VSA_EState9, ABC, ZMIC0, RNCG, SlogP_VSA2, GATS2c, IC1, nBase, nRot, RPCG, BCUTd-1h, Xch-3d, BCUTZ-1l, VSA_EState6, SlogP_VSA1, EState_VSA9, SpAbs_DzZ, ATSC2dv, NaaN, ATS1dv, ATSC2c, ATS3dv, ATSC2se, VSA_EState4, BCUTdv-1l, ATSC4d, VE2_A, SpAbs_A, VE1_A, BalabanJ, ATS4dv, NssN, ATSC3dv, ATSC3c, PEOE_VSA8, VSA_EState7, VR1_A, ATSC1dv, ETA_dEpsilon_B, PEOE_VSA2, VSA_EState5, EState_VSA8, nN, SpAbs_Dzv, ATSC3se, VSA_EState8, ATSC2d, ATS5se, PEOE_VSA10, FilterItLogS, ATSC4c, ATS4v, GGI3, ZMIC1, ATSC5c, ATS4se, SlogP_VSA5, ATSC6c, ATSC4dv, ATSC3Z, ETA_eta_R, MIC2, JGI4, nBridgehead, ATS3Z, ATSC4se, ATSC5d, ATS0dv, PEOE_VSA9, nBondsT, ATSC5dv, NdssC, ZMIC2, PEOE_VSA7, PEOE_VSA3, ATSC5se, ATS4d, ATSC2Z, PEOE_VSA11, ATSC6dv, SlogP_VSA4, ATSC6d, Xc-5d, ATS5d, SMR_VSA5, GGI4, EState_VSA4, nBondsKD, C2SP2, IC2, EState_VSA3, PEOE_VSA6, SMR_VSA6, nFHRing, IC3, ATS6d, ATSC7c, EState_VSA2, nBondsD, SsssC, nC, nHRRing, n4HRRing, EState_VSA5, ECIndex, PEOE_VSA13, SMR_VSA4, SlogP_VSA11, n5Ring, n5HRRing, ATS5dv, EState_VSA7, ATSC7dv, ETA_shape_p, NaaO, NssNH, ATSC7d, piPC8, PEOE_VSA12, C2SP3, EState_VSA6
CFS	piPC4, AETA_beta_ns_d, SMR_VSA7, ATS1i, VSA_EState9, GATS1Z, SpDiam_Dt, VSA_EState2, GATS1se, piPC2, ATSC3i, SdssC, PEOE_VSA8, ETA_beta_ns_d, GGI2, VE2_Dzp, ETA_dEpsilon_B, ETA_beta_s, NdO, SaaC, BCUTc-1l, GATS2se, BCUTse-1h, ETA_dPsi_B, RNCG, ETA_dEpsilon_D, GATS2c, nBondsD, BertzCT, SlogP_VSA10, VE2_Dt, VE3_Dzp, SlogP_VSA11, n4HRRing, NdsN, PEOE_VSA7, GATS1are, SdO, PEOE_VSA9, VSA_EState5, BCUTse-1l, ETA_dEpsilon_C, ETA_epsilon_2, ATSC1c, BIC1, n5HRRing, nH, EState_VSA10, n6AHRing, StsC, Xch-5d, n5aRing, C2SP2, SlogP_VSA2, EState_VSA5, SlogP_VSA1, AATSC0p, SpMAD_Dzp, MATS2Z, GATS2v, GATS2i, fMF, Vabc, GATS2i, AATSC2i, Xch-4d, nAHRing, ATSC3p, AETA_eta, VMcGowan, ATS2i, GATS1m, nO, MATS1c, NsNH2, ATSC4i, NdssC, Xch-6d, AETA_eta_FL, AMID, EState_VSA8, NsOH, ATSC3v, C4SP3, GATS1i, C3SP3, ATSC4c, ATSC5c, ATSC6se, ATSC3d, ATSC5v, ATSC7dv, ATSC6p, ATSC4dv, ATSC3se, ATSC2Z, EState_VSA4, SpMAD_Dzv, ATSC0p, AETA_eta_R, C2SP3, GATS1v, AATSC1v, EState_VSA7, SMR_VSA6, SM1_Dzv, PEOE_VSA1, MATS2pe, AATSC1c, nHBDOn, EState_VSA3, ATSC2d, SsssN, SlogP_VSA4, SlogP_VSA3, Xc-6d, SlogP_VSA6, NssNH, SaaNH, AETA_eta_L, ETA_shape_p, NssCH2, EState_VSA6, ETA_shape_y, C2SP1, BCUTm-1l, VSA_EState7, EState_VSA2, SaaCH, nBondsS, AETA_eta_F, VE2_Dzv, VR1_Dzare, ATS7are, ATS4se, NdsCH, ATS6i, ATSC5m, SaaO, nBondsKS, NtsC, BCUTi-1l, SMR, EState_VSA1, VE2_D, VE1_Dt, ATSC4pe, ATSC6dv, AMID_X, ATSC8dv, NaasN, ATSC5dv
LAGSSO	VR1_A, ATS2dv, ATS2d, ATS3d, ATS0m, ATS1m, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS2se, ATS4se, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS8i, ATSC3dv, ATSC1d, ATSC4d, ATSC2m, ATSC4m, ATSC5m, ATSC7m, ATSC8m, ATSC0v, ATSC1v, ATSC2v, ATSC3v, ATSC5v, ATSC7v, ATSC8v, ATSC1i, ATSC3i, ATSC4i, SpAD_Dzp, SpAD_Dzi, BertzCT, DetourIndex, TIC1, TIC2, TIC5, PEOE_VSA1, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, SMR_VSA3, SMR_VSA4, SMR_VSA5, SMR_VSA7, SlogP_VSA2, SlogP_VSA4, SlogP_VSA5, SlogP_VSA11, EState_VSA2, EState_VSA3, EState_VSA4, EState_VSA7, VSA_EState2, VSA_EState3, TMPC10, TopoPSA(NO), Vabc, TSRW10

A.3.9 Diferença de energias do HOMO e LUMO

Metódos	Atributos selecionados
Pcorr	nBondsKD, nBondsD, FCSP3, VSA_EState2, nBondsKS, C2SP2, TIC1, NdssC, NdsCH, ETA_eta_R, SMR_VSA7, ZMIC1, C1SP2, PEOE_VSA9, SssO, nAHRing, PEOE_VSA2, NssCH, ATS4v, GATS1se, EState_VSA10, nFHRing, EState_VSA9, ATS4Z, SMR_VSA5, PEOE_VSA11, GATS2pe, SIC2, GATS2Z, nFRing, FilterItLogS, C2SP3, MIC2, LabuteASA, EState_VSA7, ATSC2dv, SdssC, ATS5pe, TSRW10, PEOE_VSA6, SpMAD_A, C3SP2, ETA_dEpsilon_B, ATSC5dv, BalabanJ, n3AHRing, nBridgehead, SsCH2, SlogP_VSA2, PEOE_VSA1, WPol, ZMIC2, C3SP3, Radius, ATSC3Z, VE2_A, n5AHRing, BCUTd-1l, nRot, JGI3, n4HRing, AATS2se, NsCH3, bpol, BCUTc-1h, EState_VSA6, VR2_A, nSpiro, ETA_beta_ns_d, ATSC5d, SsOH, PEOE_VSA10, ATSC3dv, n9FAHRing, n7FAHRing, ATSC3pe, ATSC1se, ATSC4d, GGI4, ATSC3dv, n8FHRing, RPCG, ATSC4dv, C1SP1, ATSC4p, SaaN, n6FAHRing, PEOE_VSA12, SMR_VSA4, AMID_O, n5FARing, ATSC6se, n5ARing, n5FHRing, SlogP_VSA10, SlogP_VSA4, EState_VSA5, SaaO, MID_N, SaaNH, SMR_VSA6, TopoPSA(NO), VSA_EState9, VSA_EState7, SssssC, VSA_EState4, RNCG, NdsN, ATSC0i, EState_VSA8, SlogP_VSA3, n6aRing, NtN, SlogP_VSA11, ATSC4d, PEOE_VSA5, piPC8, EState_VSA4, ATSC7p, SlogP_VSA1, ZMIC0, BCUTse-1l, SlogP_VSA5, VSA_EState5, n6FRing, NaasN, GGI5, EState_VSA3, SpMAD_Dzp, IC3, SdNH, n6AHRing, n7Ring, BCUTdv-1h, GATS1c, ATSC7d, ATSC4i, SsNH2, BCUTp-1l, PEOE_VSA3, C1SP3, ATSC8p, VSA_EState6, ATSC3c, n8Ring, ATSC6p, ATSC4c, ATSC7dv, NssNH, n9Ring, C4SP3, ATSC6c, NsF, GGI6, EState_VSA1, JGI4, PEOE_VSA7, SssN, ECIndex, BCUTi-1h, ATSC7c, ATSC8c, PEOE_VSA8, NddC, ATSC5c, nBase, ATSC6d, EState_VSA2, ATSC8dv, BCUTc-1l, PEOE_VSA13, nAcid, VSA_EState8, GGI7
SCorr	nBondsKD, nBondsD, nBondsM, VSA_EState2, C2SP2, NdsCH, CIC1, MIC1, NdssC, C1SP2, ETA_eta_R, SMR_VSA7, PEOE_VSA9, nAHRing, ZMIC1, SssO, BCUTdv-1h, GATS1se, ATSC4v, NssCH, nFHRing, ATSC4Z, PEOE_VSA2, BCUTd-1l, GATS2se, EState_VSA9, SssCH, TIC2, FilterItLogS, PEOE_VSA11, GATS2Z, SdssC, SlogP_VSA3, ETA_dEpsilon_B, ETA_shape_x, VR2_A, C2SP3, SpMAD_A, ATSC2dv, EState_VSA10, EState_VSA7, VE2_Dzpe, BCUTi-1h, SlogP_VSA2, C3SP2, PEOE_VSA10, MIC2, ATS5pe, PEOE_VSA6, ATSC1dv, LabuteASA, ATSC5dv, SMR_VSA5, n3HRing, BCUTc-1h, Xc-6dv, C3SP3, SssCH2, BCUTdv-1l, nH, nBridgehead, Radius, WPol, ZMIC2, ATSC3Z, nRot, MATS2se, n4HRing, AATS2pe, MPC8, AMID_O, NaasC, VE2_A, EState_VSA6, bpol, ETA_beta_ns_d, n5AHRing, PEOE_VSA1, nSpiro, SsOH, JGI3, ATSC2v, BCUTc-1l, piPC1, n7FHRing, n9FAHRing, AATSC1dv, ATSC3pe, RPCG, GGI4, VE3_Dzp, ATSC3dv, AETA_eta_RL, BCUTse-1l, n8FHRing, ATSC4d, NaaN, EState_VSA5, ETA_shape_y, C1SP1, ATSC6se, GATS1m, ATSC3dv, SlogP_VSA5, ATSC4p, n6FAHRing, n5FHRing, SssC, SMR_VSA4, SaaO, n5FARing, SMR_VSA6, SlogP_VSA4, SlogP_VSA10, PEOE_VSA12, ATSC4dv, GATS1c, SaaNH, ATSC6dv, SpMAD_Dzp, n6aRing, VSA_EState5, ATSC4d, GGI5, SlogP_VSA11, NtN, StCH, NdsN, ATSC1c, VSA_EState9, VSA_EState7, NaasN, n6FRing, n7Ring, EState_VSA4, BIC5, n5ARing, ATSC6c, SlogP_VSA1, AMID_N, SdNH, ATSC7p, EState_VSA3, VE3_Dzpe, ATSC4c, ATSC4i, ATSC6i, SsNH2, ATSC3c, EState_VSA1, n6AHRing, EState_VSA8, C1SP3, RNCG, VSA_EState6, ZMIC5, PEOE_VSA3, n8Ring, VSA_EState4, ATSC7dv, PEOE_VSA7, ATSC8are, ATSC8p, PEOE_VSA8, ATSC5c, C4SP3, GGI6, NssNH, n9Ring, nBase, nF, ATSC7c, ETA_dPsi_B, JGI4, VE1_A, SddC, SssN, ATSC6d, ZMIC0, ATSC8dv, ATSC4se, ECIndex, VSA_EState8, ATSC8are, GGI7, nAcid, EState_VSA2, PEOE_VSA13
RF	nBondsKD, nBondsD, SMR_VSA7, C2SP2, BCUTZ-1h, BCUTi-1h, ETA_dEpsilon_B, BCUTdv-1h, nBondsT, C1SP2, ATSC2se, BCUTc-1l, BCUTd-1l, BCUTZ-1l, nN, NdO, ATSC0i, NdsCH, IC1, BCUTc-1h, ATSC2c, PEOE_VSA12, ABC, GATS1se, PEOE_VSA11, PEOE_VSA2, ZMIC1, SlogP_VSA10, GATS2c, RNCG, nBase, RPCG, BertzCT, ATSC3dv, VE2_A, BCUTd-1h, ZMIC0, nAromAtom, ATSC1dv, PEOE_VSA10, n4HRing, n3HRing, SlogP_VSA2, ATSC2dv, PEOE_VSA3, ETA_eta_R, PEOE_VSA6, ATSC3se, C2SP1, EState_VSA8, SlogP_VSA1, VSA_EState5, ATSC3c, VSA_EState4, BCUTdv-1l, VE1_A, GGI3, ATSC4d, SpAbs_A, VSA_EState9, C3SP3, VSA_EState7, Xch-3d, SdssC, ATSC1dv, ATSC4v, VR1_A, ATSC4dv, SlogP_VSA4, SpAbs_Dzv, VSA_EState6, SMR_VSA4, SpAbs_DzZ, nC, FilterItLogS, PEOE_VSA9, PEOE_VSA7, NssO, ATSC3dv, ATSC4c, MIC2, JGI4, nBridgehead, ATSC3Z, ATSC2d, PEOE_VSA8, ATSC5se, VSA_EState8, ATSC1se, NdssC, ATSC5dv, ATSC4dv, BalabanJ, ATSC3Z, ATSC5d, ATSC0dv, ATSC4d, ATSC4se, ATSC5c, ATSC5se, ATSC4se, Xc-5d, nHRing, ATSC6dv, ATSC6c, NdsN, EState_VSA9, ZMIC2, NssN, ETA_beta_ns_d, ETA_shape_p, ATSC6d, ATSC2Z, ATSC5d, n5Ring, nO, GGI4, EState_VSA4, SssssC, ATSC5dv, EState_VSA3, SMR_VSA5, ATSC6d, ECIndex, SMR_VSA6, n5HRing, IC3, EState_VSA10, IC2, NtN, C2SP3, EState_VSA2, ATSC7dv, SlogP_VSA5, EState_VSA7, n6FRing, ATSC7c, EState_VSA6, nRot, EState_VSA5
CFS	FCSP3, GATS1se, CIC1, C1SP1, ATSC1are, MATS2c, VSA_EState1, GATS1m, GATS2m, Xc-3dv, mZagreb2, nBondsKS, ETA_dBeta, BCUTare-1h, BCUTp-1l, BCUTpe-1h, ATSC1i, StsC, VE2_A, BCUTc-1h, AETA_eta_L, AETA_eta_FL, AETA_eta, AETA_beta_ns, FilterItLogS, MATS1se, ETA_dEpsilon_D, SssO, nH, VR1_A, PEOE_VSA4, PEOE_VSA5, ATSC1se, VE2_D, ETA_eta_R, ZMIC1, ATSC2c, n3HRing, PEOE_VSA9, VE2_Dzare, n4Ring, SaaCH, SpMax_Dzi, n5AHRing, RPCG, ATSC3c, VE2_DzZ, LabuteASA, VE2_Dzse, Sm, StCH, AETA_beta_s, SlogP_VSA5, SMR_VSA2, PEOE_VSA1, VE2_Dt, SMR_VSA4, ETA_eta_FL, fragCpx, C2SP1, BCUTse-1l, ETA_dEpsilon_B, ETA_epsilon_4, SlogP_VSA_EState3, ATSC1Z, GATS1Z, SlogP_VSA3, PEOE_VSA8, ETA_shape_x, PEOE_VSA12, NssCH, Vabc, C3SP2, PEOE_VSA3, NssCH2, SsOH
LASSO	VR1_A, ATSC2dv, ATSC1d, ATSC2d, ATSC3d, ATSC4d, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC6v, ATSC7v, ATSC8v, ATSC2se, ATSC4se, ATSC0i, ATSC2i, ATSC3i, ATSC4i, ATSC5i, ATSC6i, ATSC7i, ATSC8i, AATS2v, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC7m, ATSC0v, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC7v, ATSC2i, SpAbs_Dzare, SpDiam_Dzare, SpAD_Dzi, BertzCT, DetourIndex, SdsCH, StN, SdO, SssO, ECIndex, fragCpx, nHBAcc, TIC1, TIC2, PEOE_VSA2, PEOE_VSA4, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, SMR_VSA3, SMR_VSA4, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA3, SlogP_VSA5, EState_VSA2, EState_VSA3, EState_VSA4, EState_VSA6, EState_VSA8, EState_VSA9, MPC7, MPC8, TMPC10, TopoPSA(NO), Vabc, TSRW10

A.4 QM9

A.4.1 Energia interna de atomização para a temperatura de 0 K

Metódos	Atributos selecionados
Pcorr	apol, SlogP_VSA5, SLogP, PEOE_VSA7, VSA_EState8, SssCH2, ATS4Z, BCUTdv-1l, ATS3Z, PEOE_VSA6, BCUTi-1h, GATS2m, C2SP3, ATSC0i, FilterItLogS, AATSC0i, ETA_dEpsilon_B, ATS4dv, C1SP2, NaaN, AATS0d, BertzCT, AETA_beta_s, TIC0, ATSC3v, EState_VSA4, BCUTp-1l, SMR_VSA4, C3SP3, ZMIC1, WPath, SlogP_VSA1, PEOE_VSA3, IC5, fMF, PEOE_VSA4, ATSC6d, GATS2v, VSA_EState2, TIC1, EState_VSA5, PEOE_VSA11, GATS1pe, AETA_beta_ns_d, PEOE_VSA10, RPCG, VSA_EState7, BCUTc-1l, C4SP3, EState_VSA1, NaaO, AATSC1i, VR1_A, PEOE_VSA12, ECIndex, NssCH, SlogP_VSA10, PEOE_VSA13, ATSC4p, ATSC4d, ATSC6i, PEOE_VSA2, SssssC, ZMIC5, GATS2pe, ATSC7d, RNCG, ATSC2d, SMR_VSA7, nBondsD, ATSC7i, ATSC1Z, SdssC, ATSC1dv, EState_VSA10, n6aHRing, GGI2, nHRing, n5HRing, BCUTv-1h, EState_VSA9, ATSC3dv, VSA_EState4, nRot, NsNH2, SMR_VSA9, SlogP_VSA11, NaaNH, PEOE_VSA1, C1SP1, VSA_EState5, SMR_VSA2, VSA_EState3, GGI3, EState_VSA2, ATS5dv, Xch-3dv, VSA_EState1, ATSC2dv, NdssC, SlogP_VSA3, ATSC8i, NdsN, TSRW10, ATSC5d, SlogP_VSA2, ATSC3pe, nFaRing, NaasN, SlogP_VSA4, EState_VSA8, VSA_EState6, ATSC5dv, n6HRing, PEOE_VSA8
Scorr	Sp, SlogP_VSA5, ATSC0i, SpAbs_Dzare, PEOE_VSA7, BCUTdv-1l, SLogP, VSA_EState8, SssCH2, AATSC0i, BCUTi-1h, ZMIC0, ATS4Z, PEOE_VSA6, AATSC1se, ETA_epsilon_5, ATS3Z, C2SP3, GATS2m, BCUTdv-1h, FilterItLogS, ATS4dv, SssCH, C1SP2, BertzCT, EState_VSA4, AETA_beta_s, C3SP3, IC5, NaaN, fMF, ZMIC5, ATSC3v, SMR_VSA4, RNCG, SlogP_VSA1, ETA_dEpsilon_B, ATSC6d, ZMIC1, PEOE_VSA3, PEOE_VSA11, PEOE_VSA10, VSA_EState2, VE2_Dt, VE2_A, PEOE_VSA4, VE2_Dzm, TIC0, AETA_beta_ns_d, VSA_EState7, EState_VSA1, NssCH, GATS2v, ATSC6i, C4SP3, GATS1pe, PEOE_VSA12, EState_VSA5, SlogP_VSA10, ATSC4d, LogEE_Dzv, EState_VSA9, VR1_A, VSA_EState5, SssssC, NaaO, ATSC4p, PEOE_VSA13, ETA_dPsi_B, nHRing, BCUTd-1l, VE3_Dzp, SMR_VSA7, PEOE_VSA2, TIC1, ATSC1Z, ETA_eta_R, ATSC1dv, GATS2pe, AATSC1i, n5HRing, SMR_VSA9, VE3_D, nBondsD, RPCG, EState_VSA2, NsNH2, GGI2, BCUTv-1h, ATSC3dv, n6aHRing, SlogP_VSA11, PEOE_VSA1, Xc-6dv, ECIndex, NaaNH, SMR_VSA2, SlogP_VSA3, VSA_EState3, AATSC2i, SdssC, ATSC7i, GGI3, SaasC, nRot, ETA_shape_y, AMID_X, Xch-3dv, ATSC3pe, NdssC, AATSC2c, ATSC8i, EState_VSA10, BCUTc-1l, Xp-1d, NdsN, ATSC3c, C2SP2, n6HRing, PEOE_VSA5, ATSC5d, NaasN
RF	nH, nC, ATSC0i, SlogP_VSA5, TIC0, SpAbs_DzZ, ATS0dv, TIC1, ATS0Z, AATS0i, PEOE_VSA7, SLogP, BCUTi-1h, VE2_A, nN, IC1, SpAbs_A, BCUTdv-1l, ATSC0dv, AATS0d, SlogP_VSA2, VR1_A, IC3, VE1_A, ATS4Z, nO, ETA_dEpsilon_B, PEOE_VSA6, ATS3Z, ABC, GATS1se, FilterItLogS, BCUTdv-1h, BertzCT, SMR_VSA7, nBase, ECIndex, RNCG, SpAbs_Dzv, ATSC1se, PEOE_VSA1, nBondsD, ATSC2se, n5Ring, VSA_EState6, ZMIC2, BCUTZ-1l, nF, C2SP2, ATSC2c, RPCG, BCUTd-1l, n4HRing, ATSC1v, ZMIC1, ATSC1Z, SpMAD_A, VSA_EState1, BCUTc-1l, VSA_EState5, NdO, IC2, EState_VSA8, BalabanJ, SlogP_VSA4, nBridgehead, n5FRing, BCUTd-1h, ATSC2d, ATS6d, SMR_VSA6, GATS2c, n6Ring, C2SP1, nAromAtom, Xc-5d, SdssC, VSA_EState7, NsCH3, SssssC, piPC8, GGI3, EState_VSA9, NdsCH, Xch-3d, ATSC2dv, EState_VSA10, VSA_EState4, SlogP_VSA3, nRot, ETA_shape_y, ATSC2Z, nHRing, ATSC4d, SaasC, ATSC1dv, ATSC3dv, nBondsT, ATSC2i, n5HRing, ATS4dv, NdssC, nSpiro, ETA_shape_p, n6aRing, ATSC3se, ATSC2v, ATSC5d, GGI4, ATSC3c, PEOE_VSA9, SlogP_VSA1, ATSC3Z, PEOE_VSA2, PEOE_VSA4, n5AHRing, n7FRing, JGI4, PEOE_VSA12, ATSC6d, NaaN, C1SP2, ATSC4Z, ATSC4c, NtN, PEOE_VSA3, PEOE_VSA8, ATSC4dv, NdsN, ATSC5c, SlogP_VSA10, ATSC4se, ATSC5se, GGI2, ATSC5dv, NssO, EState_VSA4, EState_VSA7, NsOH, n8FRing
CFS	SIC0, Mm, VE2_Dzare, AATS1Z, AATSC0v, AMID_h, CIC0, MZ, VE2_Dzpe, SIC1, AMW, FCSP3, AETA_eta_FL, ETA_dEpsilon_C, ETA_epsilon_4, VE2_Dzi, Mv, TopoPSA(NO), n4Ring, AETA_eta_F, VE2_D, AATS1m, n3Ring, AATSC0dv, C2SP1, VE2_A, SM1_Dzi, AATSC1c, BIC1, CIC1, AATS1pe, SM1_Dzp, AATSC0Z, AATS0v, GATS1m, BIC0, AETA_eta_L, ETA_dBeta, FilterItLogS, SlogP, MID_h, SM1_Dzv, ETA_dAlpha_B, AETA_alpha, ETA_dPsi_A, ETA_psi_1, AATS1are, VE2_Dt, ATS0dv, SpMax_A, ETA_eta_FL, VE2_Dzp, AATS0Z, AETA_eta_RL, JGI1, SpMAD_Dzm, AMID_N, AATSC0p, AATSC0m, VE2_Dzv, AATS1dv, Mp, ETA_dEpsilon_D, GATS1v, AMID, ETA_shape_p, GATS2m, AATS2Z, AATSC1se, TopoPSA, AMID_C, BCUTc-1h, VE2_DzZ, AATS0m, IC1, VE2_Dzse, ATS1dv, SMR_VSA6, SlogP_VSA4, GATS2i, AATSC2i, ATSC3i, ATSC4d, AATS2dv, VE2_Dzm, AATS0dv, BCUTv-1h, BCUTi-1l, SdssC, ATSC4i, ATSC3p, AATSC1d, ETA_dEpsilon_A, SssssC, BalabanJ, MATS1v, ATSC3d, MATS1m, NsOH, AATSC1dv, NdsCH, StsC, CIC2, BIC2, n4ARing, AATS0i, ATSC5i, ATSC4v, SaaCH, ATSC6v, ATSC3v, ATSC5v, ATSC4p, CIC3, StCH, BIC4, CIC4, NdssC, SRW07, ATSC2d, ATSC5d, GATS2Z, MATS2m, ATSC4dv, ATSC3Z, JGT10, SaasC, ETA_epsilon_1, ATSC1c
LASSO	VR1_A, nAromAtom, nBridgehead, ATS2dv, ATS2d, ATS4d, ATS5d, ATS4Z, ATS1m, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS2se, ATS4se, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, AATS2v, AATS8i, ATSC2dv, ATSC3dv, ATSC3d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC0v, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC7v, ATSC8v, ATSC1i, ATSC2i, ATSC3i, ATSC5i, BertzCT, SpDiam_Dt, DetourIndex, StCH, SaaCH, StN, SssO, SsF, ECIndex, ETA_beta, fragCpx, TIC0, TIC1, TIC3, TIC4, PEOE_VSA2, PEOE_VSA3, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, PEOE_VSA13, SMR_VSA1, SMR_VSA2, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA5, SlogP_VSA11, EState_VSA3, EState_VSA5, EState_VSA6, EState_VSA8, VSA_EState2, VSA_EState5, VSA_EState7, MPC3, MPC5, MPC7, MPC8, TopoPSA(NO), TSRW10

A.4.2 Energia interna de atomização para a temperatura de 298.15 K

Metódos	Atributos selecionados
Pcorr	apol, SlogP_VSA5, SLogP, PEOE_VSA7, VSA_EState8, SssCH2, ATS4Z, BCUTdv-1l, ATS3Z, PEOE_VSA6, BCUTi-1h, GATS2m, C2SP3, ATSC0i, FilterItLogS, ETA_dEpsilon_B, AATSC0i, ATS4dv, C1SP2, NaaN, AATS0d, BertzCT, AETA_beta_s, TIC0, ATSC3v, EState_VSA4, BCUTp-1l, SMR_VSA4, C3SP3, ZMIC1, WPath, SlogP_VSA1, PEOE_VSA3, IC5, PEOE_VSA4, fMF, ATSC6d, GATS2v, VSA_EState2, EState_VSA5, TIC1, PEOE_VSA11, GATS1pe, AETA_beta_ns_d, PEOE_VSA10, RPCG, VSA_EState7, BCUTc-1l, C4SP3, EState_VSA1, NaaO, AATSC1i, VR1_A, PEOE_VSA12, NsssCH, ECIndex, SlogP_VSA10, PEOE_VSA13, ATSC6i, ATSC4p, ATSC4d, PEOE_VSA2, SssC, ZMIC5, GATS2pe, ATSC7d, RNCG, ATSC2d, SMR_VSA7, nBondsD, ATSC7i, ATSC1Z, SdssC, ATSC1dv, EState_VSA10, n6aHRRing, GGI2, nHRRing, n5HRRing, BCUTv-1h, EState_VSA9, ATSC3dv, VSA_EState4, nRot, SMR_VSA9, NsNH2, SlogP_VSA11, NaaNH, C1SP1, PEOE_VSA1, SMR_VSA2, VSA_EState5, VSA_EState3, GGI3, EState_VSA2, ATS5dv, Xch-3dv, VSA_EState1, ATSC2dv, NdssC, SlogP_VSA3, ATSC8i, TSRW10, NdsN, ATSC5d, SlogP_VSA2, ATSC3pe, nFaHRRing, NaasN, VSA_EState6, EState_VSA8, SlogP_VSA4, ATSC5dv, n6HRRing, PEOE_VSA8
Scorr	Sp, SlogP_VSA5, ATSC0i, SpAbs_Dzare, PEOE_VSA7, BCUTdv-1l, SLogP, VSA_EState8, SssCH2, AATSC0i, BCUTi-1h, ZMIC0, ATS4Z, PEOE_VSA6, AATSC1se, ETA_epsilon_5, ATS3Z, C2SP3, GATS2m, BCUTdv-1h, FilterItLogS, ATS4dv, SssCH, C1SP2, BertzCT, EState_VSA4, AETA_beta_s, C3SP3, IC5, NaaN, fMF, ZMIC5, ATSC3v, SMR_VSA4, RNCG, SlogP_VSA1, ETA_dEpsilon_B, ATSC6d, ZMIC1, PEOE_VSA3, PEOE_VSA10, PEOE_VSA11, VSA_EState2, PEOE_VSA4, VE2_A, VE2_Dt, TIC0, VE2_Dzm, AETA_beta_ns_d, VSA_EState7, NsssCH, EState_VSA1, GATS2v, ATSC6i, C4SP3, GATS1pe, PEOE_VSA12, EState_VSA5, SlogP_VSA10, ATSC4d, EState_VSA9, LogEE_Dzv, VR1_A, SssC, VSA_EState5, NaaO, ATSC4p, PEOE_VSA13, ETA_dPsi_B, BCUTd-1l, nHRRing, VE3_Dzp, SMR_VSA7, PEOE_VSA2, ATSC1Z, TIC1, ETA_eta_R, GATS2pe, ATSC1dv, AATSC1i, n5HRRing, SMR_VSA9, VE3_D, nBondsD, RPCG, EState_VSA2, NsNH2, GGI2, BCUTv-1h, ATSC3dv, n6aHRRing, SlogP_VSA11, Xc-6dv, PEOE_VSA1, ECIndex, NaaNH, SMR_VSA2, SlogP_VSA3, VSA_EState3, AATSC2i, SdssC, ATSC7i, GGI3, SaasC, AMID_X, nRot, Xch-3dv, ETA_shape_y, ATSC3pe, NdssC, AATSC2c, EState_VSA10, ATSC8i, BCUTc-1l, Xp-1d, NdsN, ATSC3c, C2SP2, PEOE_VSA5, ATSC5d, n6HRRing, NaasN
RF	nH, nC, ATSC0i, SlogP_VSA5, TIC0, SpAbs_DzZ, ATS0dv, TIC1, ATS0Z, SLogP, AATS0i, VE2_A, BCUTi-1h, nN, IC1, SpAbs_A, PEOE_VSA7, ATSC0dv, BCUTdv-1l, VE1_A, SlogP_VSA2, AATS0d, IC3, nO, VR1_A, ATS4Z, ETA_dEpsilon_B, GATS1se, ABC, FilterItLogS, BertzCT, ATS6d, ECIndex, BCUTdv-1h, SMR_VSA7, RNCG, nBase, ATSC1se, nBondsD, ATSC2se, ATS3Z, PEOE_VSA1, SpAbs_Dzv, n5Ring, VSA_EState6, VSA_EState1, PEOE_VSA6, RPCG, BCUTZ-1l, ZMIC2, C2SP2, ATSC2c, ZMIC1, BCUTd-1l, n4HRRing, ATSC1v, VSA_EState5, ATSC1Z, SpMAD_A, BCUTc-1l, NsCH3, NdO, nBridgehead, BalabanJ, n5FRing, BCUTd-1h, n6Ring, EState_VSA8, SlogP_VSA4, IC2, GATS2c, ATSC2d, nAromAtom, SdssC, piPC8, Xc-5d, C2SP1, SlogP_VSA3, GGI3, Xch-3d, ATSC2i, EState_VSA9, nBondsT, ATSC2dv, VSA_EState7, EState_VSA10, SssC, nRot, NdsCH, VSA_EState4, ATSC1dv, ETA_shape_y, nHRRing, ATSC2Z, SaasC, NaaN, n6aRing, ATSC4d, ATSC3dv, n5HRRing, nSpiro, NdssC, ETA_shape_p, ATSC3se, nF, ATSC3c, ATSC5d, GGI4, ATS4dv, ATSC2v, ATSC3Z, PEOE_VSA12, PEOE_VSA4, n7FRing, n5AHRing, JGI4, PEOE_VSA9, PEOE_VSA2, SlogP_VSA1, EState_VSA1, ATSC6d, PEOE_VSA8, ATSC4c, ATSC4Z, GGI2, C1SP2, NtN, ATSC4dv, NdsN, EState_VSA2, ATSC4se, ATSC5se, ATSC5c, ATSC5dv, PEOE_VSA3, n8FRing, ATS7d, EState_VSA4, n3HRRing, PEOE_VSA11
CFS	SIC0, Mm, VE2_Dzare, AATS1Z, AATSC0v, AMID_h, CIC0, MZ, VE2_Dzpe, SIC1, AMW, FCSP3, AETA_eta_FL, ETA_dEpsilon_C, ETA_epsilon_4, VE2_Dzi, Mv, TopoPSA(NO), n4Ring, AETA_eta_F, VE2_D, AATS1m, n3Ring, AATSC0dv, C2SP1, VE2_A, SM1_Dzi, AATSC1c, BIC1, CIC1, AATS0Z, VE2_Dt, IC1, FilterItLogS, AETA_beta, AETA_eta_L, AATS1dv, CIC2, SIC2, CIC3, AETA_eta_RL
LASSO	VR1_A, nAromAtom, nBridgehead, ATS2dv, ATS2d, ATS4d, ATS5d, ATS4Z, ATS1m, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS2se, ATS4se, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, AATS2v, AATS2i, ATSC2dv, ATSC3dv, ATSC3d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC0v, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC7v, ATSC8v, ATSC1i, ATSC2i, ATSC3i, ATSC5i, BertzCT, SpDiam_Dt, DetourIndex, StCH, SaaCH, StN, SssO, SsF, ECIndex, ETA_beta, fragCpx, TIC0, TIC1, TIC3, TIC4, PEOE_VSA2, PEOE_VSA3, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, PEOE_VSA13, SMR_VSA1, SMR_VSA2, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA5, SlogP_VSA11, EState_VSA3, EState_VSA5, EState_VSA6, EState_VSA8, VSA_EState2, VSA_EState5, VSA_EState7, MPC3, MPC5, MPC7, MPC8, TopoPSA(NO), TSRW10

A.4.3 Entalpia de atomização para a temperatura de 298.15 K

Metódos	Atributos selecionados
Pcorr	apol, SlogP_VSA5, SLogP, PEOE_VSA7, VSA_EState8, SssCH2, ATS4Z, BCUTdv-1l, ATS3Z, PEOE_VSA6, BCUTi-1h, GATS2m, C2SP3, ATSC0i, FilterItLogS, ETA_dEpsilon_B, AATSC0i, ATSC4dv, C1SP2, NaaN, AATS0d, BertzCT, AETA_beta_s, TIC0, ATSC3v, EState_VSA4, BCUTp-1l, SMR_VSA4, C3SP3, ZMIC1, WPath, SlogP_VSA1, PEOE_VSA3, IC5, PEOE_VSA4, fMF, ATSC6d, GATS2v, VSA_EState2, EState_VSA5, TIC1, PEOE_VSA11, GATS1pe, AETA_beta_ns_d, PEOE_VSA10, RPCG, BCUTc-1l, VSA_EState7, C4SP3, EState_VSA1, NaaO, AATSC1i, VR1_A, PEOE_VSA12, NsssCH, ECIndex, SlogP_VSA10, PEOE_VSA13, ATSC6i, ATSC4p, ATSC4d, PEOE_VSA2, SssssC, ZMIC5, GATS2pe, ATSC7d, RNCG, ATSC2d, SMR_VSA7, nBondsD, ATSC7i, ATSC1Z, SdssC, ATSC1dv, EState_VSA10, n6aHRRing, GGI2, nHRRing, n5HRRing, EState_VSA9, BCUTv-1h, ATSC3dv, VSA_EState4, nRot, SMR_VSA9, NsNH2, SlogP_VSA11, NaaNH, C1SP1, PEOE_VSA1, SMR_VSA2, VSA_EState5, VSA_EState3, GGI3, EState_VSA2, ATS5dv, Xch-3dv, VSA_EState1, ATSC2dv, NdssC, SlogP_VSA3, ATSC8i, TSRW10, NdsN, ATSC5d, SlogP_VSA2, ATSC3pe, nFaRing, NaasN, VSA_EState6, EState_VSA8, SlogP_VSA4, ATSC5dv, n6HRRing, PEOE_VSA8
Scorr	Sp, SlogP_VSA5, ATSC0i, SpAbs_Dzare, PEOE_VSA7, BCUTdv-1l, SLogP, VSA_EState8, SssCH2, AATSC0i, BCUTi-1h, ZMIC0, ATS4Z, PEOE_VSA6, AATSC1se, ETA_epsilon_5, ATS3Z, C2SP3, GATS2m, BCUTdv-1h, FilterItLogS, ATS4dv, SssCH, C1SP2, BertzCT, EState_VSA4, AETA_beta_s, C3SP3, IC5, NaaN, fMF, ZMIC5, ATSC3v, SMR_VSA4, RNCG, SlogP_VSA1, ETA_dEpsilon_B, ATSC6d, ZMIC1, PEOE_VSA3, PEOE_VSA11, PEOE_VSA10, VSA_EState2, PEOE_VSA4, TIC0, VE2_A, VE2_Dt, VE2_Dzm, AETA_beta_ns_d, VSA_EState7, NsssCH, EState_VSA1, GATS2v, ATSC6i, C4SP3, GATS1pe, PEOE_VSA12, EState_VSA5, SlogP_VSA10, ATSC4d, LogEE_Dzv, EState_VSA9, VR1_A, SssssC, VSA_EState5, NaaO, ATSC4p, PEOE_VSA13, ETA_dPsi_B, BCUTd-1l, nHRRing, SMR_VSA7, VE3_Dzp, PEOE_VSA2, ATSC1Z, TIC1, GATS2pe, ETA_eta_R, ATSC1dv, AATSC1i, SMR_VSA9, n5HRRing, VE3_D, nBondsD, RPCG, EState_VSA2, NsNH2, GGI2, BCUTv-1h, ATSC3dv, n6aHRRing, SlogP_VSA11, Xc-6dv, PEOE_VSA1, ECIndex, NaaNH, SMR_VSA2, SlogP_VSA3, VSA_EState3, AATSC2i, SdssC, ATSC7i, GGI3, SaasC, nRot, AMID_X, Xch-3dv, ETA_shape_y, ATSC3pe, NdssC, AATSC2c, EState_VSA10, ATSC8i, BCUTc-1l, Xp-1d, NdsN, C2SP2, ATSC3c, PEOE_VSA5, ATSC5d, n6HRRing, NaasN
RF	nH, nC, ATSC0i, SlogP_VSA5, TIC0, SpAbs_DzZ, ATS0dv, TIC1, ATS0Z, AATS0i, BCUTi-1h, SLogP, VE2_A, SpAbs_A, nN, BCUTdv-1l, IC1, ATSC0dv, PEOE_VSA7, IC3, VE1_A, AATS0d, SlogP_VSA2, nO, VR1_A, ETA_dEpsilon_B, ATS4Z, PEOE_VSA6, ABC, ECIndex, BertzCT, GATS1se, nBase, BCUTdv-1h, SMR_VSA7, RNCG, ATSC1se, ATS3Z, ATSC2se, SpAbs_Dzv, nBondsD, nF, PEOE_VSA1, n5Ring, VSA_EState1, ZMIC2, FilterItLogS, RPCG, ZMIC1, BCUTZ-1l, VSA_EState6, C2SP2, ATSC2c, ATS6d, BCUTd-1l, n4HRRing, ATSC1v, VSA_EState5, ATSC1Z, SpMAD_A, BCUTc-1l, NdO, BalabanJ, nAromAtom, nBridgehead, n5FRing, BCUTd-1h, IC2, SlogP_VSA4, EState_VSA8, C2SP3, n6Ring, ATSC2d, GATS2c, SdssC, NsCH3, Xc-5d, C2SP1, GGI3, EState_VSA9, Xch-3d, VSA_EState7, ATSC2dv, EState_VSA10, NdsCH, SssssC, piPC8, nRot, SlogP_VSA3, VSA_EState4, ETA_shape_y, ATSC1dv, ATSC2Z, ATSC2i, nHRRing, ATSC4d, n5HRRing, n6aRing, ATSC3dv, SaasC, nSpiro, NdssC, nBondsT, ATSC3se, ATSC3c, ETA_shape_p, ATSC5d, ATS4dv, ATSC2v, PEOE_VSA2, GGI4, ATSC3Z, n7FRing, PEOE_VSA12, JGI4, n5AHRing, PEOE_VSA4, SlogP_VSA1, ATSC4c, ATSC6d, PEOE_VSA9, PEOE_VSA8, C1SP2, ATSC4se, ATSC4dv, ATSC4Z, SMR_VSA6, NssCH2, NaaN, ATSC5c, NdsN, ATSC5dv, ATSC5se, EState_VSA4, GGI2, PEOE_VSA3, NtN, NsOH, n8FRing, EState_VSA7
CFS	SIC0, Mm, VE2_Dzare, AATS1Z, AATSC0v, AMID_h, CIC0, AMW, VE2_Dzpe, SIC1, MZ, FCSP3, AETA_eta_FL, ETA_dEpsilon_C, ETA_epsilon_4, VE2_Dzi, Mv, TopoPSA(NO), n4Ring, AETA_eta_F, VE2_D, AATS1m, n3Ring, AATSC0dv, C2SP1, VE2_A, SM1_Dzi, AATSC1c, BIC1, CIC1, AATS0Z, VE2_Dt, IC1, FilterItLogS, AETA_beta, AETA_eta_L, AATS1dv, CIC2, SIC2, CIC3, AETA_eta_RL
LASSO	VR1_A, nAromAtom, nBridgehead, ATS2dv, ATS2d, ATS4d, ATS5d, ATS4Z, ATS1m, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS2se, ATS4se, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, ATS2v, AATS2i, ATSC2dv, ATSC3dv, ATSC3d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC0v, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC7v, ATSC8v, ATSC1i, ATSC2i, ATSC3i, ATSC5i, BertzCT, SpDiam_Dt, DetourIndex, StCH, SaaCH, StN, SssO, SsF, ECIndex, ETA_beta, fragCpx, TIC0, TIC1, TIC3, TIC4, PEOE_VSA2, PEOE_VSA3, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, PEOE_VSA13, SMR_VSA1, SMR_VSA2, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA5, SlogP_VSA11, EState_VSA3, EState_VSA5, EState_VSA6, EState_VSA8, VSA_EState2, VSA_EState5, VSA_EState7, MPC3, MPC5, MPC7, MPC8, TopoPSA(NO), TSRW10

A.4.4 Energia livre de Gibbs de atomização para a temperatura de 298.15 K

Metódos	Atributos selecionados
Pcorr	Sp, SlogP_VSA5, SLogP, PEOE_VSA7, VSA_EState8, SssCH2, ATS4Z, BCUTdv-1l, PEOE_VSA6, ATS3Z, BCUTi-1h, GATS2m, C2SP3, ATSC0i, FilterItLogS, AATSC0i, ETA_dEpsilon_B, ATS4dv, C1SP2, NaaN, AATS0d, AETA_beta_s, BertzCT, ATSC3v, TIC0, EState_VSA4, BCUTp-1l, SMR_VSA4, C3SP3, WPath, ZMIC1, SlogP_VSA1, PEOE_VSA3, IC5, fMF, PEOE_VSA4, ATSC6d, GATS2v, EState_VSA5, TIC1, VSA_EState2, PEOE_VSA11, AETA_beta_ns_d, GATS1pe, PEOE_VSA10, VSA_EState7, RPCG, BCUTc-1l, C4SP3, EState_VSA1, AATSC1i, NaaO, VR1_A, ECIndex, PEOE_VSA12, NsssCH, SlogP_VSA10, PEOE_VSA13, ATSC4p, ATSC6i, ATSC4d, PEOE_VSA2, SssssC, ZMIC5, RNCG, ATSC7d, GATS2pe, ATSC2d, ATSC7i, SMR_VSA7, SdssC, ATSC1Z, nBondsD, BCUTv-1h, ATSC1dv, nHRing, EState_VSA10, EState_VSA9, GGI2, n5HRing, n6aHRing, ATSC3dv, VSA_EState4, nRot, NsNH2, PEOE_VSA1, VSA_EState5, SlogP_VSA11, NaaNH, SMR_VSA9, VSA_EState3, SMR_VSA2, EState_VSA2, C1SP1, GGI3, ATS5dv, Xch-3dv, VSA_EState1, SlogP_VSA3, ATSC2dv, NdssC, ATSC8i, SlogP_VSA2, NdsN, TSRW10, ATSC5d, ATSC3pe, SlogP_VSA4, nFaRing, EState_VSA8, NaasN
Scorr	Sp, SlogP_VSA5, ATSC0i, SpAbs_Dzare, PEOE_VSA7, BCUTdv-1l, SLogP, VSA_EState8, SssCH2, AATSC0i, BCUTi-1h, ZMIC0, ATS4Z, PEOE_VSA6, AATSC1se, ETA_epsilon_5, ATS3Z, C2SP3, FilterItLogS, GATS2m, BCUTdv-1h, SssCH, ATS4dv, C1SP2, AETA_beta_s, EState_VSA4, BertzCT, C3SP3, IC5, NaaN, fMF, ZMIC5, SMR_VSA4, ATSC3v, RNCG, SlogP_VSA1, ATSC6d, PEOE_VSA3, ETA_dEpsilon_B, ZMIC1, PEOE_VSA10, PEOE_VSA11, VSA_EState2, VE2_Dt, VE2_A, VE2_Dzm, PEOE_VSA4, VSA_EState7, AETA_beta_ns_d, TIC0, EState_VSA1, GATS2v, NsssCH, ATSC6i, C4SP3, GATS1pe, VSA_EState5, EState_VSA5, PEOE_VSA12, SlogP_VSA10, EState_VSA9, ATSC4d, VR1_A, SssssC, NaaO, LogEE_Dzv, ATSC4p, nHRing, PEOE_VSA13, ETA_dPsi_B, VE3_Dzp, PEOE_VSA2, TIC1, BCUTd-1l, ATSC1Z, SMR_VSA7, ETA_eta_R, ATSC1dv, AATSC1i, GATS2pe, n5HRing, VE3_D, SMR_VSA9, nBondsD, RPCG, EState_VSA2, BCUTv-1h, NsNH2, GGI2, PEOE_VSA1, ATSC3dv, n6aHRing, SlogP_VSA11, ECIndex, Xc-6dv, NaaNH, SlogP_VSA3, SMR_VSA2, VSA_EState3, AATSC2i, SdssC, ATSC7i, ETA_shape_y, GGI3, SaasC, ATSC3pe, AMID_X, nRot, Xch-3dv, AATSC2c, ATS8i
RF	nH, nC, ATSC0i, SlogP_VSA5, TIC0, SpAbs_DzZ, TIC1, ATS0dv, ATS0Z, AATS0i, BCUTi-1h, SLogP, VE2_A, BCUTdv-1l, IC1, nN, SpAbs_A, ATSC0dv, SlogP_VSA2, VE1_A, AATS0d, IC3, PEOE_VSA7, VR1_A, nO, ATS4Z, ABC, BCUTdv-1h, ETA_dEpsilon_B, GATS1se, BertzCT, nBase, SMR_VSA7, nBondsD, SpAbs_Dzv, RNCG, ATSC1se, PEOE_VSA1, ATSC2se, ECIndex, PEOE_VSA6, VSA_EState1, ATS3Z, n5Ring, VSA_EState6, FilterItLogS, ZMIC2, BCUTz-1l, BCUTd-1l, C2SP2, RPCG, ATSC2c, n4HRing, ZMIC1, ATSC1v, ATSC1Z, VSA_EState5, NdO, SpMAD_A, BCUTc-1l, BalabanJ, ATSC6d, nBridgehead, n5FRing, BCUTd-1h, IC2, nF, ATSC2d, EState_VSA8, GATS2c, SlogP_VSA4, Xc-5d, SdssC, n6Ring, SlogP_VSA3, EState_VSA9, nAromAtom, C2SP1, GGI3, VSA_EState7, Xch-3d, SssssC, NdsCH, EState_VSA10, ATSC2dv, VSA_EState4, PEOE_VSA4, nHRing, ATSC2Z, piPC8, ATSC4d, ATSC1dv, n6aRing, ETA_shape_y, nRot, NdssC, ATSC3dv, ETA_shape_p, ATSC2i, nSpiro, SaasC, n5HRing, ATSC5d, ATSC3se, GGI4, ATS4dv, ATSC6d, NaaN, ATSC3Z, ATSC3c, ATSC2v, nBondsT, PEOE_VSA12, PEOE_VSA9, PEOE_VSA2, SlogP_VSA1, n7FRing, n5AHRing, JGI4, C1SP2, ATSC4c, NtN, ATSC4dv, NsCH3, PEOE_VSA8, ATSC4se, ATSC5c, PEOE_VSA11, ATSC4Z, NsOH, ATSC5se, NssO, GGI2, ATSC5dv, EState_VSA4, PEOE_VSA3, NsssCH, NdsN, EState_VSA7, SMR_VSA6
CFS	SIC0, Mm, VE2_Dzare, AATS1Z, SIC1, VE2_Dzpe, AATSC0dv, Mv, FCSP3, AMID_N, AATSC0v, AATS1se, VE2_DzZ, AMW, C2SP1, AMID_h, AMID_C, TopoPSA(NO), VE2_Dt, AATS1m, MZ, AATSC1c, AATSC0Z, SM1_Dzp, AATSC0c, CIC0, BIC0, AETA_alpha, ETA_dAlpha_B, ETA_psi_1, ETA_dPsi_A, ETA_dEpsilon_C, ETA_epsilon_4, SM1_Dzv, SM1_Dzi, VE2_Dzm, AATS0Z, AATS0v, FilterItLogS, AETA_eta_FL, VE2_D, AATS1dv, n4Ring, nRot, AATSC1dv, MATS1p, ATSC3d, ATSC5d, SMR_VSA6, AATS0m, VE2_Dzp, BIC1, ATSC4p, AATSC1d, ATSC3p, AATS2Z, VE2_Dzse, AETA_beta, AETA_eta_L, AETA_eta_F, ATSC4d, AETA_eta_RL, ETA_shape_p, AETA_beta_s, ETA_dEpsilon_A, VE2_Dzi, BCUTv-1h, MID_h, CIC1, IC1, CIC2, SIC2
LASSO	VR1_A, nAromAtom, nBridgehead, ATS2dv, ATS2d, ATS4d, ATS5d, ATS4Z, ATS0m, ATS1m, ATS2m, ATS3m, ATS4m, ATS5m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS2se, ATS4se, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, AATS2v, AATS2i, ATSC2dv, ATSC3dv, ATSC3d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC0v, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC7v, ATSC8v, ATSC1i, ATSC2i, ATSC3i, ATSC5i, SpAD_Dzp, BertzCT, SpDiam_Dt, DetourIndex, StCH, SaaCH, StN, SssO, SsF, ECIndex, ETA_beta, fragCpx, TIC0, TIC1, TIC3, TIC4, PEOE_VSA2, PEOE_VSA3, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, PEOE_VSA13, SMR_VSA1, SMR_VSA2, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA5, SlogP_VSA11, EState_VSA3, EState_VSA5, EState_VSA6, EState_VSA8, VSA_EState2, VSA_EState5, VSA_EState7, MPC3, MPC5, MPC7, MPC8, TopoPSA(NO), TSRW10

A.4.5 Capacidade térmica para a temperatura de 298.15 K

Metódos	Atributos selecionados
Pcorr	VMcGowan, ATS4Z, fMF, NsCH3, ETA_eta_R, VE2_A, AATS0d, AETA_beta_s, ETA_shape_p, BCUTdv-1l, TIC0, CIC5, SlogP_VSA5, TIC1, TpiPC10, nRot, SpMAD_A, nHRing, ATS3Z, BalabanJ, PEOE_VSA6, GATS1pe, ATSC0i, n5HRing, nFHRing, piPC8, ETA_epsilon_5, AMID_N, n5Ring, AATSC0i, ATSC7d, ETA_shape_y, SLogP, NaaN, SMR_VSA7, AATSC1m, Xc-4dv, PEOE_VSA7, ATSC4d, BCUTd-1l, JGI2, FilterItLogS, ATSC1d, ATSC6d, BCUTi-1h, SsOH
Scorr	VMcGowan, SpMax_Dzare, fMF, ATS4Z, BCUTdv-1l, VE2_Dt, VE2_Dzm, AETA_beta_s, CIC5, ETA_shape_p, ETA_eta_R, SpMax_Dzv, SlogP_VSA5, piPC6, AATSC0i, TIC0, ATSC0i, ZMIC0, nHRing, SpMAD_A, nRot, TIC1, GATS1pe, PEOE_VSA6, ATS3Z, nFHRing, PEOE_VSA7, ETA_shape_y, AMID_N, BalabanJ, n5HRing, n5Ring, piPC8, SMR_VSA7, ETA_epsilon_5, AATSC1m, ATSC8pe, ATSC6d, nBondsKD, NaaN, SLogP, ATSC4d, ATSC7i, JGI2, ATSC1d, FilterItLogS, nFHRing, Xc-3dv, ATSC6dv, BCUTdv-1h, SsOH
RF	SpAbs_DzZ, ATS4Z, nH, ETA_shape_p, NsCH3, BCUTdv-1l, AATS0d, ATS6d, VE2_A, TIC1, ATSC1Z, BalabanJ, ETA_shape_y, NsOH, ATS3Z, BCUTd-1h, piPC8, TIC0, GATS1se, PEOE_VSA1, ATS0dv, ATSC0i, SlogP_VSA5, SpMAD_A, ATSC1v, BCUTZ-1l, ATS7d, ATSC2Z, Xch-3d, BCUTi-1h, SpAbs_Dzv, BCUTd-1l, BertzCT, VSA_EState7, ABC, nHRing, NssCH2, BCUTdv-1h, NsNH2, VE1_A, nFRing, C2SP1, GGI2, SpAbs_A, Ssssc, FilterItLogS, ATS0Z, ATSC2d, VSA_EState5, NtCH, GGI3, VR1_A, EState_VSA9, BCUTc-1l, ETA_dEpsilon_B, ATSC2v, ATS4dv, ATSC1se, NssCH, ATSC1dv, nRot, nC, ATSC5d, VSA_EState4, ATSC4d, ATSC0dv, ECIndex, AATS0i, n5Ring, SLogP, ZMIC1, ATS8d, RNCG, IC1, EState_VSA10, RPCG, ATSC4Z, IC3, JGI4, Xc-5d, ATSC3Z, ATSC6d, GATS2c, SMR_VSA7, SlogP_VSA2, NaaN, ZMIC2, ATSC2se, GGI4, ATSC2i, n6FRing, ETA_beta_ns_d, IC2, VSA_EState1, ATSC3dv, ATSC5c, ATSC6c, ATSC5dv, n7FRing, EState_VSA8, ATSC2c, ATSC4dv, ATSC2dv, ATSC4c, VSA_EState6, ATSC3c, nN, SlogP_VSA4, ATSC5se, SlogP_VSA3, SaasC, ATSC6dv, ATSC7d, nBridgehead, PEOE_VSA7, ATSC3se, ATSC4se, EState_VSA2, SdssC, n4HRing, NdO, EState_VSA4, EState_VSA3, ATS5dv, PEOE_VSA9, PEOE_VSA8, ATSC7c, PEOE_VSA3, SlogP_VSA1, EState_VSA5, n5HRing, C1SP3, nAromAtom, n5AHRing, PEOE_VSA6, SMR_VSA4, nBondsD, SMR_VSA6, EState_VSA1, GGI5
CFS	VMcGowan, ATS0are, ETA_shape_p, SpMax_Dzare, ATS4v, nHBDOn, Vabc, AATSC1d, ATS6v, AETA_eta_R, ATS5p, Xc-4d, SpMAD_Dzi, ETA_beta_ns_d, SpMAD_Dzare, ATS5se, NtCH, ATSC0Z, ETA_eta_R, NsCH3, Xch-3dv, BalabanJ, GATS1m, SpMax_D, SsOH, SMR, JGI1, ATS5v, TIC2, StsC, ATS3m, ATS7p, ATS5are, SsNH2, SdO, CIC0, VSA_EState3, SsCH3, SpMax_Dzi, SaasC, ETA_shape_x, ATS4Z, StCH, GATS1v, n8FaRing, FilterItLogS, SdssC, ATS6are, MATS2Z, SpMAD_Dzpe, MATS1p, GATS1p, AATS0v, ATS4pe, SpMAD_D, ATSC5d, ATS7i, AETA_beta_ns_d, AATSC0v, ETA_dPsi_A, GATS1Z, WPol, SpMAD_Dzm, ATSC2Z, Xch-3d, ATSC3d, JGI3, ATS4p, mZagreb2, VE2_Dzare, AATS2v, ATSC0m, C2SP1, ATS0pe, NdsN, JGI2, ATS5Z, CIC2, TIC1, ETA_dBeta, AETA_eta, PEOE_VSA1, VSA_EState8, nBondsT, NdO, ATS0i, VSA_EState9, AATSC1p, LogEE_D, ATS5i, ATS0se, ETA_dEpsilon_B, GGI1, ATS6pe, StN, Sare
LASSO	VR1_A, ATS1dv, ATS2dv, ATS3dv, ATS3d, ATS3Z, ATS0m, ATS1m, ATS2m, ATS4m, ATS5m, ATS6m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS3se, ATS4se, ATS5se, ATS0i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, AATS1v, AATS2i, ATSC1dv, ATSC3d, ATSC4d, ATSC0m, ATSC1m, ATSC2m, ATSC3m, ATSC4m, ATSC0v, ATSC1v, ATSC2v, ATSC3v, ATSC5v, ATSC6v, ATSC8v, ATSC0i, ATSC1i, SpAbs_Dzp, BertzCT, SpAbs_Dt, SpDiam_Dt, SpAD_Dt, StCH, SaaCH, SdNH, SaaN, SsOH, ECIndex, fragCpx, TIC1, TIC2, PEOE_VSA1, PEOE_VSA2, PEOE_VSA3, PEOE_VSA5, PEOE_VSA8, PEOE_VSA10, PEOE_VSA12, SMR_VSA3, SMR_VSA5, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA3, SlogP_VSA4, SlogP_VSA5, SlogP_VSA6, EState_VSA3, EState_VSA4, EState_VSA5, EState_VSA6, EState_VSA8, VSA_EState4, VSA_EState7, MPC3, MPC7, TopoPSA(NO), Vabc, TSRW10, Zagreb2

A.4.6 Momento dipolar

Metodos	Atributos seleccionados
Pcorr	IC0, BCUTc-1h, BCUTi-1h, nN, StN, ATS4dv, GATS2pe, VSA_EState2, SLogP, PEOE_VSA7, ETA_dPsi_B, SssCH2, GATS1pe, PEOE_VSA4, GATS2Z, MATS1c, SlogP_VSA5, SlogP_VSA1, VSA_EState7, PEOE_VSA6, ZMIC1, EState_VSA5, ATSC1m, BCUTc-1l, nBondsD, VSA_EState8, PEOE_VSA2, C2SP3, C1SP2, VSA_EState5, piPC1, EState_VSA10, C3SP3, ATS5dv, ATSC3v, VSA_EState6, RPCG, MATS1are, PEOE_VSA12, ATSC6dv, C1SP1, mZagreb2, SMR_VSA4, NssCH, ATSC3dv, NdssC, ATSC5dv, ATSC6d, EState_VSA4, MATS2se, BCUTv-1h, BCUTd-1l, ATSC2dv, NssNH, EState_VSA2, ETA_dEpsilon_D, ATSC1v, PEOE_VSA3, AATSC1i, VSA_EState3, ZMIC0, nFARing, AATSC0i, ATSC7dv, NsNH2, BCUTdv-1l, ZMIC5, PEOE_VSA10, SpMAD_Dzi, SssssC, C4SP3, ATSC0i, AETA_eta_R, ATSC4v, nBridgehead, NaaN, ATSC7d, TSRW10, ATSC2d, PEOE_VSA11, ATS4m, EState_VSA1, Xch-3dv, SlogP_VSA3, FilterItLogS, RNCG, SpMAD_A, PEOE_VSA13, VSA_EState4, ATSC4d, GGI4, ATS6dv, Radius, AATSC1dv, GATS2v, EState_VSA8, n5ARing, SssO, BalabanJ, SlogP_VSA10, TIC0, n9FARing, ETA_beta_ns_d, n6aHRing, NaaNH, GGI2, SpAbs_Dzp, ATS3Z, ATSC5d, StCH, nSpiro, SdssC, ETA_dEpsilon_B, EState_VSA7, ETA_shape_p, ATSC8dv, NaasN, SMR_VSA7, JGI3, EState_VSA9, VR2_A, n8FARing, PEOE_VSA1, ATSC3pe, NaaO, TIC1, ATSC7c, SdsCH, VE2_A, VSA_EState1, n7FARing, ATSC6c, GGI5, ATSC8c, nAHRing, n6HRing, n8FaRing, ATSC4c, ATS7dv, n5HRing, ATSC3c, n6FARing, nBase, NdssN, n6ARing, AATSC2i, nRot, ATSC4pe, SlogP_VSA2, n4HRing, n3HRing, SlogP_VSA4, SssssN, n7aRing, ECIIndex, n7ARing, NdNH, piPC8, EState_VSA6, n8ARing, SaasC, n5FRing, GGI6, JGI4, EState_VSA3, SMR_VSA6, VE1_A, nAcid, n9ARing, C2SP2, NdsN, PEOE_VSA9, ATS8dv, ETA_shape_y, n5AHRing, PEOE_VSA8, n9FaHRing, ATSC5c, C3SP2, n9aRing, GGI7, SddC, NdCH2, SlogP_VSA11, VSA_EState9
Scorr	BCUTc-1h, BCUTi-1h, TopoPSA(NO), AETA_beta, BCUTd-1l, BCUTdv-1h, ATS4dv, NdO, StN, GATS2pe, SlogP, GATS1pe, AATSC0i, ETA_dPsi_B, GATS2Z, PEOE_VSA7, SssCH2, SssssCH, SlogP_VSA1, C1SP2, nBondsD, ETA_epsilon_5, VSA_EState7, ZMIC1, AATSC1se, VSA_EState5, EState_VSA5, PEOE_VSA6, ATSC1m, ATSC6dv, ZMIC5, PEOE_VSA2, PEOE_VSA12, ATSC3m, VSA_EState6, C2SP3, BCUTv-1h, VSA_EState8, RPCG, ATS5dv, NdssC, ZMIC0, EState_VSA2, ATSC0i, SlogP_VSA3, C3SP3, EState_VSA10, ATSC5dv, C1SP1, mZagreb2, ATSC3dv, PEOE_VSA10, SssssC, NssCH, ATSC6d, NssNH, ATSC7dv, PEOE_VSA3, SMR_VSA4, EState_VSA4, ATSC1v, MATS2se, ATSC2dv, AATSC1i, IC5, RNCG, VSA_EState3, EState_VSA1, PEOE_VSA11, Xc-6dv, AETA_eta_R, BCUTc-1l, NsNH2, nBridgehead, SssO, ATSC4v, BalabanJ, C4SP3, NaaN, ATSC4d, FilterItLogS, SpMAD_Dzi, SpMAD_A, ATS4m, VE2_D, ATS7d, GGI4, PEOE_VSA13, EState_VSA8, TIC0, Radius, EState_VSA9, EState_VSA7, SlogP_VSA10, SpAbs_Dzp, ATSC8dv, PEOE_VSA1, VR2_A, n9FARing, SsOH, ATSC3pe, AATSC1dv, ATS3Z, StCH, MATS2c, MPC8, n5ARing, nSpiro, GATS2v
RF	BCUTi-1h, BCUTd-1l, BCUTdv-1h, NtN, PEOE_VSA4, NdO, nN, BCUTc-1l, SMR_VSA7, RPCG, PEOE_VSA3, VSA_EState4, SlogP_VSA2, ATSC6dv, RNCG, ATSC2c, ATSC4c, ATSC2se, PEOE_VSA1, GATS2c, ATSC5dv, ATSC3c, ATSC1se, NaaN, ATSC5c, ATS4dv, ATSC3dv, ATSC7dv, ATSC6c, BCUTZ-1l, ATSC3se, ATSC2dv, VSA_EState6, ATSC4se, PEOE_VSA12, SlogP_VSA1, ATSC5se, EState_VSA10, BertzCT, VSA_EState7, ATSC2i, ATS0dv, NtCH, VSA_EState5, PEOE_VSA2, ATSC4dv, VSA_EState1, ATSC4Z, SlogP, PEOE_VSA10, nBase, PEOE_VSA9, ATSC3Z, GATS1se, PEOE_VSA8, ATSC4d, ATSC5d, PEOE_VSA11, SdssC, IC1, BCUTdv-1l, SpAbs_Dzv, ATSC2d, nC, ATS4Z, EState_VSA2, BCUTd-1h, ATS5dv, ZMIC1, SpAbs_DzZ, ATS3Z, ZMIC2, ATSC1Z, ATSC6d, ATSC1dv, ATSC2v, FilterItLogS, ATSC2Z, VE1_A, ATSC7c, ATSC1v, ETA_beta_ns_d, nO, IC2, IC3, EState_VSA3, SpMAD_A, nHRing, JGI4, BalabanJ, ATSC0dv, AATS0i, GGI3, TIC1, nBondsD, ATS6d, SlogP_VSA10, VR1_A, EState_VSA9, PEOE_VSA13, EState_VSA8, VE2_A, ETA_shape_y, SaasC, SpAbs_A, AATS0d, ETA_dEpsilon_B, EState_VSA4, SlogP_VSA3, GGI4, EState_VSA5, Xc-5d, NdsN, ETA_shape_p, ATS0Z, GGI2, ABC, ATSC7d, SMR_VSA6, PEOE_VSA7, nRot, NsNH2, ECIIndex, C1SP2, ATSC8dv, VSA_EState9, EState_VSA6, EState_VSA1, SssssC, C2SP2, ATSC0i, NaaNH, nAromAtom, SlogP_VSA4, ATS7d, piPC8, SMR_VSA4, NdssC, n3HRing, TIC0
CFS	IC0, StN, SdO, nN, BCUTc-1h, AATSC0v, NtN, BCUTdv-1h, RPCG, ATSC4pe, TopoPSA(NO), ETA_dEpsilon_C, SMR_VSA3, NdO, SMR_VSA2, ATSC0dv, fMF, GATS1m, ATSC3dv, ATSC5dv, MATS2are, RNCG, MIC1, CIC0, AATSC0dv, BCUTp-1l, GATS2m, ETA_dPsi_A, SM1_Dzv, TopoPSA, ATSC2dv, ATSC4dv, SaasN, ATSC3are, VSA_EState8, AATS1v, ATSC0pe, SM1_Dzp, ETA_dAlpha_B, AETA_alpha, AMID_h, CIC1, ETA_dEpsilon_B, FCSP3, nBondsKD, VE2_Dzi, LabuteASA, VE2_Dt, VE3_Dzi, EState_VSA9, VSA_EState2, BCUTi-1h, AETA_eta_L, ETA_dBeta, AETA_eta, VE2_Dzare, MIC2, ATSC0v, AETA_beta, PEOE_VSA1, ETA_psi_1, ETA_epsilon_4, nBondsS, AETA_eta_FL, AMID_C
LASSO	VR1_A, ATS1dv, ATS2dv, ATS3dv, ATS4dv, ATS5dv, ATS2d, ATS3d, ATS4Z, ATS0m, ATS1m, ATS2m, ATS3m, ATS5m, ATS6m, ATS7m, ATS8m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS3se, ATS4se, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS7i, ATS8i, AATS2m, ATSC0dv, ATSC2dv, ATSC4dv, ATSC2d, ATSC3d, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC0v, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC6v, ATSC7v, ATSC8v, ATSC0i, ATSC1i, ATSC3i, ATSC5i, ATSC6i, BCUTv-1l, SpAbs_Dzv, SpDiam_Dzare, BertzCT, SpAbs_Dt, SpDiam_Dt, SpAD_Dt, StCH, SaaCH, StN, SdsN, SsOH, SdO, SaaO, ECIIndex, fragCpx, TIC1, TIC2, TIC3, TIC4, MIC1, MIC2, ZMIC1, PEOE_VSA1, PEOE_VSA4, PEOE_VSA6, PEOE_VSA8, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, SMR_VSA3, SMR_VSA6, SMR_VSA7, SlogP_VSA2, SlogP_VSA3, SlogP_VSA4, SlogP_VSA11, EState_VSA1, EState_VSA3, EState_VSA6, EState_VSA7, VSA_EState7, MPC6, MPC7, MPC8, TMPC10, TopoPSA(NO), Vabc

A.4.7 Polarizabilidade isotrópica

Metódos	Atributos selecionados
Pcorr	SMR, SLogP, SlogP_VSA5, PEOE_VSA7, WPath, SsCH3, PEOE_VSA6, FilterItLogS, ATS4Z, ECIndex, ATSC0i, SssCH2, ETA_eta_R, VSA_EState7, VSA_EState5, BCUTpe-1h, BCUTdv-1l, BCUTv-1h, TIC1, AETA_beta_s, BCUTi-1h, VR1_A, EState_VSA8, EState_VSA1, PEOE_VSA1, ATSC1d, ATSC4d, AATS0d, IC5, GATS2m, ATS3Z, ATSC7d, EState_VSA10, SssssC, EState_VSA5, VSA_EState2, SlogP_VSA3, VSA_EState1, EState_VSA4, SlogP_VSA4, PEOE_VSA10, SssCH, SlogP_VSA2, ATSC6d, GATS1pe, SMR_VSA4, SdssC, EState_VSA2, AATS1p, fMF, AATSC0i, ATS4dv, ATSC4p, C2SP3, C2SP1, ATSC1dv, PEOE_VSA11, RPCG, C4SP3, C1SP2, TIC0, BCUTc-1l, C3SP3, PEOE_VSA3, ATSC7i, ATSC3v, NaaN, Radius, AATSC1i, PEOE_VSA4, ATSC8i, GGI5, VSA_EState3, ZMIC5, PEOE_VSA12, ZMIC1, nHRRing, PEOE_VSA13, GATS2pe, SlogP_VSA1, NaaO, MATS2p, n5HRRing, EState_VSA9, NssO, SlogP_VSA10, AETA_beta_ns_d, RNCG, PEOE_VSA2, GGI4, GGI6, ATSC1Z, nBondsD, ATSC6i, C3SP2, GGI1, EState_VSA7, PEOE_VSA8, SdsCH, ETA_shape_y, SlogP_VSA11, AATSC2d, NdssC, ATSC3dv, StCH, nRot, PEOE_VSA9, Xch-3dv, piPC1, NsNH2, GGI3, n5AHRing, GGI2, ATSC4se, nFHRing, ATSC3c, GGI7, TSWW10
Scorr	SMR, ATSC0i, SpAbs_Dzare, SLogP, SlogP_VSA5, PEOE_VSA7, ETA_epsilon_5, SsCH3, AATSC1se, FilterItLogS, PEOE_VSA6, BCUTm-1h, VE2_Dzare, SssCH2, VSA_EState7, ATS4Z, VSA_EState5, ZMIC0, BCUTv-1h, AETA_beta_s, VE2_Dt, VE2_A, EState_VSA8, EState_VSA1, SssCH, ECIndex, AATS0d, PEOE_VSA1, ATSC4d, EState_VSA2, ATSC1d, TIC1, SlogP_VSA3, BCUTi-1h, ATSC6d, VE3_Dzp, fMF, PEOE_VSA10, EState_VSA4, RNCG, SlogP_VSA2, NdO, GATS2m, VR1_A, IC5, SssssC, ETA_eta_RL, ETA_eta_R, ZMIC5, SlogP_VSA4, EState_VSA5, SMR_VSA4, ATS4dv, EState_VSA10, PEOE_VSA11, ETA_dPsi_B, ATSC4p, C1SP2, GATS1pe, ATSC1dv, ATS8p, nBondsKD, nHRRing, C4SP3, PEOE_VSA3, C2SP3, ETA_shape_y, VE3_D, C3SP3, ATS3Z, SdssC, LogEE_Dzv, ATSC7i, SlogP_VSA1, ATSC6i, MATS2p, EState_VSA9, VSA_EState3, ATSC3v, C2SP1, PEOE_VSA12, n5HRRing, PEOE_VSA13, VSA_EState1, AATSC0i, NaaN, AATSC2c, GGI5, Radius, nBondsD, GATS2pe, ZMIC1, PEOE_VSA4, ATSC3c, RPCG, AATSC1i, NaaO, ATSC1Z, TIC0, GGI6, ATSC7dv, BCUTc-1l, SlogP_VSA10, ATS6dv, AATSC2se, AETA_beta_ns_d, PEOE_VSA9, PEOE_VSA2, NdssC, C3SP2, ATSC8dv, GGI4, AMID_X, EState_VSA7, SlogP_VSA11, ATS7dv, VSA_EState4, PEOE_VSA8, n5AHRing, nFHRing, ATSC3dv, StsC, AATSC2d, NsNH2, NsOH, ATSC4se, Xch-3dv, nRot, ATSC3pe, n6HRRing, NssCH, AATSC2dv, Xc-6dv, n4AHRing, MPC8, ETA_shape_p, GGI7, SMR_VSA2, ATSC5p, GGI2, n6AHRing, GGI3, n5Ring, BCUTd-1l, n8FHRing, NssNH, piPC4, n8FaRing, ATS8dv, VSA_EState9, NaaNH, SssssN, BalabanJ
RF	nC, ATSC0i, nH, SpAbs_DzZ, ATS0dv, ATS6d, TIC1, BCUTZ-1l, AATS0i, SLogP, VE2_A, ATS0Z, TIC0, ATS4Z, AATS0d, VSA_EState5, BCUTi-1h, ETA_dEpsilon_B, BCUTdv-1h, ATST7d, C2SP1, ECIndex, GATS1se, SlogP_VSA5, ABC, SMR_VSA7, VR1_A, ATSC1v, SpAbs_A, nFRRing, FilterItLogS, nN, VE1_A, BCUTd-1l, nO, ATS3Z, SdssC, SpAbs_Dzv, PEOE_VSA7, NssssCH, VSA_EState4, RNCG, BertzCT, ATSC1se, VSA_EState1, IC3, ATSC1Z, VSA_EState6, BalabanJ, VSA_EState7, NdsN, nBondsD, BCUTdv-1l, ETA_shape_y, EState_VSA8, ATS8d, ATSC4d, ZMIC1, ATSC2d, BCUTc-1l, BCUTd-1h, ATSC7d, ATSC6d, RPCG, ATSC0dv, ATSC5d, SaasC, nF, ATS4dv, SlogP_VSA2, IC1, ZMIC2, NdsCH, ATS5dv, PEOE_VSA6, C2SP2, ATSC2v, SpMAD_A, PEOE_VSA4, EState_VSA9, ETA_shape_p, ATSC2dv, SMR_VSA4, NtN, ATSC6dv, ATSC2i, GGI4, ATSC1dv, PEOE_VSA1, GATS2c, JGI4, ATS6dv, ATSC2c, ATSC2Z, nHRRing, PEOE_VSA8, ATSC2se, nAromAtom, C3SP2, ATSC3dv, C2SP3, ATSC6c, ATSC4c, GGI3, ATSC5c, ATSC3Z, NssCH2, SlogP_VSA4, EState_VSA1, ATSC5dv, ATSC3c, IC2, ATSC5se, PEOE_VSA9, ATSC4Z, ATSC4dv, SlogP_VSA1, ATSC3se, SssssC, piPC8, ATSC7c, ATSC4se, GGI2, nBase, EState_VSA2, PEOE_VSA11, n5Ring, EState_VSA6, ATSC7dv, EState_VSA3, PEOE_VSA2, EState_VSA4, EState_VSA5, ETA_beta_ns_d, EState_VSA7, ATSC8dv, EState_VSA10, nBondsT, C1SP2, GGI5
CFS	SMR, ATS0v, VMcGowan, ETA_alpha, ETA_dEpsilon_C, AATS1are, ETA_psi_1, ATS6v, LabuteASA, SM1_Dzp, ETA_dEpsilon_A, Sp, nBondsKD, ATS0p, C2SP1, ATS1i, VSA_EState4, apol, SdssC, SpAbs_D, AETA_alpha, AATSC2v, AATSC1p, ATSC3v, Sv, ATSC2p, ECIndex, PEOE_VSA7, Vabc, SMR_VSA7, Xch-4d, NssNH, StsC, ATS4d, NdsN, n6aRing, CIC0, C1SP2, nC, ATS5p, FilterItLogS, SaasC, SsCH2, AETA_eta_R, Mp, ATS8v, AETA_eta, SMR_VSA3, SM1_Dzv, ETA_dPsi_A, C2SP2, CIC1, NtsC, BCUTZ-1l, SIC1, n4Ring, ATS4v, VSA_EState5, nBondsD, Si, ETA_dEpsilon_B, ATS2i, SMR_VSA2, n7aRing, PEOE_VSA8, SlogP_VSA5, AATS1se, AATSC0v, MID_h, SLogP, SlogP_VSA2, BCUTv-1h, TopoPSA(NO), VE2_Dt, VE3_A, VE2_D, VE3_Dt, ATSC7d, ATS3v, MZ, ATSC0v
LASSO	VR1_A, nAromBond, nSpiro, nBridgehead, ATS2dv, ATS3dv, ATS4dv, ATS5dv, ATS6dv, ATS1d, ATS2d, ATS3d, ATS4d, ATS6d, ATS2Z, ATS3Z, ATS0m, ATS1m, ATS2m, ATS4m, ATS6m, ATS8m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS8v, ATS3se, ATS4se, ATS5se, ATS6se, ATS0i, ATS1i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, AATS0v, AATS2v, AATS1i, AATS2i, ATSC0dv, ATSC1dv, ATSC2dv, ATSC3dv, ATSC2d, ATSC3d, ATSC5d, ATSC6d, ATSC2Z, ATSC0m, ATSC2m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC0v, ATSC1v, ATSC2v, ATSC3v, ATSC5v, ATSC6v, ATSC7v, ATSC8v, ATSC0i, ATSC1i, ATSC2i, ATSC5i, ATSC6i, AATSC1v, BCUTm-1h, BCUTv-1l, SpDiam_Dzv, SpDiam_Dzp, SpDiam_Dzi, SpAD_Dzi, BertzCT, nBondsA, C1SP3, DetourIndex, SpAbs_D, SpDiam_D, SpAD_D, NdsCH, SssCH2, StCH, SdssC, SsNH2, SssNH, StN, SdsN, SdO, SaaO, ECIndex, fragCpx, TIC0, TIC2, TIC4, TIC5, PEOE_VSA1, PEOE_VSA3, PEOE_VSA4, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, SMR_VSA5, SMR_VSA7, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA3, SlogP_VSA5, SlogP_VSA6, SlogP_VSA8, SlogP_VSA10, EState_VSA1, EState_VSA2, EState_VSA4, EState_VSA5, EState_VSA6, EState_VSA7, EState_VSA8, EState_VSA10, VSA_EState4, VSA_EState5, VSA_EState7, MPC5, MPC7, MPC8, TMPC10, piPC6, n5ARing, n5AHRing, SMR, TopoPSA(NO), TMWC10, SRW09

A.4.8 Energia HOMO

Metodos	Atributos seleccionados
Pcorr	NtN, piPC4, NaasC, ETA_epsilon_5, SaasC, C1SP1, ETA_beta_ns_d, SpMAD_Dt, SaaNH, n5aRing, SMR_VSA7, ATS4dv, ETA_dEpsilon_C, SaaCH, C2SP2, PEOE_VSA8, TIC1, SMR, PEOE_VSA4, ATS5dv, SpMAD_Dzi, ATSC3dv, SlogP_VSA11, ATSC4d, ATS1i, EState_VSA1, SMR_VSA1, VE1_A, GATS1are, n5HRRing, ATSC1d, SlogP_VSA4, SlogP_VSA10, ATSC6dv, n5Ring, SpMAD_A, nHRRing, SpMAD_D, ATSC6d, ETA_shape_y, nHBDdon, BalabanJ, MATS1v, ZMIC5, IC5, SsNH2, SRW09, SsssN, SdssC, ATSC4v, C3SP2, ATS6dv, nRot, NdO, SssssC, SpMAD_Dzv, SMR_VSA3, SlogP_VSA1, SlogP_VSA3, ATSC7dv, EState_VSA9, IC2, VSA_EState9, GATS1p, nBondsKD, EState_VSA2, piPC8, ETA_shape_p, ETA_alpha, nFHRing, ATSC1m, EState_VSA10, MATS2v, C1SP2, VSA_EState3, ATSC4dv, SaasN, VSA_EState4, NssO, RPCG, EState_VSA4, AATSC2d, BCUTd-1l, EState_VSA8, ATSC3m, ATS4m, n8FRing, ATSC1dv, GATS1c, MATS2m, BCUTdv-1h, nBase, JGI1, ZMIC1, SMR_VSA6, VSA_EState1, TIC0, BCUTp-1h, FilterItLogS, ATSC7d, PEOE_VSA10, PEOE_VSA1, SssNH, SdsCH, ETA_dEpsilon_B, AATSC2c, ATSC0i, ATS7dv, RNCG, ATSC7c, EState_VSA5, ATSC1c, VSA_EState7
Scorr	BCUTd-1l, piPC4, NtN, SpMAD_Dt, C1SP1, ETA_epsilon_5, ATS4dv, SlogP_VSA6, SpMAD_Dzi, PEOE_VSA8, ATSC1d, ETA_dEpsilon_C, ATSC3dv, SMR, EState_VSA1, ATS5dv, C2SP2, NaasC, ETA_beta_ns_d, nO, AETA_beta_s, ATS1i, ATSC4d, SpMAD_A, nHRRing, VE1_A, SaasC, EState_VSA9, SaaNH, ATSC0i, n5Ring, EState_VSA2, ATSC6d, ATSC6dv, VSA_EState4, SsssN, ATSC1m, SsssCH, AATSC1v, SdssC, TIC1, n5HRRing, SpMAD_Dzv, nRot, ZMIC4, ETA_shape_y, BCUTm-1h, GATS1are, MPC8, SlogP_VSA11, SlogP_VSA4, ATSC4v, TIC2, IC5, ATS6dv, BalabanJ, ATSC3m, SlogP_VSA3, NdO, nBase, SssssC, ZMIC2, MID_N, SlogP_VSA10, C3SP2, GATS1p, ETA_shape_p, ATS4m, EState_VSA8, EState_VSA4, NssO, ETA_dBeta, BCUTp-1h, MATS2v, ATSC1dv, PEOE_VSA10, n8FRing, ATSC7dv, VE2_A, SsNH2, IC2, EState_VSA5, VSA_EState9, Vabc, nHBDdon, RNCG, ETA_dEpsilon_B, EState_VSA10, RPCG, AATSC2m, nBondsKD, VSA_EState3, AETA_eta_R, Radius, VSA_EState7, Xp-1d, PEOE_VSA11, VSA_EState5, AATSC1se, ATS7dv, ATS2dv, GGI5, VR1_A, ATSC5d, n9FARing, GGI6, AATSC2d, SlogP_VSA1, FilterItLogS, SsCH3, JGI1, SaasN, PEOE_VSA7, SssNH, ATSC5dv, JGI3, ZMIC1, C2SP1, AATSC0i, AATSC2dv, n7FRing, GATS2are, VE2_Dzv, SLogP, n4HRRing, BCUTc-1l, SlogP_VSA8, PEOE_VSA2, SssCH2, AATSC2c, n6Ring, TIC0, GATS1c, SMR_VSA6, n5AHRing, ATSC3pe, SaaO, nBondsD, PEOE_VSA9, nBridgehead, ATS8m, n6FRing, AATSC2pe, C1SP2, n8FaHRRing, PEOE_VSA3, JGI4, ATS3Z, PEOE_VSA1, Xc-6dv, Sm, EState_VSA6, nSpiro, NdssC, ATSC7d, GGI7, GGI4, ATSC8dv, BIC5, nAcid, ATSC5p, ATSC4c, SdsCH, n6aRing, PEOE_VSA12, SlogP_VSA5, PEOE_VSA13, ATSC5c, ATSC8pe, SddsN, NsOH, AMID_X, VE1_Dzare, EState_VSA3, SlogP_VSA2, n7ARing, VE1_Dzp, NdNH, n9Ring, NssCH, EState_VSA7, n7aHRRing, C2SP3, n9aHRRing, n8ARing, n3HRRing, n5FRing, VSA_EState6, ETA_dPsi_B, ATSC3c, n9FaRing, ZMIC0, PEOE_VSA6, C4SP3, SddC, NdCH2
RF	NtN, SaasC, PEOE_VSA4, BCUTdv-1h, SlogP_VSA10, SMR_VSA7, ETA_beta_ns_d, GATS1se, BCUTi-1h, BCUTd-1l, VSA_EState6, ATS0dv, Naan, nBase, PEOE_VSA1, nO, BCUTc-1l, ATSC0i, nH, BertzCT, PEOE_VSA8, RPCG, SlogP_VSA3, TIC1, n3HRRing, ATS4dv, ATSC1v, SlogP_VSA1, GATS2c, SlogP_VSA11, Xch-3d, IC1, BCUTZ-1l, BCUTd-1h, SlogP_VSA2, nN, NaaNH, VSA_EState9, VSA_EState4, RNCG, ETA_dEpsilon_B, GGI2, ATSC0dv, nAromAtom, PEOE_VSA3, ATSC1se, SdssC, ABC, ATSC2c, ATSC1Z, C2SP2, SLogP, AATS0i, EState_VSA9, AATS0d, nFRing, PEOE_VSA13, NssN, TIC0, ATSC2se, ZMIC2, ATSC3c, ATSC2dv, ATSC2i, PEOE_VSA2, ATSC3dv, VSA_EState1, BCUTdv-1l, VSA_EState5, VSA_EState7, ATSC1dv, EState_VSA10, NsNH2, VE1_A, SMR_VSA6, ATS0Z, ATSC4d, SpMAD_A, IC3, ZMIC1, ATSC2d, ATSC3se, ATSC4c, ATS3Z, SpAbs_DzZ, SpAbs_A, n5HRRing, EState_VSA8, FilterItLogS, PEOE_VSA10, GGI3, IC2, nBridgehead, ATS5dv, ATSC5c, SlogP_VSA4, VR1_A, SpAbs_Dzv, ATSC3Z, ATSC4se, n5Ring, ATSC4dv, NdO, PEOE_VSA9, ATS4Z, PEOE_VSA7, VE2_A, PEOE_VSA11, ATSC5dv, ATSC5se, ATSC5d, ATSC4Z, BalabanJ, ATSC6c, ATSC2v, nRot, ETA_shape_y, nBondsT, EState_VSA3, ATSC2Z, nBondsD, ATSC6d, PEOE_VSA6, ATSC6dv, nHRRing, NdsCH, C1SP2, JGI4, EState_VSA2, nC, NssNH, ETA_shape_p, SlogP_VSA5, Xc-5d, n4HRRing, GGI4, SMR_VSA4, EState_VSA4, EState_VSA1, EState_VSA6
CFS	piPC4, ETA_dEpsilon_C, ATSC0c, PEOE_VSA8, NaasC, ATS1i, VE2_Dt, TIC1, SpMAD_Dt, AETA_beta_ns_d, NdssC, GATS2are, AATSC0v, ETA_dEpsilon_D, SaaNH, NssN, SsCH3, SlogP_VSA6, MATS1v, ATSC3i, AATSC2i, GATS2i, VSA_EState9, AATSC0p, SlogP_VSA1, n5aRing, RNCG, EState_VSA10, ETA_dPsi_B, PEOE_VSA1, ATSC1Z, C2SP1, BCUTse-1l, C2SP2, SsNH2, SdssC, ETA_shape_x, GATS1i, VSA_EState5, SlogP_VSA11, n6AHRing, MATS1p, SaasN, SMR_VSA6, NdsCH, NaaNH, SM1_Dzv, BCUTare-1h, NssNH, ETA_shape_y, SaaCH, CIC0, NdsN, SlogP_VSA10, JGI2, VE2_A, ATS6v, n4HRRing, n5HRRing, NtCH, NdO, PEOE_VSA7, NsOH, GATS2c, VSA_EState4, NdNH, SddsN, NddsN, fMF, BalabanJ, Xch-4d, ATSC3d, C4SP3, SaasC, ATSC5dv, ATSC4c, ETA_beta_ns_d, TIC2, ATSC3c, AATSC2c, SsssN, VE2_D, ATSC0dv, AATS1p, AATSC2d, nHBDdon, AMID_C, nFaRing, nBondsD, PEOE_VSA9, ATSC7dv
LASSO	VR1_A, ATS2dv, ATS2d, ATS0m, ATS1m, ATS2m, ATS3m, ATS4m, ATS5m, ATS6m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS2se, ATS4se, ATS0i, ATS1i, ATS2i, ATS3i, ATS5i, ATS6i, ATS7i, ATS8i, AATS2v, ATSC4d, ATSC0m, ATSC2m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC0v, ATSC1v, ATSC3v, ATSC4v, ATSC5v, ATSC6v, ATSC7v, ATSC1i, ATSC3i, ATSC4i, SpAbs_Dzi, SpDiam_Dzi, SpAD_Dzi, BertzCT, C1SP3, DetourIndex, SaaN, TIC1, TIC2, PEOE_VSA1, PEOE_VSA3, PEOE_VSA6, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, SMR_VSA3, SMR_VSA4, SMR_VSA5, SlogP_VSA2, SlogP_VSA4, SlogP_VSA5, SlogP_VSA6, SlogP_VSA10, SlogP_VSA11, EState_VSA2, EState_VSA3, EState_VSA4, EState_VSA6, EState_VSA7, VSA_EState2, MPC4, MPC5, TMPC10, TopoPSA(NO), TSRW10

A.4.9 Diferença de energias do HOMO e LUMO

Metódos	Atributos selecionados
Pcorr	nBondsKD, CIC1, nBondsD, VSA_EState2, C1SP2, SMR_VSA7, C2SP2, GATS2m, SssCH2, BertzCT, C2SP3, ETA_dEpsilon_B, PEOE_VSA2, SlogP_VSA5, NdssC, SpMAD_Dzi, NsssCH, MID_N, ATS4dv, NdsCH, ZMIC1, EState_VSA10, PEOE_VSA7, VSA_EState8, SsssCH, GATS1pe, GATS2pe, C3SP3, GATS2v, PEOE_VSA11, ATSC2d, BCUTi-1h, ATSC3m, SLogP, SssO, SMR_VSA4, NaaN, ETA_beta_ns_d, ATSC1m, SlogP_VSA1, mZagreb2, PEOE_VSA12, ATSC2dv, n6aRing, BCUTdv-1h, BCUTdv-1l, PEOE_VSA10, AATSC0i, RNCG, PEOE_VSA9, EState_VSA4, PEOE_VSA3, TIC1, NaaO, C4SP3, ATSC0i, AATSC1i, JGI3, BCUTc-1l, EState_VSA5, VSA_EState4, nFARing
Scorr	nBondsKD, nBondsD, VSA_EState2, C2SP2, C1SP2, SMR_VSA7, BCUTdv-1h, GATS2m, SssCH2, AATSC0i, ETA_dEpsilon_B, C2SP3, NdsCH, NsssCH, NdssC, SpMAD_Dzi, BCUTi-1h, PEOE_VSA2, ATS4dv, nAromAtom, C3SP3, SsssCH, GATS1pe, GATS2pe, ZMIC1, VSA_EState8, PEOE_VSA7, GATS2v, SssO, PEOE_VSA11, ETA_beta_ns_d, EState_VSA10, PEOE_VSA10, ATSC0i, BCUTdv-1l, SMR_VSA4, BCUTd-1l, ATSC3m, ATSC1m, SlogP_VSA1, ATSC2dv, SlogP_VSA3, RNCG, ETA_epsilon_5, ZMIC5, SlogP, mZagreb2, Xc-6dv, n6aRing, EState_VSA4, PEOE_VSA12, PEOE_VSA9
RF	nBondsD, SMR_VSA7, BCUTdv-1h, IC1, C2SP2, C1SP2, ETA_dEpsilon_B, nAromAtom, BCUTi-1h, nN, BertzCT, BCUTc-1l, NdO, BCUTZ-1l, ATSC2c, ATSC2se, NdsCH, nBondsT, BCUTd-1l, nH, PEOE_VSA12, PEOE_VSA2, SlogP, NdssC, SlogP_VSA3, NaaN, ATSC3se, AATS0d, nBase, PEOE_VSA11, BCUTd-1h, SlogP_VSA2, SlogP_VSA10, SdssC, RPCG, GATS1se, ATS0dv, GATS2c, ATSC3c, ABC, n4HRing, VSA_EState4, PEOE_VSA10, n6aRing, PEOE_VSA13, RNCG, PEOE_VSA1, SlogP_VSA1, PEOE_VSA3, EState_VSA8, Xch-3d, EState_VSA9, AATS0i, VSA_EState5, SaasC, GGI2, PEOE_VSA6, VSA_EState1, ATSC2dv, n3HRing, VSA_EState6, VE1_A, ATSC0dv, GGI3, SlogP_VSA4, BCUTdv-1l, n6Ring, ATSC3dv, NsssN, ATSC4c, SpAbs_DzZ, C2SP1, n5Ring, ATSC2d, NssCH2, JGI4, ATS3Z, VSA_EState7, ATSC1v, BalabanJ, ETA_beta_ns_d, VR1_A, ATSC2i, NssO, NdsN, ATSC1se, ATSC1Z, ATSC4d, ATS4Z, SpMAD_A, nBridgehead, ZMIC1, ATS4dv, SpAbs_Dzv, SpAbs_A, n5HRing, EState_VSA10, ATSC3Z, PEOE_VSA9, PEOE_VSA8, ATS5dv, ATSC5c, ATSC5dv, TIC0, ETA_shape_y, FilterItLogS, nRot, ATSC5d, SlogP_VSA5, ATSC4se, VE2_A, ATSC5se, ATSC4dv, piPC8, ATSC1dv, ATSC6dv, PEOE_VSA7, ZMIC2, ATSC6c, ATSC4Z, GGI4, ATSC6d, nO, ATSC2v, PEOE_VSA4, nHRing, ATSC2Z, TIC1, Xc-5d, IC2, SMR_VSA6, IC3, VSA_EState9, ETA_shape_p, SsssC, ATSC0i, EState_VSA2, EState_VSA3, NsssCH, EState_VSA4
CFS	FCSP3, CIC1, GATS1se, VSA_EState1, SlogP_VSA4, SMR_VSA5, VE2_Dzse, TIC0, AETA_beta, AETA_eta_L, AETA_eta_FL, ETA_eta_L, VSA_EState3, CIC2, NssO, PEOE_VSA5, ATSC2c, GATS2Z, VE2_Dt, ATSC0d, n5aHRing, SpMAD_Dzv, ATSC4d, ATSC6d, ATSC5p, ATSC3d, Xpc-5dv, ATSC2d, BCUTd-1h, NtsC, BCUTse-1l, GATS1Z, NsssCH, NaaCH, AATSC0are, SssCH2, nBondsT, GATS1c, MATS1se, PEOE_VSA9, AETA_eta, VE2_Dzare, ATSC0c, BCUTv-1h, bpol, MZ, SsssCH, TopoPSA(NO), nH, C1SP1, VE2_A, ATS1are, ETA_epsilon_4, SM1_Dzv, ETA_dPsi_A, AETA_alpha, ETA_eta_FL, nBondsS, ETA_dEpsilon_A, CIC0
LASSO	VR1_A, ATS2dv, ATS4dv, ATS5dv, ATS6dv, ATS1d, ATS2d, ATS3d, ATS5d, ATS0m, ATS1m, ATS2m, ATS3m, ATS4m, ATS6m, ATS7m, ATS0v, ATS1v, ATS2v, ATS3v, ATS4v, ATS5v, ATS6v, ATS7v, ATS8v, ATS2se, ATS0i, ATS1i, ATS2i, ATS3i, ATS4i, ATS5i, ATS6i, ATS7i, ATS8i, AATS2m, AATS2v, AATS1i, ATSC0dv, ATSC2dv, ATSC3dv, ATSC3d, ATSC1m, ATSC3m, ATSC4m, ATSC5m, ATSC6m, ATSC7m, ATSC8m, ATSC0v, ATSC1v, ATSC2v, ATSC3v, ATSC4v, ATSC5v, ATSC7v, ATSC0i, ATSC2i, ATSC3i, ATSC4i, SpDiam_Dzp, SpDiam_Dzi, SpAD_Dzi, BertzCT, SpAbs_Dt, SaaCH, SssNH, StN, SsOH, SdO, ECIndex, fragCpx, nHBAcc, TIC0, TIC1, TIC2, TIC3, MIC1, PEOE_VSA3, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, SMR_VSA1, SMR_VSA4, SMR_VSA6, SMR_VSA9, SlogP_VSA1, SlogP_VSA2, SlogP_VSA3, SlogP_VSA4, SlogP_VSA5, SlogP_VSA11, EState_VSA1, EState_VSA3, EState_VSA4, EState_VSA6, EState_VSA7, EState_VSA9, VSA_EState1, VSA_EState2, VSA_EState4, MPC5, MPC7, TopoPSA(NO), Vabc, TSRW10

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.