sid.inpe.br/mtc-m21c/2021/04.19.19.48-TDI

ANÁLISE DE EMOÇÕES EM MÍDIAS SOCIAIS UTILIZANDO APRENDIZADO DE MÁQUINA E SÉRIES TEMPORAIS CONSIDERANDO INFORMAÇÕES DE EVENTOS EXTREMOS SOCIAIS E NATURAIS

Marilyn Minicucci Ibañez

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Lamartine Nogueira Frutuoso Guimarães, e Reinaldo Roberto Rosa, aprovada em 15 de abril de 2021.

URL do documento original:
<http://urlib.net/8JMKD3MGP3W34R/44H7S82>

INPE
São José dos Campos
2021

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE
Coordenação de Ensino, Pesquisa e Extensão (COEPE)
Divisão de Biblioteca (DIBIB)
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 INTELECTUAL DO INPE - CEPPII (PORTARIA Nº 176/2018/SEI-INPE):
Presidente:
Dra. Marley Cavalcante de Lima Moscati - Coordenação-Geral de Ciências da Terra (CGCT)
Membros:
Dra. Ieda Del Arco Sanches - Conselho de Pós-Graduação (CPG)
Dr. Evandro Marconi Rocco - Coordenação-Geral de Engenharia, Tecnologia e Ciência Espaciais (CGCE)
Dr. Rafael Duarte Coelho dos Santos - Coordenação-Geral de Infraestrutura e Pesquisas Aplicadas (CGIP)
Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)
BIBLIOTECA DIGITAL:
Dr. Gerald Jean Francis Banon
Clayton Martins Pereira - Divisão de Biblioteca (DIBIB)
REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:
Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)
André Luis Dias Fernandes - Divisão de Biblioteca (DIBIB)
EDITORAÇÃO ELETRÔNICA:
Ivone Martins - Divisão de Biblioteca (DIBIB)
André Luis Dias Fernandes - Divisão de Biblioteca (DIBIB)

sid.inpe.br/mtc-m21c/2021/04.19.19.48-TDI

ANÁLISE DE EMOÇÕES EM MÍDIAS SOCIAIS UTILIZANDO APRENDIZADO DE MÁQUINA E SÉRIES TEMPORAIS CONSIDERANDO INFORMAÇÕES DE EVENTOS EXTREMOS SOCIAIS E NATURAIS

Marilyn Minicucci Ibañez

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Lamartine Nogueira Frutuoso Guimarães, e Reinaldo Roberto Rosa, aprovada em 15 de abril de 2021.

URL do documento original:
<http://urlib.net/8JMKD3MGP3W34R/44H7S82>

INPE
São José dos Campos
2021

Dados Internacionais de Catalogação na Publicação (CIP)

Ibañez, Marilyn Minicucci.

Ib1a Análise de emoções em mídias sociais utilizando aprendizado de máquina e séries temporais considerando informações de eventos extremos sociais e naturais / Marilyn Minicucci Ibañez. – São José dos Campos : INPE, 2021.

xxx + 194 p. ; (sid.inpe.br/mtc-m21c/2021/04.19.19.48-TDI)

Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2021.

Orientadores : Drs. Lamartine Nogueira Frutuoso Guimarães, e Reinaldo Roberto Rosa.

1. Eventos extremos. 2. Mídias sociais. 3. Análise de sentimento. 4. Séries Temporais. 5. Aprendizado de Máquina. I.Título.

CDU 004.85:316.454.52

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.

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS
Serviço de Pós-Graduação

DEFESA FINAL DE TESE DE MARILYN MINICUCCI IBANEZ

BANCA Nº 047/2021

No dia 15 de abril de 2021, as 09h15min, por teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Doutora em Computação Aplicada. O trabalho precisa da incorporação das correções sugeridas pela Banca Examinadora e revisão final pelo(s) orientador(es).

Título: "Análise de Emoções em Mídias Sociais utilizando Aprendizado de Máquina e Séries Temporais Considerando Informações de Eventos Extremos Sociais e Naturais"

Eu, Haroldo Fraga de Campos Velho, como Presidente da Banca Examinadora, assino esta ATA em nome de todos os membros, com o consentimento dos mesmos.

Dr. Haroldo Campos Velho - Presidente - INPE

Dr. Lamartine Nogueira Frutuoso Guimarães - Orientador - IEAv

Dr. Reinaldo Roberto Rosa - Orientador - INPE

Dr. Elcio Hideiti Shiguemori - Membro Interno – IEAv

Dr. Paulo Henrique Barchi - Membro Externo – Instituto de Pesquisa Eldorado (IPE)

Dr. Jurandy Gomes de Almeida Junior - Membro Externo – UNIFESP

Dr. Marco Antonio Leonel Caetano - Membro Externo - INSPER-SP

---

sei! assinatura eletrônica

Documento assinado eletronicamente por **Haroldo Fraga de Campos Velho**, **Pesquisador Titular**, em 19/04/2021, às 15:26 (horário oficial de Brasília), com fundamento no art. 6º do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site http://sei.mctic.gov.br/verifica.html, informando o código verificador **6601308** e o código CRC **5A911DD3**.

"A vida será mais complicada se você possuir uma curiosidade ativa, além de aumentarem as chances de você entrar em apuros, mas será mais divertida".

Edward Speyer
em *"Seis Caminhos a Partir de Newton"*, 1994

Ao meu esposo ***Fabiano****, aos meus pais* ***Aurora*** *e* ***Elar****,*
ao meu irmão ***Delano****, a minha cunhada* ***Stefani*** *e*
sobrinhos ***Nina*** *e* ***Ravi****.*

AGRADECIMENTOS

Agradeço a Deus e a Nossa Senhora por terem me abençoado com saúde e perseverança para sempre continuar, mesmos nos momentos mais difíceis.

Agradeço aos orientadores, Dr. Reinaldo Roberto Rosa e Dr. Lamartine Nogueira Frutuoso Guimarães. Em especial, agradeço ao professor Lamartine pelo conhecimento passado, pelos conselhos dados, amizade gerada e por sempre acreditar que eu conseguiria desenvolver e finalizar esta tese.

Agradeço aos professores, Dr. Rafael Duarte Coelho dos Santos, Dr. Thales Sehn Körting, Dra. Margarete Oliveira Domingues, Dr. Odim Mendes Júnior, Dr. Fernando Manuel Ramos e Dr. Adenilson Roberto Carvalho pelos conselhos, apoio e incentivo, para que eu sempre me mantivesse no objetivo de finalizar o doutorado.

Agradeço a minha família, meus pais e meu esposo, Fabiano, pela paciência, apoio, amor e carinho nos momentos de cansaço e desânimo, que me fizeram sempre permanecer concentrada para o desenvolvimento da tese.

Agradeço aos membros da banca pelo tempo dedicado à análise do texto da tese e pelo conhecimento passado.

Agradeço ao INPE/LAC/CAP pela disponibilização da insfraestrutura para o desenvolvimento do conhecimento e trabalhos.

Agradeço à Biblioteca do INPE pela agilidade, dedicação e tempo dispensado na ajuda da padronização e melhoria desta tese.

Agradeço ao Serviço de Pós-Graduação (SPG) do INPE por todo o atendimento fornecido sempre com muita agilidade e respeito aos alunos.

Agradeço ao Instituto Federal de Educação Ciência e Tecnologia de São Paulo – IFSP câmpus São José dos Campos, pela licença qualificação concedida para o desenvolvimento dos estudos.

RESUMO

Nas últimas décadas, o crescimento do uso da Internet gerou um aumento substancial na circulação de informações nas redes sociais. Devido ao grande interesse de diversas áreas da sociedade na análise de dados de redes sociais, estabeleceu-se a busca por melhores técnicas para a manipulação e compreensão desse conteúdo, permitindo que este enorme volume de informações possa ser interpretado de forma rápida e precisa. Dentro da grande variedade de informações que circulam na internet, a ocorrência de eventos extremos pode ser considerada uma área de grande interesse público devido a sua grande influência direta na sociedade. Assim, compreender a eclosão desses eventos extremos continua sendo um dos grandes desafios científicos contemporâneos, cujo progresso depende fortemente de abordagens multidisciplinares. Assim, nesta dissertação, são analisados dados coletados em mídias sociais, de grande circulação nacional e mundial, relacionados a eventos sociais e naturais extremos, a fim de identificar a emoção de ameaça definida para cada tema do evento abordado. Como estudos de caso, foram considerados dados sobre eventos sociais extremos relacionados a conflitos armados, entre os países Síria e EUA, Irã e EUA e Global (considerando os países China, Índia, Paquistão, Reino Unido, Japão, EUA, Coreia do Norte, Coreia do Sul, Taiwan e Indonésia). Em eventos naturais extremos foram selecionados dados relativos à ocorrência de secas, incêndios e desmatamentos na região da Floresta Amazônica para os anos de 2015, 2016, 2017, 2018, 2019 e 2020. A coleta dessas informações foi realizada considerando a evolução crescente de eventos, buscando entender como as ameaças ao longo do tempo podem gerar uma evolução endógena resultando em um evento extremo. O processamento dessas informações é realizado por meio da técnica de Análise de Sentimentos, para identificar o grau de ameaça de cada notícia coletada. O endereço eletrônico das notícias coletadas é armazenado em arquivo .csv juntamente com as informações sobre a data de publicação e o grau de ameaça, que formam um portfólio de ameaças para cada modelo de dados abordado. Os portfólios foram utilizados para validar o algoritmo "*P-Model*" como gerador de séries temporais endógenas para eventos extremos. O resultado desta validação é a geração de séries temporais de ameaças endógenas, que são utilizadas para prever a variação de ameaça futura dos eventos sociais e naturais extremos analisados. Para realizar a predição de séries temporais endógenas, utiliza-se a técnica de "*Deep Learning*" em uma estrutura da rede que aplica a rede neural "*Long-Short Term Memory - LSTM*". Os resultados alcançados com base no LSTM, mostraram uma acurácia entre 46% e 71% na previsão do padrão de flutuação interpretado como ameaças, quando considerados os dados coletados para os dois estudos de caso abordados.

Palavras-chave: Eventos Extremos. Mídias Sociais. Análise de Sentimento. Séries Temporais. Aprendizado de Máquina. Aprendizado Profundo. LSTM.

ANALYSIS OF EMOTIONS IN SOCIAL MEDIA USING MACHINE LEARNING AND TIME SERIES CONSIDERING INFORMATION FROM EXTREME SOCIAL AND NATURAL EVENTS

ABSTRACT

In the last decades, the growth of Internet access has generated a substantial increase in the circulation of information on social networks. Due to the great interest of several areas of society in the analysis of social network data, the search for better techniques for the manipulation and understanding of this content has been established, allowing this huge volume of information to be interpreted quickly and accurately. Within the wide variety of information circulating on the internet, the occurrence of extreme events can be considered an area of great public interest due to their great direct influence on society. Thus, understanding the outbreak of these extreme events remains one of the great contemporary scientific challenges, whose progress depends heavily on multidisciplinary approaches. Thus, in this thesis, data collected from social media, of great national and worldwide circulation, related to extreme social and natural events are analyzed in order to identify the emotion of defined threat for each event theme addressed. As case studies, data on extreme social events related to armed conflicts were considered, between the countries Syria and USA, Iran and USA and Global (considering the countries China, India, Pakistan, United Kingdom, Japan, USA, North Korea , South Korea, Taiwan and Indonesia). On extreme natural events were selected data related to the occurrence of drought, fires and deforestation in the Amazon Forest region for the years 2015, 2016, 2017, 2018, 2019 and 2020. The collection of this information was carried out considering the increasing evolution of events, searching to understand how threats along time can generate an endogenous evolution resulting in an extreme event. The processing of this information is performed using the technique of Sentiment Analysis, to identify the degree of threat of each news collected. The electronic address of the news collected is stored in a .csv file together with the information on the date of publication and the degree of threat, which form a threat portfolio for each data model addressed. The portfolios were used to validate the algorithm P-Model as a generator of endogenous time series for extreme events. The result of this validation is the generation of endogenous threat time series, which are used to predict the future threat variation of the analyzed extreme social and natural events. To perform the prediction of endogenous time series, the Deep Learning technique is used in one structure of the network that applies the neural network Long-Short Term Memory − LSTM. The results achieved based on the LSTM, showed an accuracy between 46% and 71% in the prediction of the fluctuation pattern interpreted as threats, when considering the data collected for the two case studies addressed.

Keywords: Extreme Events. Social Media. Sentiment Analysis. Time Series. Machine Learning. Deep Learning. LSTM.

LISTA DE FIGURAS

Pág.

3.1 As fases do processo de descoberta de conhecimento (KDD) aplicadas na tese. 28
3.2 Estágios da NLP. 29
3.3 Ilustração de uma célula de memória LSTM. 35
3.4 Cascata multiplicativa para conflitos armados. (a) Hierarquia de escala para conflitos armados; (b) Respectivo padrão de densidade de espectro de energia esperada mostrando a transição da faixa inercial para o evento extremo que é a resposta ao alto regime dissipativo. 41
3.5 Exemplo de duas séries temporais típicas timuladas a partir do algoritmo "*P-Model*" fixando $p = 0.25$ e variando o valor de β. (a) Um padrão $XEendo$ de $\beta = -0.39$. (b) Um padrão $XEexo$ de $\beta = -0.72$. Os respectivos PSDs são mostrados usando imagens menores incorporadas. As Figuras (c) e (d) mostram a respectiva energia cumulativa que expressa os padrões funcionais típicos para endogenia (log) e exogenia (exponencial). 44
3.6 Exemplo de assimetria negativa, neutra e positiva 48
3.7 Exemplo das distribuições leptocúrtica, platicúrtica e mesocúrtica. 49

4.1 Reflexão para a coleta das notícias das mídias sociais. 51
4.2 Metodologia desenvolvida para a tese. 52
4.3 Aplicação da análise de sentimento para o cálculo do grau de similaridade de ameaça. 58
4.4 Esquematização do processo para gerar a série endógena de ameaças utilizando o algoritmo "*P-Model*". 62
4.5 Representação das camadas da arquitetura de aprendizado profundo e rede neural utilizada na tese. 64
4.6 Variação da perda para os dados de treinamento e validação da predição do conflito entre a Síria e USA. 65
4.7 Variação do RMSE para os dados de treinamento e validação da predição do conflito entre a Síria e USA. 65

5.1 Visão geral do portfólio de ameaças para o conflito armado entre a Síria e USA no período de janeiro de 2016 e abril de 2017. 68
5.2 Visão geral do portfólio de ameaças para o conflito armado entre Irã e USA no período de janeiro de 2019 e janeiro de 2020. 69

5.3 Visão geral do portfólio de ameaças para o conflito armado global no período de julho de 2020 e dezembro de 2020. 70
5.4 Visão geral do portfólio de ameaças para a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para o ano de 2015. . . 71
5.5 Visão geral do portfólio de ameaças para a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para o ano de 2016. . . 72
5.6 Visão geral do portfólio de ameaças para a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para o ano de 2017. . . 73
5.7 Visão geral do portfólio de ameaças para a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para o ano de 2018. . . 74
5.8 Visão geral do portfólio de ameaças para a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para o ano de 2019. . . 75
5.9 Visão geral do portfólio de ameaças para a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para o ano de 2020. . . 76
5.10 Série temporal de ameaça de eventos extremos sociais referentes ao conflito armado entre a Síria e USA. 82
5.11 Análise do cálculo dos parâmetros média, variância, assimetria e curtose para a série temporal endógena do conflito armado entre a Síria e USA. . 84
5.12 Predição de uma fração da série temporal endógena de ameaças do conflito entre a Síria e USA no período de janeiro de 2016 e abril de 2017. . 85
5.13 Sobreposição da predição da série endógena com o cálculo do valor de DTW. 86
5.14 Série temporal de ameaças de eventos extremos sociais referentes ao conflito armado entre Irã e USA. 87
5.15 Variação do cálculo dos parâmetros média, variância, curtose e assimetria para a série temporal endógena do conflito armado entre Irã e USA. . . . 89
5.16 Predição de uma fração da série temporal endógena de ameaças do conflito entre Irã e USA no período de janeiro de 2019 a janeiro de 2020. . . 90
5.17 Sobreposição da predição da série endógena com o cálculo do valor de DTW para o conflito entre Irã e USA. 90
5.18 Série temporal de ameaças de eventos extremos sociais referentes ao conflito armado global. 91
5.19 Variação do cálculo dos parâmetros média, variância, curtose e assimetria para a série temporal endógena do conflito armado Global. 93
5.20 Predição de uma fração da série temporal endógena de ameaças do conflito Global no período de julho de 2020 e dezembro de 2020. 94
5.21 Sobreposição da predição da série endógena com o cálculo do valor de Similaridade para o conflito armado Global. 94

5.22 Predição do futuro da série temporal endógena de ameaças do conflito entre a Síria e USA. 95
5.23 Predição do futuro da série temporal endógena de ameaças do conflito entre Irã e USA. 96
5.24 Predição do futuro da série temporal endógena de ameaças do conflito entre Global e USA. 96
5.25 Série temporal de ameaças de eventos extremos naturais referentes às notícias sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2015. 98
5.26 Predição de uma fração da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para ano de 2015. 99
5.27 Comparação da predição do trecho da série endógena com o trecho original pelo valor de similaridade para o ano de 2015. 99
5.28 Série temporal de ameaças de eventos extremos naturais referentes às notícias sobre ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2016. 100
5.29 Predição de uma fração da série temporal endógena sobre ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2016. 101
5.30 Comparação da predição do trecho da série endógena com o trecho original pelo valor de similaridade para o ano de 2016. 102
5.31 Série temporal de ameaça de eventos extremos naturais referentes às notícias sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2017. 103
5.32 Predição de uma fração da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2017. 104
5.33 Comparação da predição do trecho da série endógena com o trecho original pelo valor de similaridade para o ano de 2017. 104
5.34 Série temporal de ameaça de eventos extremos naturais referentes às notícias sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2018. 105
5.35 Predição de uma fração da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2018. 106
5.36 Comparação da predição do trecho da série endógena com o trecho original pelo valor de similaridade para o ano de 2018. 107

5.37 Série temporal de ameaça de eventos extremos naturais referentes às notícias sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2019. 108
5.38 Predição de uma fração da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2019. 109
5.39 Comparação da predição do trecho da série endógena com o trecho original pelo valor de similaridade para o ano de 2019. 109
5.40 Focus de queimadas x desmatamentos para o ano de 2020. 110
5.41 Série temporal de ameaça de eventos extremos naturais referentes às notícias sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2020. 111
5.42 Variação do cálculo dos parâmetros média, variância, curtose e assimetria para a série temporal endógena de ocorrência de seca, queimadas e desmatamentos do ano de 2020. 113
5.43 Predição de uma fração da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2020. 113
5.44 Comparação da predição do trecho da série endógena com o trecho original pelo valor de similaridade para o ano de 2020. 114
5.45 Predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos para o ano de 2015. 115
5.46 Sobreposição dos 3276 pontos preditos com os 3276 pontos iniciais da série de 2016 e cálculo do valor de similaridade para esses elementos. . . 116
5.47 Predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos para o ano de 2016. 116
5.48 Sobreposição dos 3276 pontos preditos com os 3276 pontos iniciais da série de 2017 e cálculo do valor de similaridade para esses elementos. . . 117
5.49 Predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos no ano de 2017. 118
5.50 Sobreposição dos 3276 pontos preditos com os 3276 pontos iniciais da série de 2018 e cálculo do valor de similaridade para esses elementos. . . 118
5.51 Predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos para o ano de 2018. 119
5.52 Sobreposição dos 3276 pontos preditos com os 3276 pontos iniciais da série de 2019 e cálculo do valor de similaridade para esses pontos. 119
5.53 Predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos para o ano de 2019. 120

5.54 Sobreposição dos 3276 elementos preditos com os 3276 elementos iniciais da série de 2020 e cálculo do valor de similaridade para esses elementos. . 121
5.55 Predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamento para o ano de 2020. 121

A.1 MultiLayer Perceptron Representation. 153
A.2 Structure of Auto-encoder neural network. 154
A.3 Structure of Deep Auto-encoder neural network. 155
A.4 Methodology for analyzing the data collected from Twitter about Brazil's presidential election in 2018. 156
A.5 Methodology flow chart of the phases sequence followed for the development of the project. 161
A.6 Methodology to Application of Sentiment Analysis for Calculating the Level of Threat. 165
A.7 Time series of the threat level in the analyzed news of the conflict between Syria and the USA. 168

C.1 Análise dos parâmetros de média, variância, curtose e assimetria considerando 32 séries com 512 elementos para o conflito armado entre a Síria e USA. 187
C.2 Análise dos parâmetros de média, variância, curtose e assimetria considerando 32 séries com 512 elementos para o conflito armado entre Irã e USA. 188
C.3 Análise dos parâmetros de média, variância, curtose e assimetria considerando 32 séries com 512 elementos para o conflito armado Global. . . . 188
C.4 Análise dos parâmetros de média, variância, curtose e assimetria considerando 8 séries com 2048 elementos para o conflito armado entre a Síria e USA. 189
C.5 Análise dos parâmetros de média, variância, curtose e assimetria considerando 8 séries com 2048 elementos para o conflito armado entre Irã e USA. 189
C.6 Análise dos parâmetros de média, variância, curtose e assimetria considerando 8 séries com 2048 elementos para o conflito armado Global. . . . 190
C.7 Análise dos parâmetros de média, variância, curtose e assimetria considerando 32 séries com 512 elementos para a ocorrência de seca, queimadas e desmatamentos para o ano de 2020. 191
C.8 Análise dos parâmetros de média, variância, curtose e assimetria considerando 8 séries com 2048 elementos para a ocorrência de seca, queimadas e desmatamentos para o ano de 2020. 191

C.1 Predição de parte da série temporal endógena de ameaças do conflito entre a Síria e USA utilizando a série endógena de 1024 elementos. 193
C.2 Predição de parte da série temporal endógena de ameaças do conflito entre a Síria e USA utilizando a série endógena de 2048 elementos. 193
C.3 Predição de parte da série temporal endógena de ameaças do conflito entre a Síria e USA utilizando a série endógena de 4096 elementos. 194
C.4 Predição de parte da série temporal endógena de ameaças do conflito entre a Síria e USA utilizando a série endógena de 8192 elementos. 194

LISTA DE TABELAS

Pág.

2.1 Sumário das referências analisadas sobre análise de sentimento. 13
2.2 Sumário das referências analisadas sobre análise de sentimento. 14
2.3 Sumário das referências analisadas sobre análise de sentimento. 15
2.4 Sumário das referências analisadas sobre análise de sentimento. 16
2.5 Sumário das referências analisadas sobre predição de conflitos armados. . 22
2.6 Sumário das referências analisadas sobre predição de conflitos armados. . 23
2.7 Sumário das referências analisadas sobre predição de fenômenos naturais. 24
2.8 Sumário das referências analisadas sobre predição de fenômenos naturais. 25

3.1 "*Layers*" disponíveis na API "*Keras*" 37

4.1 Caracterização dos termos ameaça, evento e evento extremo para os eventos extremos sociais. 53
4.2 Amostra de algumas palavras sobre conflitos armados usadas na busca de notícias de forma manual e pelo "*chatbot Newsbot*". 53
4.3 Caracterização dos termos ameaça, evento e evento extremo para os eventos extremos naturais. 54
4.4 Amostra de algumas palavras sobre ocorrência de seca, queimadas e desmatamento usadas na busca de notícias de forma manual e pelo "*chatbot Newsbot*". 54
4.5 Exemplo da organização das notícias sobre eventos extremos sociais armazenadas no arquivo "*.csv*". 55
4.6 Exemplo da organização das notícias sobre eventos extremos naturais (ocorrência de seca, queimadas e desmatamentos na Amazônia) armazenadas no arquivo "*.csv*". 55
4.7 Notícias sobre o conflito entre a Síria e USA com valor GSA, que representa o grau de ameaça. 60
4.8 Notícias sobre a ocorrência de seca, queimadas e desmatamentos na Amazônia com valor GSA que representa o grau de ameaça. 60
4.9 Exemplo da série temporal de ameaças para o conflito entre a Síria e USA. 61
4.10 Exemplo da série temporal de ameaças dos dados de ocorrência de seca, queimadas e desmatamentos no ano de 2017. 61

5.1 Valores de DFA para as séries dos conflitos armados entre a Síria e USA, Irã e USA e Global. 78

5.2 Comparação dos valores de DFA e autocorrelação das séries geradas pelo algoritmo "*P-Model*" com as séries dos conflitos armados entre a Síria e USA, Irã e USA e Global para os valores de $\beta = \{-0,3; -0,4\}$. 78
5.3 Comparação dos valores de DFA e autocorrelação das séries geradas pelo algoritmo "*P-Model*" com as séries dos conflitos armados entre a Síria e USA, Irã e USA e Global para os valores de $\beta = \{-0,5; -0,8; -1,0\}$. . . 79
5.4 Comparação dos valores de DFA e autocorrelação das séries geradas pelo algoritmo "*P-Model*" com as séries dos conflitos armados entre a Síria e USA, Irã e USA e Global para $\beta = \{-1,34; -2,0\}$. 80
5.5 Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena do conflito armado entre a Síria e USA. 83
5.6 Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena do conflito armado entre Irã e USA. 87
5.7 Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena do conflito armado global. 92
5.8 Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2015. 98
5.9 Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2016. 101
5.10 Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena sobre ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2017. 103
5.11 Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2018. 106
5.12 Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2019. 108
5.13 Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2020. 111
5.14 Resultados das similaridades ou precisões alcançadas com as predições dos dados sobre os conflitos armados entre a Síria e USA, Irã e USA e Global. 123

5.15 Resultados das similaridades ou precisões alcançadas com as predições dos dados sobre a ocorrência de seca, queimadas e desmatamentos para os anos de 2015, 2016, 2017, 2018, 2019 e 2020. 123

A.1 Sample of Vocabulary with Words Classified in Sentiment Positive or Negative. 159
A.2 The Architecture of MLP Neural Network applied in the work. 160
A.3 The Architecture of Auto-encoder - MLP Neural Network applied in the work 160
A.4 The Architecture of Deep Auto-encoder - MLP Neural Network applied in the work 161
A.5 Sample of .csv file structure with Date and URL of the Collected News. . 162
A.6 Comparison of the results of the application of machine learning techniques MLP, Auto-encoder-MLP and Deep Auto-encoder-MLP of the candidate Bolsonaro data. 166
A.7 Comparison of the results of the application of machine learning techniques MLP, Auto-encoder-MLP and Deep Auto-encoder-MLP of the candidate Haddad data. 166
A.8 Showing of the accuracy results of the application of machine learning techniques MLP, Auto-encoder-MLP and Deep Auto-encoder-MLP. . . . 167
A.9 Sample of Data analysis of the result of the armed conflict between Syrian and the USA to determine the probability of threats. 168

LISTA DE ABREVIATURAS E SIGLAS

ACLED	–	*Armed Conflict Location and Event Data Project*
API	–	*Application Programming Interface*
AUC	–	*Area Under Curve*
AVI	–	*Advanced Vegetation Index*
BLR	–	*Binary Logistic Regression*
BSI	–	*Bare Soil Index*
BS4	–	*Beautiful Soup 4.0*
CCI	–	*Cognos Consumer Insight*
CFG	–	*Context Free Grammar*
CharSCNN	–	*Character to Sentence Convolutional Neural Network*
CNN	–	*Convolutional Neural Network*
CSV	–	*Comma-Separated Values*
CVG	–	*Compositional Vector Grammar*
DVD	–	*Digital Versatile Disc*
DFA	–	*Detrended Fluctuation Analysis*
FCD	–	*Forest Canopy Density*
GCRI	–	*Global Conflict Risk Index*
GSA	–	*Grau de Similaridade de Ameaça*
HTML	–	*Hypertext Markup Language*
IA	–	*Inteligência Artificial*
IBM	–	*International Business Machines*
IG	–	*Information Gain*
KDD	–	*Knowledge Discovery in Databases*
LDA	–	*Latent Dirichlet Allocation*
LIWC	–	*Linguistic Inquiry and Word Count*
LSTM	–	*Long Short-Term Memory*
MI	–	*Mutual Information*
MLP	–	*MultiLayer Perceptron*
MODIS	–	*Moderate Resolution Imaging Spectroradiometer*
MRV	–	*Monitoring, Reporting and Verification*
NLP	–	*Natural Language Processing*
NLTK	–	*Natural Language Tool Kit*
NRT	–	*Near-Real-Time*
NP	–	*Noun Phrases*
PCFG	–	*Probabilistic Context-Free Grammar*
PNN	–	*Probabilistic Neural Network*
PP	–	*Prepositional Phrases*
RBM	–	*Restricted Boltzmann Machine*
REDD	–	*Reducing Emissions from Deforestation and Forest Degradation*
RF	–	*Random Forest*

RNN	–	*Recursive Neural Network*
RNTN	–	*Recursive Neural Tensor Network*
ROC	–	*Receiver Operating Characteristics*
RTF-REPTree	–	*Rotational Forest and Reduced Error Pruning Trees*
SASA	–	*SailAil Sentiment Analyze*
SI	–	*Shadow Index*
SSTb	–	*Stanford Sentiment Tree-bank*
STS	–	*Stanford Twitter Sentiment*
SVM	–	*Support Vector Machine*
SMO	–	*Minimal Optimization*
TSS	–	*True Skill Statistics*
UCDP	–	*Upsala Conflict Data Program*
URL	–	*Uniform Resource Locator*
USA	–	*United State of America*
VADER	–	*Valence Aware Dictionary for sEntiment Reasoning*
VD	–	*Scaled Vegetation Density*
ViEWS	–	*Violew Early Warning System*
WWW	–	*World Wide Web*

SUMÁRIO

Pág.

1 INTRODUÇÃO **1**

2 REVISÃO BIBLIOGRÁFICA **5**

2.1 Análise de sentimento 5

2.2 Eventos extremos: eventos sociais e eventos naturais 17

2.2.1 Eventos sociais - conflitos armados 17

2.2.2 Eventos naturais - fenômenos naturais 19

3 FUNDAMENTAÇÃO TEÓRICA **27**

3.1 Mídia social 27

3.2 Ciência de dados 28

3.3 Processamento da linguagem natural 29

3.4 Análise de sentimento 31

3.5 Aprendizado de máquina 32

3.5.1 Aprendizado profundo 34

3.5.2 "*Long short-term memory*" - LSTM 35

3.5.2.1 API de aprendizado profundo "*Keras TensorFlow*" 36

3.6 Série temporal 40

3.6.1 Algoritmo "*P-Model*" 40

3.6.1.1 "*Detrended fluctuation analysis*" - DFA 44

3.6.2 Média aritmética 45

3.6.3 Variância 46

3.6.4 Desvio padrão 46

3.6.5 Assimetria 47

3.6.6 Curtose 48

3.6.7 "*Dynamic time warping*" - DTW 50

4 METODOLOGIA **51**

4.1 Mídias sociais 52

4.2 Ciência de dados 54

4.3 Análise de sentimento 55

4.4 Série temporal 60

4.5 Série temporal endógena 61

4.6 Arquitetura de apresendizado profundo com rede neural LSTM 63

5 ANÁLISE E RESULTADOS 67

5.1 Resultado 1: portifólio de ameaças 67

5.1.1 Eventos extremos sociais - conflitos armados 67

5.1.1.1 Conflito armado entre a Síria e USA 67

5.1.1.2 Conflito armado entre Irã e USA 68

5.1.1.3 Conflito armado global 69

5.1.2 Eventos extremos naturais - ocorrência de seca, queimadas e desmatamentos 70

5.1.2.1 Ocorrência de seca, queimadas e desmatamentos para o ano de 2015 71

5.1.2.2 Ocorrência de seca, queimadas e desmatamentos para o ano de 2016 71

5.1.2.3 Ocorrência de seca, queimadas e desmatamentos para o ano de 2017 72

5.1.2.4 Ocorrência de seca, queimadas e desmatamentos para o ano de 2018 73

5.1.2.5 Ocorrência de seca, queimadas e desmatamentos para o ano de 2019 74

5.1.2.6 Ocorrência de seca, queimadas e desmatamentos para o ano de 2020 75

5.2 Resultado 2: validação "*P-Model*" 76

5.3 Resultado 3: predição dos eventos extremos sociais e naturais 81

5.3.1 Série temporal, série temporal endógena e predição dos eventos extremos sociais 81

5.3.1.1 Primeira fase de predição 82

5.3.1.2 Segunda fase de predição 95

5.3.2 Série temporal, série temporal endógena e predição dos eventos extremos naturais 97

5.3.2.1 Primeira fase de predição 97

5.3.2.2 Segunda fase de predição 114

5.3.2.3 Resumo dos resultados das predições realizadas para os eventos extremos sociais e eventos extremos naturais 122

6 CONCLUSÕES 125

REFERÊNCIAS BIBLIOGRÁFICAS 129

ANEXO A - APLICAÇÕES DE ANÁLISE DE SENTIMENTO . . . 145

A.1 *Sentiment analysis applied to analyze society's emotion in two different context of social media data* 145

A.1.1 Introduction 145

A.1.2 Related works 147

A.1.3 Theoretical foundation 149

A.1.3.1 Natural language processing - NLP 149
A.1.3.2 Sentiment analysis 150
A.1.3.3 Machine learning 152
A.1.4 Methodology 156
A.1.4.1 Phase 1: Methodology applied to analyze the brazilian president 2018 election 156
A.1.4.2 Phase 2: Methodology applied to analyze the armed conflict social media data 161
A.1.5 Results 164
A.1.5.1 Data analysis results of the twitter for the 2018 presidential election 164
A.1.5.2 Data analysis results of the social media about armed conflict 167
A.1.6 Conclusion 168

ANEXO B - TEXTO BASE PARA ANÁLISE DA EMOÇÃO DE AMEAÇAS 171
B.0.1 Texto base para os eventos extremos sociais 171
B.0.1.1 Texto base para o conflito armado entre Síria e USA 171
B.0.1.2 Texto base para o conflito armado entre Irã e USA 174
B.0.1.3 Texto base para o conflito armado global 175
B.0.2 Texto base para os eventos extremos naturais 178

ANEXO C - ANÁLISES ESTATÍSTICAS PARA OS PARÂMETROS DE MÉDIA, VARIÂNCIA, CURTOSE E ASSIMETRIA PARA A QUANTIDADE DE 8 E 32 SÉRIES. 187
C.1 Eventos extremos sociais - conflitos armados entre Síria e USA, Irã e USA e Global 187
C.2 Eventos extremos naturais - ocorrência de seca, queimadas e desmatamentos para o ano de 2020 190

ANEXO C - PREDIÇÃO DOS EVENTOS EXTREMOS SOCIAIS PARA A QUANTIDADE DE ELEMENTOS DAS SÉRIES ENDÓGENAS DE 1024, 2048, 4096 E 8192 193
C.1 Predição do conflito armado entre Síria e USA 193

1 INTRODUÇÃO

A evolução da internet possibilitou o advento das mídias sociais como um dos principais meios de circulação de informações pessoais, políticas e de divulgação. Como consequência, a quantidade de informações geradas diariamente nesses meios de comunicação aumenta gradativamente a cada ano. Esse excesso de informações chama a atenção de estudiosos e autoridades de diferentes áreas do conhecimento que perceberam a importância de usar essas informações de forma adequada para verificar as opiniões e sentimentos da população sobre um determinado assunto. Desse modo, identifica-se a necessidade do uso de técnicas modernas, como aprendizado de máquina e análise de sentimento, para auxiliar na verificação precisa de informações específicas entre esses enormes volumes de dados.

Uma dessas áreas da sociedade que chama atenção para uma análise mais profunda das suas causas e consequências são os eventos extremos. Um evento extremo caracteriza-se por uma sequência de pequenos eventos gerados por emoções humanas ou alguma reação da natureza que podem evoluir para um evento maior chegando até a um fato catastrófico (ROSA et al., 2019) (CLAUSET, 2018). Assim, nesta tese realiza-se a análise de informações coletadas das mídias para identificar a variação da emoção de ameaça sobre dois modelos de eventos extremos, sendo categorizados na tese como eventos extremos sociais e eventos extremos naturais. Para os eventos extremos sociais analisaram-se as informações sobre conflitos armados e para os eventos extremos naturais, analisaram-se as informações sobre a ocorrência de seca, queimadas e desmatamentos.

Os conflitos armados são eventos extremos sociais que fazem parte da história do desenvolvimento da humanidade. Os conflitos armados, dentro desse contexto, são sequências de ameaças seguidas de ataques até que se chega ao seu ápice com um conflito armado ou uma guerra. Todos os problemas gerados pelos conflitos armados, chama a atenção para uma solução que auxilie na análise, previsão e possível alerta da população para uma redução dos danos que tais eventos costumam causar. Na tese, utiliza-se como casos de estudos sobre evento extremo social, três modelos de conflitos armados: conflito entre a Síria e USA, Irã e USA e um possível conflito Global que envolve os países, China, Índia, Paquistão, Reino Unido, Japão, USA, Coreia do Norte, Coreia do Sul, Taiwan e Indonésia.

A ocorrência de seca, queimadas e desmatamentos são eventos extremos naturais que da mesma forma trazem desiquilíbrio e perdas para a vida da sociedade (SANTOS et al., 2017). Para esse evento extremo considera-se a seca como uma ameça, as

queimadas como um evento maior e o desmatamento como o evento extremo. Na tese, utilizou-se como casos de estudos para o evento extremo natural dados coletados sobre a ocorrência de seca, queimadas e desmatamentos nos anos de 2015, 2016, 2017, 2018, 2019 e 2020 na região da Floresta Amazônica.

Conforme o contexto apresentado, esta tese desenvolve uma abordagem de análise e predição de variação de ameaças de eventos extremos social e natural endógenos que utiliza informações coletadas das mídias sociais de referência mundial, nacional e específicas dos temas analisados, como (REUTERS, 2019) (CNN, 2020) (GUARDIAN, 2020), (GLOBO, 2006), (BUTLER, 1999). A coleta dessas informações é realizada utilizando-se o buscador da "*web*" Google (GOOGLE, 2016) e por meio da utilização de um "*bot*" (NEWSBOT, 2019), que realizam uma busca de notícias relacionadas com os temas abordados na tese sobre ameaças sociais, políticas, secas, queimadas e desmatamento. Cada notícia coletada, é armazenada e agrupada considerando a ordem crescente da sua data de publicação.

As notícias coletadas nas mídias sociais são processadas utilizando técnicas de aprendizado de máquina que permitem identificar as emoções humanas presentes em um documento de texto. Conforme o contexto da tese, a emoção identificada nos textos das notícias analisadas é a ameaça natural, social, política e armada dentro de cada tópico de evento extremo abordado. A técnica de aprendizado de máquina utilizada para a identificação da emoção de ameaça nas notícias é a análise de sentimento, sendo aplicada na tese em conjunto com a técnica de Processamento da Linguagem Natural, que realiza a análise de textos. A análise de sentimento possibilita que se identifique quão similar um texto é de determinado contexto, utilizando um texto base com palavras referentes a um domínio (BIRD et al., 2009), neste caso, de ameaça desses eventos extremos. Assim, cada notícia coletada, é analisada e identificada uma porcentagem de ameaça existente no seu texto.

O resultado dessa coleta é um portfólio de ameaças com séries temporais com o grau de ameaça referentes a esses eventos extremos. Esse portfólio é utilizado na tese como entrada do algoritmo "*P-model*" (ROSA et al., 2019) para a geração de uma série temporal com características endógenas. Esse modelo de série temporal se caracteriza por considerar somente elementos que exercem alguma influência interna ao domínio analisado (ROSA et al., 2019) (SORNETTE, 2006). A série endógena gerada na tese foi utilizada como entrada de um sistema desenvolvido com aprendizado de máquina, mais especificamente aprendizado profundo, para a criação do modelo de predição de eventos extremos sociais e naturais.

A técnica de aprendizado profundo está sendo utilizada na tese devido as suas propriedades de extração automática de características e de correlações não lineares existentes nos dados (GOODFELLOW et al., 2016). Os conceitos de aprendizado profundo são aplicados utilizando uma estrutura de rede neural que utiliza as redes neurais recursivas "*Long Short Term Memory - LSTM*" em um sistemas desenvolvido por meio da API de aprendizado profundo "*Tesorflow Keras*" (CHOLLET, 2015). As LSTM são adequadas para classificar, processar e prever séries temporais, devido aos atrasos de duração desconhecidas de algumas séries (HOCHREITER; SCHMIDHUBER, 1997).

Os resultados obtidos na tese pretendem apresentar uma metodologia que utiliza a técnica de análise de sentimento para identificar o grau da emoção de ameaça em cada notícia coletada. Esse grau de ameaça, com as informações de endereço de publicação da notícia e a data em ordem crescente formam um portfólio de ameaças. Esses dados do portfólio são processados pelo algoritmo "*P-Model*"que em conjunto com uma arquitetura baseada em aprendizado profundo, que utiliza a rede LSTM, permitem realizar a previsão da variação de ameaça futura desses eventos extremos sociais e naturais endógenos, podendo assim identificar a possibilidade de acontecimento futuro desses fenômenos. Dessa forma, a metodologia criada pode ser aplicada em dados das mais diversas áreas, tais como, eventos políticos, naturais, sociais, etc. Nesta tese, apresentamos exemplos de aplicação da metodologia desenvolvida nas áreas política-social e natural. A metodologia desenvolvida na tese apresenta uma nova abordagem de análise de informações de mídias sociais de grande interesse e utilidade pública.

Desta forma, a principal hipótese desta tese consiste em analisar se a técnica de análise de sentimento permite construir séries temporais a partir de notícias publicadas nas mídias sociais que antecedem o acontecimento de eventos extremos. Para realizar essa análise, defende-se que:

- a análise de sentimento é adequada para a construção de séries temporais.
- o "*P-Model*" pode representar as flutuações presentes nos dados das mídias sociais.
- a arquitetura baseada em aprendizado profundo que utiliza a rede LSTM pode predizer as séries endógenas geradas.

Esta tese está organizada de forma a apresentar no Capítulo 2 uma revisão bibli-

ográfica referente ao estado da arte dos trabalhos relacionados ao tema da tese, no Capítulo 3 algumas definições das teorias utilizadas, no Capítulo 4 a metodologia desenvolvida, no Capítulo 5 os resultados alcançados na tese e no Capítulo 6 a Conclusão.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo apresenta o estudo de trabalhos relacionados à análise de sentimento e previsão de eventos extremos sociais e naturais com a intenção de mostrar o estado da arte dessas áreas. Na Seção 2.1, apresentam-se trabalhos relacionados à análise de sentimento em mídias sociais com aprendizado de máquina e técnicas de processamento de linguagem natural desde 2012 e nas Tabelas 2.1, 2.2, 2.3 e 2.4, apresentam-se as sumarizações das principais características das referências estudadas. Na Seção 2.2, apresentam-se trabalhos relacionados a eventos extremos com abordagem para conflitos armados (Seção 2.2.1) e fenômenos naturais (Seção 2.2.2) desde 2015. Nas Tabelas 2.5 e 2.6, apresentam-se os resumos dos trabalhos aplicados na área de predição de conflitos armados e nas Tabelas 2.7 e 2.8, mostram-se os resumos com as principais características das referências estudadas para predição de fenômenos naturais.

2.1 Análise de sentimento

O artigo "*Comparative Experiments for Multilingual Sentiment Analysis Using Machine Translation*" (BALAHUR; TURCHI, 2012) propõe e avalia o uso de tradução automática e métodos supervisionados para realizar a análise do sentimento em um contexto multilíngue. Os métodos de tradução utilizados foram "*Bing Translator*", "*Google Translate*" e "*Moses to translate*" para as três línguas abordadas: Alemão, Espanhol e Francês. Para a análise de sentimento também foi utilizado técnica de aprendizado de máquina "*Support Vector Machines Sequential Minimal Optimization (SVM SMO)*".

O artigo "*Parsing with Compositional Vector Grammars*" (SOCHER et al., 2013) ressalta que até então os trabalhos apresentados na área de análise de linguagem natural utilizam-se de pequenos grupos de categorias gramaticais como frases nominais, verbais e preposicionais. Ressalta também que essa abordagem não captura toda a riqueza sintática e semântica das frases linguísticas e as tentativas de aprimorá-las por meio de frases léxicas ou de divisão de categorias apenas abordam parcialmente o problema com grandes custos computacionais. O trabalho apresenta um modelo denominado analisador "*Compositional Vector Grammar (CVG)*", que aborda o problema de representar frases e categorias. Ao contrário dos trabalhos passados, aprende em conjunto como analisar e como representar frases como categorias discretas e vetores contínuos. A CVG combina as vantagens do padrão probabilístico de Gramáticas livres de contexto (PCFG) com as redes neurais recursivas (RNNs). O primeiro pode capturar a categorização discreta de frases em NP "(*noun phrases*)"

ou PP "(*prepositional phrases*)", enquanto o segundo pode capturar informações sintáticas finas e composição-semântica em frases e palavras. Essa informação pode ajudar nos casos em que a ambiguidade sintática só pode ser resolvida com informações semânticas. A CVG baseia-se na representação dos textos em vetores de palavras contínuos e discretos. Deses vetores são gerados conjuntos de árvores que possuem a classificação dos vetores de palavras. A árvore com o menor número de nós com rótulos incorretos é considerada a melhor classificação. Uma RNN original é parametrizada por uma única matriz de peso W. Em contraste, a CVG utiliza uma RNN sintaticamente desamarrada (SU-RNN) que tem um conjunto de tais pesos. O tamanho desse conjunto depende do número de combinações de categorias de irmãos no PCFG. Esse artigo teve como referência principal o artigo publicado em 1997 por (MANNING; CARPENTER, 1997), que apresentou um trabalho intitulado "*Probabilistic Parsing Using Left Corner Language Models*" que aborda o desenvolvimento de um classificador "*Probabilistics Context-Free Grammars*(PCFG)" para a aplicação em conceitos de linguagem livre de contexto no cálculo da probabilidade de determinada classe de palavra na frase. Como resultado foi apresentado um modelo (CVG) que escolhe diferentes funções de composição para um nó pai com base nas categorias sintáticas de seus filhos. A CVG obteve uma precisão de 90,44% na classificação.

O artigo "*The New Eye of Government: Citizen Sentiment Analysis in Social Media*" publicado por (ARUNACHALAM; SARKAR, 2013) apresenta uma abordagem para análise de sentimento no contexto do Governo Americano em redes sociais. Essa abordagem apresenta um modelo considerando "*hotwords*" e não "*hotwords*" para analisar os textos sobre as políticas do governo e os sentimentos da população sobre as políticas entre outros. A análise de modelagem e sentimento é realizada pelo sistema da IBM "*Cognos Consumer Insight (CCI)*". A arquitetura da CCI funciona com base no modelo de "*hotwords*". O resultado considera os seguintes domínios: Conceitos Distribuição do sentimento entre conceitos; Distribuição do sentimento por meio de "*hotwords*"; Análise de afinidade de conceito "*hotwords*" e Análise de Causa Raiz.

O "*blog*" de (HIDALGO, J. M. G., 2013) apresenta um estudo denominado "*Baseline Sentiment Analysis with WEKA*" sobre a utilizazção do software "*Weka*". O "*Weka*" é uma coleção de algoritmos de aprendizado de máquina para aplicação em tarefas de mineração de dados (THE UNIVERSITY OF WAIKATO, 2006). O estudo destacou a utilização das classes do "*Weka*":

a) "*weka.core.converters.TextDirectoryLoader*",

b) "*weka.filters.unsupervised.attribute.StringToWordVector*",

c) "*weka.core.tokenizers.NGramTokenizer*",

d) "*weka.filters.supervised.attribute.AttributeSelection*",

e) "*weka.attributeSelection.Ranker*",

f) "*weka.classifiers.bayes.NaiveBayes*",

na classificação de textos pequenos disponíveis na "*Web*". Para a realização da classificação ressalta-se a necessidade dos dados de entrada estarem rotulados como uma afirmação positiva ou uma afirmação negativa sobre o assunto analisado. O estudo compara o grupo de palavras "*n-grams*" versus palavras simples (ou "*unigrams*") para as línguas inglesa e espanhola da seguinte forma: "*unigrams*", "*3-grams*" e "*1-a-3-grams*". Como resultado, verificou-se um melhor desempenho da metodologia abordada para a língua inglesa em relação a espanhola.

O grupo do Laboratório de Pesquisa da IBM no Brasil (SANTOS; GATTI, 2014), apresentou um trabalho intitulado "*Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts*", que propõe uma rede neural convolucional profunda para explorar a partir de informações de caracteres a nível de sentença uma análise de sentimentos de textos. A abordagem foi aplicada em dois domínios diferentes: o "*Stanford Sentiment Tree-bank (SSTb)*", que contém frases de críticas de filmes e o "*Stanford Twitter Sentiment corpus (STS)*", que contém mensagens do Twitter. A rede proposta, chamada de "*Character to Sentence Convolutional Neural Network*" (CharSCNN), usa duas camadas convolucionais para extrair características relevantes de palavras e frases de qualquer tamanho e também explora facilmente a riqueza de sentenças embutidas produzidas pelo pré-treino não-supervisionado. A CharSCNN funciona da seguinte forma: dada uma sentença, calcula-se uma pontuação para cada rótulo de sentimento $\tau \epsilon T$. Para marcar uma sentença, a rede tem como entrada a sequência de palavras na sentença e a passa através de uma sequência de camadas onde são extraídas características com níveis crescentes de complexidade. A rede extrai recursos do nível do caractere até o nível da sentença. A principal novidade dessa arquitetura de rede é a inclusão de duas camadas convolucionais, que lhe permite manusear palavras e frases de qualquer tamanho. No trabalho foi utilizado a biblioteca "*Theano*" para a implementação da rede "*Backpropagation*".

As principais contribuições do artigo são:

a) a ideia de usar redes neurais convolucionais para extrair características de caracteres da oração;

b) a demonstração de que uma arquitetura de rede neural de "*Backpropagation*" pode ser tão eficaz quanto RNTN (SOCHER et al., 2013) para a análise de sentimentos de sentenças;

c) a definição de novos resultados de estado-da-arte para as bases SSTb e STS.

O resultado obtido com a CharSCNN alcançou uma precisão de 85,7% para o domínio SSTb e 84,6% para o domínio STS.

O trabalho Monitoramento de informação em mídias sociais: o e-monitor dengue (ANTUNES et al., 2014) apresentou a utilização de um sistema de monitoramento da Internet, chamado e-Monitor Dengue, para a verificação de rumores sobre suspeitas de dengue e assim identificar em que local do país deve-se realizar uma ação mais agressiva ao combate das causas da doença. Ressaltou-se que o sistema robô utilizado somente foi adaptado para pesquisas da palavra Dengue nas fontes de notícias. As fontes de notícias inicialmente utilizadas foram:

- Oficiais (relacionadas ao governo),
- Notícia (jornais e revistas de grande circulação),
- Produção científica e
- Redes sociais.

Dentre as diversas fontes analisadas, os melhores resultados foram alcançados utilizando os dados da rede social "*Twitter*".

O artigo "*Sentiment Analysis on Microblogging by Integrating Text and Image Features*" de (ZHANG et al., 2015) apresentou um novo classificador de vizinhança baseado em similaridade para análise de sentimento que combina recursos de texto com recursos de imagem. A fim de melhorar a eficiência da análise de sentimento, considerou-se dois grupos de características, incluindo recursos de texto e recursos de imagem:

a) **Característica do texto:** O modelo "*Bigram*" foi aplicado nas características de texto para um bom desempenho utilizando o vocabulário chinês.

Alguns métodos de medição de recurso, como MI (Informação Mútua), IG (Ganho de Informação) foram utilizados para selecionar recursos.

b) **Característica da imagem:** as características da imagem foram baseadas nas relações entre efeitos emocionais induzidos que vem da teoria da arte e combinações de cor. Utilizou-se no projeto as características de cor e textura para realizar a seleção de características de imagem.

O projeto considerou uma operação comum de reconhecimento de padrões baseada no algoritmo K-vizinhos mais próximos, que é um método não-paramétrico para classificação e regressão. Ele atribui objetos "valor"ou "classe"com base em K exemplos de treinamento mais próximo no espaço de características. O projeto considerou realizar a classificação do sentimento combinando textos com imagens no espaço de texto-imagem. O artigo também apresentou que diversos experimentos de comparação mostram a eficiência do método proposto.

O artigo *Uma Abordagem Multilíngue para Análise de Sentimento* (REIS et al., 2015) apresentou uma comparação entre diversas ferramentas de análise de sentimento utilizando nove linguagens diferentes: Português, Francês, Espanhol, Italiano, Turco, Russo, Árabe, Holandês, Alemão e Inglês. A base de dados estava inicialmente em Inglês e depois foi traduzida para as outras línguas utilizando a API do "*Python Goslate*" para tradução. As ferramentas de análise de sentimento utilizadas na comparação foram: "*Linguistic Inquiry and Word Count (LIWC)*"(PENNEBAKER; FRANCIS, 1996), "*SentiStrength*" (THELWALL et al., 2010), "*SentiWordNet*"(ESULI, 2002), "*SenticNet*" (LABORATORY, 2009), "*SASA - SailAil Sentiment Analyze*" (MALANDRAKIS et al., 2016), "*Happiness Index*"(SINGH et al., 2017), "*Emotion Lexicon*" (TABAK; EVRIM, 2016), "*Hashtag Sentiment Lexicon*" (KOTO; ADRIANI, 2015), "*Sentiment Lexicon*" (LIU, 2015), "*OpinionLexicon*" (BIRD et al., 2009), "*VADER - Valence Aware Dictionary for sEntiment Reasoning*"(BIRD et al., 2009). Na análise verificou-se que a melhor acurácia das ferramentas foi para a língua inglesa.

O artigo intitulado "*A Model for Sentiment Analysis Based on Ontology and Cases*" (CECI et al., 2016) apresentou um trabalho utilizando a técnica de Raciocínio Baseado em Caso com uso de ontologia para auxiliar no processo de classificação na análise de sentimento. O modelo desenvolvido propõe que os raciocínios já desenvolvidos sejam armazenados para serem utilizados em futura classificação. O projeto utilizou como base de dados as informações da "*Amazon*" sobre os produtos câmeras e filmes e utilizou as avaliações publicadas pelos usuários como votação por quantidade de estrelas. O artigo apresentou como resultado a comparação do modelo desenvolvido

com as técnicas "*Naive Bayes* (NB)" e "*Support Vector Machine* (SVM)". Para o domínio filmes o modelo teve uma acurácia de 85% enquanto que as técnicas NB e SVM tiveram $78,7\%$ e $78,6\%$ respectivamente. Já para o domínio câmera, o modelo teve uma acurácia de 91% enquanto que ambas as técnicas NB e SVM tiveram $84,5\%$.

A "*Google*" lançou uma API em linguagem JAVA (GOOGLE, 2016) sobre linguagem natural na nuvem, esse novo serviço oferece aos desenvolvedores acesso a análise de sentimento, reconhecimento de entidades e análises de sintaxe do Google. A parte de análise de sentimento da API apresenta as seguintes funcionalidades:

- verificação da estrutura e o significado do texto, oferecendo poderosos modelos de aprendizagem de máquina em uma fácil utilização
- extração de informações sobre pessoas, lugares, eventos e muito mais, mencionados em documentos de texto, artigos de notícias ou postagens de "*blog*"
- entendimento do sentimento sobre o produto em mídias sociais ou analisar a intenção de conversas com clientes em um "*call center*" ou em um aplicativo de mensagens
- analise do texto enviado em solicitação ou integração com o armazenamento de documentos no "*Google Cloud Storage*"

O trabalho intitulado "*Multimodal Sentimental Analysis of Telug Song*" (ABBURI et al., 2016), apresentou uma abordagem para detectar o sentimento de uma canção com base em suas naturezas de multimodalidade (texto e áudio). As características líricas textuais são extraídas do banco de palavras. Usando esses recursos, Doc2Vec gerou-se um único vetor para cada canção. "*Support Vector Machine (SVM)*", "*Navie Bayes (NB)*" e uma combinação desses dois classificadores foi desenvolvida para classificar o sentimento usando as características líricas textuais. As características de áudio foram usadas como um "*add-on*" aos líricos que incluem características temporais, características espectral, do tempo e do "*chroma*". Modelos de Mistura Gaussiana (GMM), SVM e uma combinação de ambos os classificadores foi desenvolvida para classificar o sentimento usando recursos de áudio.

O artigo "*A novel deep learning architecture for sentiment classification*" (GHOSH et al., 2016), propôs uma arquitetura híbrida de aprendizagem profunda com uma

Máquina Restrita de "*Boltzmann*" de duas camadas (RBM) e uma Rede Neural Probabilística (PNN) para a classificação de sentimentos. Na primeira etapa, a RBM executou a redução de dimensionalidade. Na próxima etapa, a PNN realiza a classificação de sentimentos. O trabalho testou cinco conjuntos diferentes de dados e comparou os resultados com os trabalhos atuais. O método proposto apresentou a melhor precisão de 93,3%, 92,7%, 93,1%, 94,9% e 93,2% para as abordagens Filmes, Livros, DVD, Eletrônicos e Utensílios de cozinha, respectivamente.

O livro "*Sentiment Analysis in Social Networks*" (POZZI et al., 2016) analisou as obras publicadas na área de análise de sentimento. Em cada capítulo se aborda um tema relacionado a análise de sentimento e ao final apresenta-se os artigos utilizados para elaboração do tema em questão. No capítulo 6 aborda-se o tema "*Sentiment Analysis in Social Network: A Machine Learning Perspective*", no qual é apresentado os trabalhos realizados utilizando aprendizado de máquina supervisionado e não supervisionado com classificação polarizada e não polarizada.

O artigo "*Social Media Sentiment Analysis: lexicon versus machine learning*" (DHAOUI et al., 2017) comparou a abordagem baseada em análise léxica e a abordagem de aprendizado de máquina para abordar três questões de pesquisa:

a) Essas duas técnicas de análise de sentimento existentes são apropriadas para a análise de conversas de redes sociais?

b) Em que medida os resultados das duas abordagens diferem quando usados em conversas de redes social?

c) Uma abordagem combinada melhora a precisão geral da classificação de sentimento das conversas nas redes sociais?

Para responder a essas perguntas utilizou-se dados da rede social "*Facebook*", o "*RTextTools*", que é um pacote de aprendizado de máquina em R para classificação automática de texto e a ferramenta de mineração de texto LIWC2015, que foi usada para conduzir uma análise de sentimento baseada em análise léxica da amostra de dados. Os resultados revelam que ambas as abordagens alcançaram maior precisão na classificação de sentimento positivos do que negativos.

O artigo " *Understanding #worldenvironmentday user opinions in twitter: A topic-based sentiment analysis approach*" (REYES-MENENDEZ et al., 2018) identificou os fatores sociais, econômicos, ambientais e culturais relacionados ao cuidado sustentável com o meio ambiente e a saúde pública que mais preocupam os usuários do

"*Twitter*". Com o objetivo de identificar os fatores relacionados ao cuidado sustentável com o meio ambiente e a saúde pública, baixou-se $n = 5.873$ "*tweets*" que utilizaram a "*hashtag #WorldEnvironmentDay*" no respectivo dia. Para a classificação de sentimentos utilizou-se uma rede neural "*Support VectorMachine (SVM)*" e para a classiificação dos sentimentos em positivo, neutro e negativo utilizou-se a ferramenta NVivo Pro 12. Como resultados, identificou-se os principais fatores que preocupam a população global no que diz respeito ao desenvolvimento sustentável do planeta, saúde pública e meio ambiente. A importância dos desses resultados é determinada pela relevância da análise da opinião pública nas redes sociais sobre o meio ambiente nas redes sociais.

O artigo "*Topic modeling and sentiment analysis of global climate change tweets.*" (DAHAL et al., 2019) abordou a utilização de dados de redes sociais para a verificação de mudanças climáticas em determinada localização. No trabalho foram utilizados dados da rede social "*Twitter*" com "*geo-tags*" que possibilitava identificar a localização, data e hora das mensagens. Para a realização da classificação dos dados utilizou-se as técnicas de "*Natural Language Processing*" tais como: análise de sentimento e "*Latent Dirichlet Allocation (LDA)*". Como resultado, observa-se que a análise de sentimento mostra que a discussão geral é negativa, especialmente quando os usuários estão reagindo a eventos políticos ou climáticos extremos. A modelagem de tópicos mostra que os diferentes tópicos de discussão sobre mudanças climáticas são diversos, mas alguns tópicos são mais prevalentes do que outros.

O artigo de (MUELLER; RAUH, 2020) intitulado "*The Hard Problem of Prediction for Conflict Prevention*" explorou o aprendizado de máquina supervisionado e não supervisionado para prever o início de conflitos em países antes pacíficos. Especificamente, o modelo de "*Natural Language Process - Latent Dirichlet Allocation (LDA)*" é usado para extração de recursos de 3,8 milhões de notícias de jornais sobre tópicos pré-definidos como: economia, política e conflitos. Essas informações são utilizadas juntamente com dados da base de dados de conflitos de UPSALA em um modelo de "*Random Forest*" para prever possíveis conflitos.

O artigo "*Social media sentiment analysis based on covid19*" de (NEMES; KISS, 2021) analisou o sentimento das pessoas em relação a pandemia do COVID-19 por meio da coleta de dados das redes sociais. Os dados utilizados foram coletados na rede social "*Twitter*" e classificados em sentimento positivo, fracamente positivo, fortemente positivo, neutro, negativo, fracamente negativo e fortemente negativo. As técnicas utilizadas para a classificação foram API "*Tensorflow Keras*" da aprendizao

profundo e a ferramenta de análise de sentimento "*TextBlob*". Como resultado foi apresentado uma comparação do desempenho de ambas técnicas na classificação, com a rede RNN de aprendizado profundo apresentando um melhor desempenho do que a "*TextBlob*" .

A seguir é realizada a sumarização dos trabalhos analisados na área de análise de sentimento. Nas Tabelas 2.1, 2.2, 2.3 e 2.4, apresentam-se os resumos dos trabalhos analisados com as informações de título do trabalho, dados de entrada, técnicas e ferramentas e resultados alcançados.

Tabela 2.1 - Sumário das referências analisadas sobre análise de sentimento.

Trabalho	Dados de Entrada	Técnicas Ferramentas	Resultados Alcançados
"*Comparative Experiments for Multilingual Sentiment Analysis Using Machine Translation*" (BALABANTARAY et al., 2012)	Textos nas linguas Alemão, Espanhol e Francês	SVM SMO	Tradutor multilingue para a análise de sentimento
"*Parsing with Compositional Vector Grammars*" (SOCHER et al., 2013)	Textos na língua Inglesa	PCFG e RNNs	Precisão de 90,4%. Melhora de 3,8% em relação ao método PCFG de Stanford
"*The New Eye of Government: Citizen Sentiment Analysis in Social Media*" (ARUNACHALAM; SARKAR, 2013)	*Hotwords*" e não "*hotwords*"	Sistema da IBM "*Cognos*" "*Consumer*" *Insight* (CCI)	Ferramenta com alta precisão de análise com duas línguas e pontencial para combinações de várias línguas
"*Baseline Sentiment Analysis with WEKA*" (HIDALGO, J. M. G., 2013)	Classificação Polarizada	WEKA: "*TextDirectoryLoader*", "*StringToWordVector*", "*NGramTokenizer*", "*AttributeSelection*",	Classificação de textos na língua inglesa e espanhola para os grupos *3-grams*, *unigrams* e *1-a-3-grams*

Tabela 2.2 - Sumário das referências analisadas sobre análise de sentimento.

Trabalho	Dados de Entrada	Técnicas Ferramentas	Resultados Alcançados
"*Deep Convolutional Neural Networks* for Sentiment Analysis of Short Texts (SANTOS; GATTI, 2014)	Sentenças	Character to Sentence e duas Convolutional Neural Network (CharSCNN)"	Rede com inclusão de camadas convolucionais
Monitoramento de informação em mídias sociais: o e-Monitor Dengue (ANTUNES et al., 2014)	Textos de mídias socias com a hashtag #dengue	Software e-Monitor Dengue	Sistema adaptado para buscar sentenças com a palavra dengue. Possibilitou uma maior precisão na localização de áreas com incidência da dengue
"*Sentiment Analysis on Microblogging by Integrating Text and Image Features*" (ZHANG et al., 2015)	Textos e Imagens	Similaridade características de imagens	A combinação de recursos de texto e recursos de imagem pode melhorar o desempenho médio da classificação de sentimento
"*Multimodal Sentimental Analysis of Telug Song Telug Song*" (ABBURI et al., 2016)	Vetor de palavras	SVM e Native Bayes (NB)	Precisão de entre 85% e 95% na identificação de sentimento
"*A novel deep learning architecture for sentiment classification*" (GHOSH et al., 2016)	Classificação por Polaridade	Aprendizado profundo com dois níveis: Máquina de Boltzmann Restrita e Rede Neural Probabilística	Redução da dimensionalidade da base de dados e melhor desempenho na classificação do sentimento em filmes

Tabela 2.3 - Sumário das referências analisadas sobre análise de sentimento.

Trabalho	Dados de Entrada	Técnicas Ferramentas	Resultados Alcançados
"*A Model for Sentiment Analysis Based on Ontology and Cases*" (CECI et al., 2016)	Comentários amazon	Raciocínio baseado em casos	Acurácia para câmeras: 91% filmes: 85%
"*Sentiment Analysis in Social Networks*" (POZZI et al., 2016)	Classificação: polarizada e não-polarizada	Aprendizado de máquina supervisionado e não supervisionado	O livro apresenta diversos trabalhos na área de análise de sentimento
"*Supervised Sentiment Analysis in Multilingual Environments*" (VILARES et al., 2017)	Classificação Polarizada	Modelo supervisionado	Depende do desempenho da tradução
"*Social media sentiment analysis: lexicon versus machine learning*" (DHAOUI et al., 2017)	Mensagens do *Facebook*	Aprendizado de máquina: RTextTools e Análise Léxica: LIWC2015	Melhor Precisão na classificação de sentimentos positivos do que negativos
"*Understanding #WorldEnvironmentDay User Opinions in Twitter: A Topic-Based Sentiment Analysis Approach*" (REYES-MENENDEZ et al., 2018)	*hashtags Twitter #WorldEnvironmentDay*	SVM e NVivo Pro 12	Principais fatores que preocupam a população global em relação à ações sustentáveis saúde pública e meio ambiente

Tabela 2.4 - Sumário das referências analisadas sobre análise de sentimento.

Trabalho	Dados de Entrada	Técnicas Ferramentas	Resultados Alcançados
"*Topic modeling and sentiment analysis of global climate change tweets*" (DAHAL et al., 2019)	Mensagens do *Twitter* com geo-tags	Análise de sentimento e *Latent Dirichlet Allocation* (LDA)	Análise de sentimento: reação negativa LDA: tópicos variados sobre mudanças climáticas
"*The Hard Problem of Prediction for Conflict Prevention*" (MUELLER; RAUH, 2020)	Notícias em jornais sobre economia, política, conflitos e base de dados de UPSALA	*Latent Dirichlet Allocation* (LDA) e *Random Forest*	Predição de iníco de conflitos em locais antes pacíficos
"*Social media sentiment analysis based on COVID-19*" (NEMES; KISS, 2021)	Mensagens do *Twitter* com *hashtags:* COVID e corona	*Keras* Aprendizado Profundo RNN e *TextBlob*	RNN com melhor desempenho que *TextBlob* na classificação de sentimentos positivos

2.2 Eventos extremos: eventos sociais e eventos naturais

Nesta seção são apresentados os trabalhos referentes a predição de eventos extremos sociais na Seção 2.2.1 com ênfase em conflitos armados e eventos extremos naturais na Seção 2.2.2 com ênfase em desmatamento.

2.2.1 Eventos sociais - conflitos armados

O artigo "*Using decision tree to predict armed conflicts in Sudan*" (ABBAS et al., 2015) apresentou um estudo que teve como objetivo gerar um modelo de predição de conflitos armados no Sudão usando árvores de decisão. O trabalho utilizou dados do "*Sudanese Armed Forces HG database*". A principal contribuição do trabalho foi a identificação de palavras chaves ou variáveis que possuem uma grande relação com a ameaça de um conflito armado, como: tensão, força, estação e arma.

Em "*Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data*" (MUCHLINSKI et al., 2016) apresentou-se uma comparação do desempenho de técnica de "*Random Forest*" com três versões de regressão logística (regressão logística clássica, regressão logística de eventos raros e regressão logística regularizada L1) e verificou-se que a abordagem algorítmica fornece previsões significativamente mais precisas do início da guerra civil em dados da amostra do que qualquer um dos modelos de regressão logística. O artigo ainda discute esses resultados e as maneiras pelas quais métodos estatísticos algorítmicos como "*Random Forest*" podem ser úteis para prever com mais precisão eventos raros em dados de conflito.

O estudo intitulado "*Predicting armed conflict: Time to adjust our expectations?*" (CEDERMAN; WEIDMANN, 2017) apresentou uma análise de diversos trabalhos de predição de conflitos armados utilizando técnicas simples de redes neurais e os problemas que cada um possuem principalmente relacionadas a obtenção de informações realistas das condições sociais e políticas da região em conflito. Por fim, o autor sugere a utilização de técnicas de aprendizado de máquina mais atuais para a análise e predição de conflitos armados devido as caraterísticas diversas de seus dados.

O projeto "*Views: A political violence early-warning system*" (HEGRE et al., 2018) apresentou um sistema de alerta precoce sobre violência política que busca ser maximamente transparente, disponível ao público e coberto de maneira uniforme. O projeto descreveu as inovações metodológicas necessárias para alcançar esses objetivos. O ViEWS ainda produz previsões mensais nacionais e subnacionais para 36

meses no futuro e todos os três tipos de violência organizada pela UCDP: conflito baseado no estado, conflito não estatal e violência unilateral na África. O projeto utiliza dados de UCDP e ACLED, bem como modelagem de "*logit*" (um modelo linear generalizado), simulação dinâmica e técnicas de "*Random Forest*".

Posteriormente, o projeto "*Improving armed conflict prediction using machine learning : Views+*" (HELLE et al., 2018) expandiu a funcional idade do software "*Violew Early Warning System* (*ViEWS*)", que usa de um grande número de variáveis para realizar previsões. O objetivo do projeto seria realizar a seleção de variáveis automaticamente e, assim, melhorar a eficiência, velocidade e precisão das previsões em comparação com a versão anterior da ferramenta. Esse projeto usou as técnicas da "*Random Forest*" e a API "*Python Scikit−learn*".

O artigo "*Trends and fluctuations in the severity of interstate wars*" (CLAUSET, 2018) apresentou uma análise baseada em dados das evidências gerais para tendências nos tamanhos e anos entre guerras interestaduais em todo o mundo e usa os modelos resultantes para caracterizar a plausibilidade de uma tendência em direção à paz desde o fim da Segunda Guerra Mundial. A variabilidade subjacente nesses dados é capturada usando uma abordagem de conjunto, que então especifica um processo estacionário pelo qual distinguir tendências de flutuações no tempo de início de guerra, a gravidade das guerras e a distribuição conjunta de ataques e gravidade.

O artigo "*Modeling and forecasting armed conflict: Automl with human-guided machine learning*" (D'ORAZIO et al., 2019) apresentou uma pesquisa que se baseia em três pontos para ciências sociais computacionais. Primeiro, modelos preditivos de conflito se beneficiariam de aplicativos mínimos do "*AutoML*" . Em segundo lugar, o aprendizado de máquina guiado por humanos oferece a opção atraente de restringir os sistemas "*AutoML*" para abordar os tipos de perguntas que os pesquisadores de conflitos avaliam com modelos preditivos. Finalmente, as implementações atuais existentes do "*AutoML*" produzem soluções divergentes e podem ser aproveitadas produtivamente em paralelo.

O capítulo "*Modeling social and geopolitical disasters as extreme events: a case study considering the complex dynamics of international armed; conflicts*" (ROSA et al., 2019), do livro "*Towards Mathematics, Computers and Environment: A Disasters Perspective*" apresentou como a dinâmica turbulenta dos conflitos armados internacionais é tratada no âmbito dos sistemas multiagentes complexos, considerando explicitamente as propriedades da cascata multiplicativa não homogênea, onde a endogenia e a exogenia são pontos-chave no modelo matemático do fenômeno. Como

resultado principal, o estudo apresentou um protótipo de autômato celular que permite caracterizar regimes de conflitos armados extremos, como os ataques terroristas de 11 de setembro e as grandes guerras mundiais.

O artigo "*The global conflict risk index: A quantitative tool for policy support on conflict prevention*" (HALKIA et al., 2020) apresentou, validou e discutiu o Índice de Risco de Conflito Global (GCRI), o ponto de partida quantitativo do Sistema de Alerta Precoce de Conflitos da União Europeia. Usando regressão logística, o GCRI calcula a probabilidade de risco de conflito nacional e subnacional. Apesar de sua metodologia padrão e simples, o modelo prevê melhor do que seis outros sistemas de alerta antecipado de conflito quantitativo publicados para dez entre doze métricas de desempenho relatadas. Desse modo, o artigo visa contribuir para uma fertilização cruzada dos esforços acadêmicos e governamentais na modelagem quantitativa de risco de conflito.

2.2.2 Eventos naturais - fenômenos naturais

O artigo "*Rainfall prediction for manaus, amazonas with artificial neural networks*" (LIMA; GUEDES, 2015) abordou o problema de previsão de chuvas em Manaus usando redes neurais artificiais multicamadas. Os dados de entrada foram obtidos de uma estação meteorológica automática durante os anos de 1970 a 2015. O fator de desempenho considerado foi o erro quadrático médio normalizado. De acordo com os resultados observados, uma rede neural "*feedforward*" com 2 camadas ocultas com 10 neurônios cada foi a que melhor desempenho teve na resolução do problema. No trabalho também foi observado que a utilização de redes neurais recorrentes não teve influência no ganho de desempenho do problema abordado.

Em "*Predicting amazon fires for policy making*" (MORELLO et al., 2016) apresentou contribuições com as políticas públicas de combate aos incêndios na Amazônia. O trabalho se concentrou na identificação em escala municipal das principais variáveis de previsão de ocorrências de incêndios. Como resultado, um painel de dados inédito na literatura foi construído a partir de imagens de satélite e dados socioeconômicos, abrangendo os anos de 2008, 2010 e 2012. O trabalho concluiu que dos 41 potenciais previsores de incêndios avaliados, apenas 9 mostraram-se significativos a um nível de incerteza tolerável, compreendendo áreas desmatadas, áreas de pastagens e de floresta, terras indígenas, temperatura e textura do solo.

O artigo "*Integrating remotely sensed fires for predicting deforestation for redd*" (ARMENTERAS et al., 2017) apresentou um trabalho no qual aborda que o programa

das Nações Unidas para a Redução de Emissões por Desmatamento e Degradação Florestal (REDD +) deixa uma lacuna na tomada de decisão sobre as intervenções de REDD + ao falhar em incluir sistematicamente os incêndios em seus dados. O trabalho abordou essa lacuna crítica de conhecimento de duas maneiras. Primeiro, revisou os projetos e programas de REDD + para avaliar a inclusão de incêndios em sistemas de monitoramento, relatório e verificação (MRV). Em segundo lugar, modelou a relação entre o fogo e a floresta para um local piloto na Colômbia usando dados de monitoramento de fogo em tempo quase real (NRT) derivados do Espectronatrômetro de Imagem de Resolução Moderada (MODIS). As previsões de desmatamento baseadas no modelo tiveram um desempenho melhor do que o sistema oficial de alerta precoce REDD. A AUC do modelo para 2013 e 2014 foi de 0,81, em comparação com 0,52 para o sistema de alerta precoce em 2013 e 0,68 em 2014. Isso demonstrou que o monitoramento de incêndios NRT é uma ferramenta poderosa para prever locais de desmatamento florestal.

A publicação "*Deforestation prediction using neural networks and satellite imagery in a spatial information system*" (AHMADI, 2018) investigou a distribuição espacial do desmatamento usando redes neurais artificiais e imagens de satélite. No artigo apresentou-se a modelagem das mudanças na cobertura da terra (florestas) para prever o desmatamento usando uma rede neural artificial "*Multilayer Perceptron*" devido ao seu potencial significativo para o desenvolvimento de modelos complexos não lineares. O procedimento envolveu o registro de imagens e correção de erros, classificação de imagens, preparação de mapas de desmatamento, determinação de camadas e desenho de uma rede neural multicamadas para prever o desmatamento. As imagens de satélite para esse estudo são de uma região de Hong Kong capturadas de 2012 a 2016. Os resultados do estudo demonstraram que a abordagem de redes neurais para prever o desmatamento pode ser utilizada e seus resultados mostraram as áreas que foram destruídas durante o período da pesquisa. Ressaltou-se que devido a fatores climáticos, econômicos e políticos a predição de áreas desmatada é de difícil precisão.

O artigo "*Deforestation prediction using time series and lstm*" (GAO, 2019) utilizou as técnicas séries temporais e rede neural LSTM, para prever o desmatamento futuro da floresta amazônica. Os cálculos revelaram uma necessidade urgente de tomar medidas para evitar uma maior deterioração. Discutiu-se no trabalho que a ajuda de técnicas avançadas de IA reduzirá o desmatamento de forma eficiente, melhorando o meio ambiente em um futuro previsível.

O artigo "*Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, reptree: A case study at the gumani river basin, india*" (SAHA et al., 2020) utilizou o acoplamento de regressão logística binária (BLR), "*Random Forest (RF)*", conjunto de floresta rotacional e árvores de poda com erro reduzido (RTF-REPTree) com FCD para descobrir a probabilidade de desmatamento. No trabalho utilizou-se o Índice de vegetação avançado (AVI), índice de solo nu (BSI), índice de sombra (SI) e densidade de vegetação em escala (VD) derivados de imagens Landsat como os principais parâmetros de entrada para identificar o FCD. Esses modelos de probabilidade de desmatamento foram validados por meio da área sob a curva (AUC), características de operação do receptor (ROC), eficiência, estatísticas de habilidade verdadeira (TSS) e coeficiente Kappa. O resultado da validação mostrou que todos os modelos como BLR (AUC = 0,874), RF (AUC = 0,886) e RTF-REPTree (AUC = 0,919) teve boa capacidade de avaliar a probabilidade de desmatamento, mas entre eles, RTF-REPTree teve a maior precisão nível.

O estudo das referências bibliográficas demonstrou que a área de análise de sentimento ainda possui bastante potencial inovador para ser aplicada utilizando dados de mídias sociais para análise de eventos extremos. Também se observou na análise que as áreas de predição de conflitos armados e fenômenos naturais ainda não foram totalmente exploradas na análise de mídias sociais. Esses fatos podem ser observados nos trabalhos de Antunes et al. (2014), Dahal et al. (2019), Mueller e Rauh (2020) e Nemes e Kiss (2021). O artigo de Mueller e Rauh (2020) apesar de utilizar dados de mídias sociais e a "*Natural Language Proccessing*" com a técnica de "*Latent Dirichlet Allocation*" para a identificação dos tópicos presentes em determinadas notícias, realizou uma predição de conflitos com características bem específicas relacionadas a países que eram pacíficos e por algum motivo econômico/político geraram algum conflito. Já os artigos Antunes et al. (2014), Dahal et al. (2019) e Nemes e Kiss (2021) apresentaram uma análise utilizando mensagens de uma única rede social, a "*Twitter*".

Essa análise das referências bibliográficas também possibilitou a identificação das limitações e problemas encontrados por cada trabalho na área de predição de eventos extremos e que estão sendo considerados para o desenvolvimento da tese. O trabalho de Abbas et al. (2015) aborda o assunto utilizando árvores de decisão e não explora as novas tecnologias de aprendizado de máquina. O trabalho de Cederman e Weidmann (2017) discute a falta de dados que reflitam as reais condições sociais e políticas dos locais em conflitos e também aborda a falta de utilização de técnicas atuais de

aprendizado de máquina na análise de eventos extremos. Na área de eventos extremos naturais, verifica-se bastante a utilização de imagens nas predições (ARMENTERAS et al., 2017), (AHMADI, 2018), (SAHA et al., 2020). No entanto, não se identificou nos trabalhos analisados a utilização de dados de mídias sociais.

A seguir é realizada a sumarização dos Trabalhos Analisados na área de Eventos Extremos Sociais e Naturais com as informações de título do trabalho, dados de entrada, técnicas e ferramentas e resultados alcançados. Nas Tabelas 2.5, 2.6 e 2.6, apresentam-se os resumos das referências sobre eventos extremos sociais com o tema predição de conflitos armados e nas Tabelas 2.7 e 2.8, apresentam-se os resumos das referências sobre eventos extremos naturais com o tema seca, queimadas e desmatamento.

Tabela 2.5 - Sumário das referências analisadas sobre predição de conflitos armados.

Trabalho	Dados Utilizados	Técnica e Ferramentas	Resultado
"*Using Decision Tree to Predict Armed Conflicts in Sudan*" (ABBAS et al., 2015)	"*Sudanese armed forces HQ*"	Árvores de Decisão	Variáveis de ameaças: tensão força, estação e arma
"*Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data*" (MUCHLINSKI et al., 2016)	"*Civil Onset Data*"	Random Forest Regressão: Logistic	Random Forest teve um melhor resultado na predição de conflitos do que a Regressão Linear
"*Predicting armed conflict:Time to adjust our expectations?*" (CEDERMAN; WEIDMANN, 2017)	Artigos sobre conflitos armados	Análise	Predições de possíveis cenários
"*ViEWS: A political violence early-warning system system* " (HEGRE et al., 2018)	Dados da UCDP e ACLED	Modelo Linear, Simulação Dinâmica Random Forest	Sistema de alerta precoce sobre violência política

Tabela 2.6 - Sumário das referências analisadas sobre predição de conflitos armados.

Trabalho	Dados Utilizados	Técnica e Ferramentas	Resultado
"*Improving Armed conflict prediction using machine learning VIEWS+*" (HELLE et al., 2018)	Dados da UCDP e ACLED	"*Random Forest Scikit Learning*"	Seleção Automática de variáveis do VIEWS
"*Trends and fluctuations in the severity of interstate wars*" (CLAUSET, 2018)	Trabalhos relacionados a conflitos armados	Análise estatística	Tendências de tempo e tamanho das guerras
"*Machine Learning Artificial Intelligence and the Use of Force by States*" (DEEKS et al., 2019)	Artigos sobre conflitos armados	Análise de textos	Identificar diferentes pontos de tomada de decisão que surgem no uso da força pelos estados
"*Modeling and Forecasting Armed Conflict: AutoML with Human-Guided Machine Learning*" (D'ORAZIO et al., 2019)	Descrição do problema a ser analisado	"*AutoML* e "*Human-GuidedML* de	Seleção da melhor aplicação de aprendizado máquina para o problema descrito
"*Modeling social and geopolitical disasters as extreme events: a case study considering the of international armed*" (ROSA et al., 2019)	Base de dados de UPSALA UCDP	Cascata Homogênea Multiplicativa séries endógenas e exógenas	Algoritmos P-Model e automato celular para caracterizar conflitos

Tabela 2.7 - Sumário das referências analisadas sobre predição de fenômenos naturais.

Trabalho	Dados Utilizados	Técnica e Ferramentas	Resultado
"*The Global Conflict Risk Index: A quantitative tool for policy support on conflict prevention*" (HALKIA et al., 2020)	Base de dados de UPSALA variáveis sobre modelos de regimes políticos	"*ten-fold cross-validation*"	Validação do GCRI para eventos de conflitos e paz com 86% de precisão para ambos
"*Rainfall Prediction for Manaus, Amazonas with Artificial Neural Networks*" (LIMA; GUEDES, 2015)	Dados do Instituto Nacional de Meteorologia INMET	Rede Neural "*feedforward*"	Melhores resultados na previsão com a rede de 2 camadas e 10 neurônios
"*Predicting Amazon fires for policy making*" (MORELLO et al., 2016)	Imagens de Satélites e dados sócio econômicos	Estatística qui-quadrado e pseudo-R$\hat{2}$	Painel com variáveis previsoras
"*Integrating remotely sensed fires for predicting deforestation for REDD*" (ARMENTERAS et al., 2017)	Imagens do MODIS	Revisão dos dados do REDD+ e análise estatística das imagens do MODIS	Predição da probabilidade de desmatamento de 81%
"*Deforestation Prediction Using Neural Networks and Satellite Imagery in a Spatial Information System*" (AHMADI, 2018)	Imagens de Satélites	Rede Neural MLP	Previsão da área desmatada

Tabela 2.8 - Sumário das referências analisadas sobre predição de fenômenos naturais.

Trabalho	Dados Utilizados	Técnica e Ferramentas	Resultado
"*Deforestation Prediction Using Time Series and LSTM*" (GAO, 2019)	dados da URL rainforests. mongabay. com/amazon	Séries temporais e Redes neurais LSTM	Apresenta que a LSTM teve um melhor desempenho na predição do que as séries temporais
"*Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, REPTree: A case study at the Gumani River Basin, India India*" (SAHA et al., 2020)	Índices derivados de imagens Landsat	Acoplamento de regressão logística binária (BLR), floresta aleatória (RF), conjunto de floresta rotacional e árvores de poda com erro reduzido (RTF-REPTree)	Modelos com precisão: BLR (AUC=0.874), RF (AUC=0.886) e RTF-REPTree (AUC=0.919)
"*Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, REPTree: A case study at the Gumani River Basin, India India*" (SAHA et al., 2020)	Índices derivados de imagens Landsat	Acoplamento de regressão logística binária (BLR), floresta aleatória (RF), conjunto de floresta rotacional e árvores de poda com erro reduzido (RTF-REPTree)	Modelos com precisão: BLR (AUC=0.874), RF (AUC=0.886) e RTF-REPTree (AUC=0.919)

3 FUNDAMENTAÇÃO TEÓRICA

O advento da internet proporcionou a circulação de uma imensa quantidade de informações públicas, por meio das mídias sociais, relacionadas aos mais diversos assuntos (CASTELLS, 2003), tais como economia, política, saúde, segurança, fenômenos naturais, etc. Esse excesso de informação sobre assuntos específicos despertou o interesse de estudiosos de diversas áreas de conhecimento que observaram a importância dessas informações como fonte de conhecimento sobre o passado, o presente e a previsão do futuro dos assuntos abordados (GABER et al., 2015). Assim, a utilização de técnicas que permitam a análise e entendimento dessas informações se demonstra de grande interesse para a sociedade(CEDERMAN; WEIDMANN, 2017). Dessa forma, nesta seção apresentam-se as técnicas utilizadas no desenvolvimento da tese para gerar um analisador e preditor de eventos extremos utilizando dados de mídias sociais. Os conceitos apresentados são de mídia social, Processamento da Linguagem Natural, análise de sentimento, aprendizado de máquina e aprendizado profundo. Nas próximas seções apresentam-se com mais detalhes essas técnicas e conceitos.

3.1 Mídia social

O conceito de mídia social surgiu em meados dos anos de 1979 com a criação do sistema de publicação de mensagens "*online*", denominado "*Usenet*", por Tom Truscott e Jim Ellis da Universidade de Duke na Carolina do Norte, Estados Unidos. Por muitos anos, o conceito foi utilizado para designar qualquer meio de comunicação como revista, jornais e rádios. Esses meios se caracterizam por realizar a comunicação de mão única, na qual existe pouca interação com usuário. Com o surgimento das redes sociais, "*blogs*", "*wikis*" e sites de compartilhamento, o conceito de mídia social foi ampliado para abranger os aplicativos de interação com o usuário. Atualmente, a mídia social é caracterizada como um conjunto de aplicativos que se baseiam na Internet e que se fundamentam nas bases ideológicas e tecnológicas da Web 2.0 permitindo a criação e troca de conteúdo gerado pelo usuário (MOYER; KAPLAN, 2020).

Na tese, o conceito de mídia social é utilizado para designar os meios de circulação de informação de mão única: jornais e revistas. Os meios de comunicação com interação do usuário são denominados redes sociais, "*blogs*", "*wikis*" e sites de compartilhamento. As mídias sociais utilizadas nesta tese são as de grande circulação mundial ((REUTERS, 2019), (CNN, 2020), (GUARDIAN, 2020), etc) para eventos sociais extremos, como conflitos armados e de circulação nacional ((GLOBO, 2006), (BUTLER, 1999), etc) para eventos naturais extremos, como desmatamento.

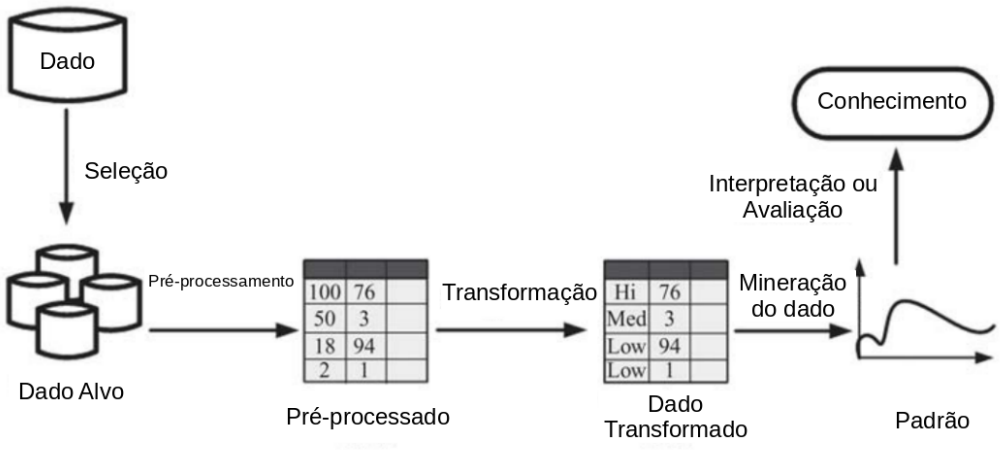
3.2 Ciência de dados

A Ciência de Dados é o estudo disciplinado de dados e informações inerentes ao negócio e todas as inferências que podem envolver um determinado assunto. É uma ciência que estuda a informação, seu processo de captura, transformação, geração e, posteriormente, análise dos dados. A ciência dos dados envolve várias disciplinas: computação, estatística, matemática e conhecimento de negócios (ZAFARANI et al., 2014).

A ciência de dados é usada nesta tese aplicando o conceito de mineração de mídia social. Esse conceito é uma disciplina emergente que realiza o processo de representação, análise e extração de padrões acionáveis de dados de mídia social (ZAFARANI et al., 2014).

Nesta tese a descoberta de padrões nos dados coletados nas mídias sociais são identificados por meio do processo denominado Descoberta de Conhecimento em Base de dados ou KDD da sigla em inglês de "*Knowledge Discovery in Databases (KDD)*". As fases básicas desse processo são mostradas na Figura 3.1.

Figura 3.1 - As fases do processo de descoberta de conhecimento (KDD) aplicadas na tese.



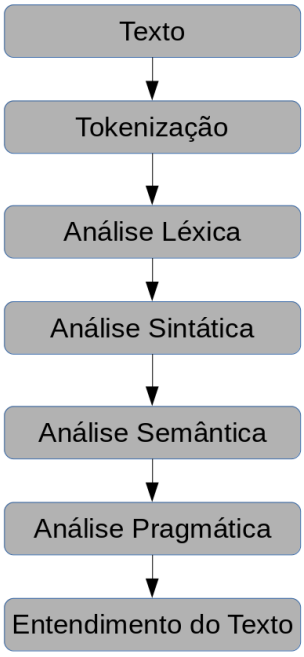
Fonte:Zafarani et al. (2014).

Na Seção 4.2, apresenta-se como o KDD foi aplicado na metodologia da tese.

3.3 Processamento da linguagem natural

O Processamento da Linguagem Natural ou NLP, sigla de ("*Natural Language Process*"), é uma subárea da inteligência artificial e possibilita o desenvolvimento de sistemas que permitem a interação entre computador e homem (JACKSON; MOULINIER, 2002) utilizando a linguagem natural humana seja por texto ou por fala (OLIVEIRA, 1990). A NLP pode ser dividida nos seguintes estágios apresentados na Figura 3.2 que serão detalhados a seguir.

Figura 3.2 - Estágios da NLP.



Fonte: Adaptado de Jackson e Moulinier (2002).

- **Tokenização:** A tokenização é muito bem caracterizada nas linguagens artificiais como as linguagens de programação, pois não possuem muitas ambiguidades. No entanto, na linguagem natural um mesmo caractere pode ter vários significados dependendo do seu contexto. Dessa forma, o uso da tokenização na análise da linguagem natural pode ser subdividida nas seguintes abordagens (JACKSON; MOULINIER, 2002):

  a) **linguagens delimitadas por espaços:** nessa categoria de linguagem, como as línguas europeias, os limites de palavras são indicados pela inserção de espaços em branco. No entanto, não necessariamente os símbolos entre os espaços são os "*tokens*" necessários para o proces-

samento seguinte. Isso ocorre devido à natureza ambígua dos sistemas de escrita quanto ao intervalo de convenções de tokenização existentes.

b) **linguagens não-segmentadas:** nessa categoria de linguagem, como chinês e tailandês, as palavras são escritas sucessivamente sem indicação de limites de palavras. Por isso, nesse caso a tokenização das linguagens não-segmentadas requer informações lexicais e morfológicas adicionais.

Assim, a tokenização para ser efetiva precisa ser processada com os estágios de análise léxica e análise sintática.

- **Análise Léxica:** A análise léxica realiza análise de texto ao nível da palavra. Uma tarefa básica da análise léxica é relacionar variantes morfológicas com sua respectiva palavra-chave. Essas palavras se encontram em um dicionário de palavras-chaves agrupadas com a sua informação semântica e sintática. No caso da análise da linguagem natural, a análise léxica desmembra as palavras de uma sentença em seus componentes gramaticais (substantivo, adjetivo, pronome, etc) (JACKSON; MOULINIER, 2002).

- **Análise Sintática:** A análise sintática é a tarefa de reconhecer uma frase e atribuir-lhe uma estrutura sintática. Essas estruturas sintáticas são atribuídas pela Gramática Livre de Contexto (CFG). A aplicação da CFG por algoritmos específicos gera uma representação em uma estrutura de árvore. Essas árvores analisam um importante estágio intermediário de representação para a análise semântica (JACKSON; MOULINIER, 2002).

- **Análise Semântica:** Na análise semântica as representações dos significados intermediários são compostas por expressões linguísticas. Essas representações de significados são atribuídas às frases com base no conhecimento adquirido com as fases léxica e gramatical. Assim, essa categoria de análise é utilizada no entendimento do significado de uma sentença. Também é muito utilizada para a eliminação de ambiguidades (JACKSON; MOULINIER, 2002).

- **Análise Pragmática:** A análise pragmática é a última etapa da análise da linguagem natural. Nessa fase, o significado é elaborado com base no conhecimento contextual e a forma lógica é mapeada para o idioma final de representação do assunto abordado. Essa análise também é utilizada para validar a análise semântica. Nessa análise considera-se que as palavras podem se associar por significados (água, nadar) ou de proximidade de assunto (água, poço) (JACKSON; MOULINIER, 2002).

3.4 Análise de sentimento

A área de Análise de Sentimento refere-se às tarefas de análise, identificação e classificação de toda informação que se caracterize de forma emocional, subjetiva ou geradora de opinião, seja a informação no formato de texto, imagem ou som (CUADRADO; GóMEZ-NAVARRO, 2011). Para a realização dessas tarefas de caracterização de sentimento normalmente utiliza-se as técnicas de Processamento da Linguagem Natural, estatísticas e/ou métodos de aprendizado de máquina. Segundo (CUADRADO; GóMEZ-NAVARRO, 2011), essas tarefas podem ser divididas da seguinte maneira.

a) **Classificação da Subjetividade:** Nesse tipo de classificação aborda-se a identificação de partes dos textos que demonstre um sentimento de subjetividade.

b) **Classificação de polaridade:** Essa classificação determina que os fragmentos de textos sejam classificados em sentimentos positivos ou negativos.

c) **Classificação de Intensidade:** Nessa classificação trabalha-se com a intensidade emocional expressa no texto. Essa abordagem costuma ser dividido nas classes: fortemente positivos, positivos, fortemente negativos, negativos ou neutros.

d) **Análise de sentimento com base em tópicos ou recursos:** Esse modelo de análise se baseia na verificação de características existentes relacionadas a sentimentos sobre o assunto abordado.

e) **Mineração de Opinião:** Essa abordagem está relacionada com a recuperação de informação de uma consulta. Assim, permite-se consultar um tópico específico e classificá-lo em determinada categoria.

Nos últimos anos, a análise de sentimento está sendo aplicada em diferentes formas de expressão de emoções/sentimentos além das formas textuais, utilizando principal-

mente informações das redes sociais. Na literatura atual da área, pode-se encontrar trabalhos que identificam sentimentos por sons (ABBURI et al., 2016), utilizando "*emoticons*" das redes sociais (BALABANTARAY et al., 2012) e imagens (ZHANG et al., 2015). Além disso, a análise de sentimento também está sendo utilizada nos mais diversos setores da sociedade. Nos últimos anos um setor que está voltando sua atenção para o estudo dos sentimentos da população é o governo. Em vários países estão sendo estudados os sentimentos da população em relação às políticas e programas lançados pelas diferentes áreas dos seus governos (municipal, estadual e federal) (ARUNACHALAM; SARKAR, 2013) (ZAVATTARO et al., 2015). No trabalho de (CEPIK, 2003) ressalta-se que uma forma de controle das atividades de inteligência é o conhecimento da opinião pública formada pela mídia. No Brasil algumas pesquisas na área envolvendo a visão da sociedade em relação a programas e políticas desenvolvidas já estão sendo realizadas. Uma dessas pesquisas é o trabalho de (ANTUNES et al., 2014) que analisa as "*hashtags*" das redes sociais para avaliar a adesão da população, por estado, à campanha de combate ao mosquito "*aedes aegypti*" transmissor do vírus da dengue entre outras doenças. Em relação aos conflitos sociais extremos, ainda não existe uma aplicação abrangente da análise de sentimento com aprendizado de máquina. O trabalho de (MUELLER; RAUH, 2020) apresenta a análise das mídias sociais como uma forma de previsão e prevenção de conflitos sociais em países pacíficos, utilizando técnicas Estatísticas, de Processamento da Linguagem Natural para a análise de sentimento e "*Random Forest*" para a previsão dos conflitos. Para a área de eventos extremos naturais, ainda não se identificou trabalhos utilizando o conjunto: mídias sociais, análise de sentimento e aprendizado de máquina.

3.5 Aprendizado de máquina

A área de Aprendizado de Máquina estuda os algoritmos computacionais que possibilitam a aprendizagem e sua consequente melhora por repetição de experiências. Dentro da área de Aprendizado de Máquina existem diversas aplicações que englobam desde Mineração de dados para a descoberta de regras gerais em um grande volume de dados até sistemas que aprendem de forma automática as necessidades do usuário (MITCHELL, 1997). O aprendizado pode ser subdividido nos seguintes paradigmas, segundo (POZZI et al., 2016) e (HAYKIN, 2008).

a) **Aprendizado Supervisionado:** neste paradigma os dados iniciais são rotulados com a verdade a ser alcançada para realização do ensinamento ou treinamento do modelo. Assim, no treinamento parte dos dados de entrada são comparados com os rótulos e corrigidos quando as saídas di-

ferem da verdade considerada. O modelo mais comum de desenvolvimento do aprendizado supervisionado é a aprendizagem por correção de erros. Essa técnica de aprendizado permite realizar os ajustes dos pesos visando a minimização do erro. Além dos problemas de regressão e classificação, o aprendizado supervisionado também aborda os problemas relacionados com o aprendizado profundo (BROWNLEE, 2019) (HAYKIN, 2008).

b) **Aprendizado Não-Supervisionado:** no paradigma não-supervisionado considera-se as características existentes nos dados de entrada, possibilitando a extração e a generalização de regras. Tal modelo pode se basear tanto em processos matemáticos que reduzem a redundância dos dados quanto pela organização dos dados por similaridade. Essa abordagem utiliza a aprendizagem competitiva, na qual os neurônios de saídas competem entre si para serem mais ativos. Como problemas nos quais o modelo é aplicável estão agrupamento de dados, redução de dimensionalidade e o aprendizado profundo (BROWNLEE, 2019) (HAYKIN, 2008).

c) **Aprendizado Semi-supervisionado:** Na categoria de aprendizado semi-supervisionado é construído um modelo híbrido no qual se utiliza o aprendizado supervisionado e o não-supervisionado. As características dos dois modelos de aprendizado são aplicadas nos dados de entrada do semi-supervisionado, que utiliza dados com rótulos e dados não rotulados em seus dados de entrada (BROWNLEE, 2019) (POZZI et al., 2016). O aprendizado semi-supervisionado também pode ser usado para agrupamento de dados. Nesse caso, a supervisão é fornecida pelo especialista via "*constraints must-link*" e "*cannot-link*". Ou seja, informa-se, a priori, alguns pares de exemplos que devem estar num mesmo grupo ("*cluster*") e pares que não devem ser agrupados ("*cannot-link*") (CHAPELLE et al., 2006). Essa categoria de aprendizado também é comumente utilizada em problemas de classificação, clusterização, regressão e aprendizado profundo.

d) **Aprendizado por Reforço:** No Aprendizado por Reforço a relação dos dados de entrada e saída (resultado) realiza-se com uma interação contínua com o ambiente. Dessa forma, o modelo aprende por reforço atrasado que representa a observação de uma sequência temporal de estímulos resul-

tando na geração do sinal de reforço heurístico. A cada iteração o algoritmo recebe um sinal de recompensa que indica a boa precisão da decisão tomada ou uma punição caso contrário. Uma aplicação comum desse modelo de aprendizado é a robótica. (MICROSOFT CORPORATION, 2017) (HAYKIN, 2008). As duas características mais importantes da aprendizagem por reforço são: tentativa, erro e recompensa atrasada. Esse modelo de aprendizado possui como elementos o agente e o ambiente e como subelementos principais, os quatro itens listados a seguir (SUTTON; BARTO, 2012):

- **Política:** define o comportamento do agente em um determinado momento.
- **Função de recompensa:** define a meta de um problema no modelo de aprendizagem por reforço.
- **Função de valor:** especifica quais ações foram boas ao longo da execução do modelo.
- **Modelo do Ambiente:** reproduz o comportamento do ambiente.

Nesta tese, estar-se-á aprofundando no estudo e aplicação das técnicas de aprendizado profundo e "*LSTM - Long short-term memory*". Nas próximas seções essas técnicas serão detalhadas.

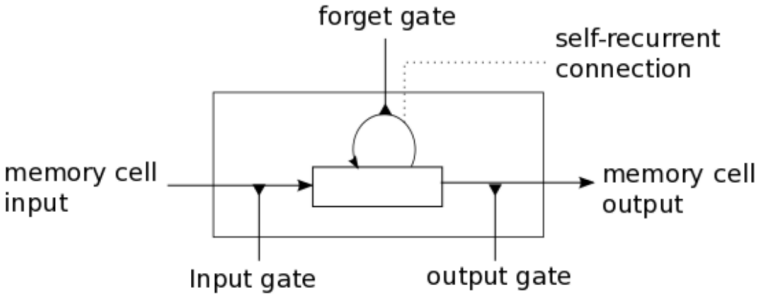
3.5.1 Aprendizado profundo

O Aprendizado Profundo é uma subárea da inteligência artificial que possibilita a criação de modelos de rede neural com grandes dimensões que podem ser ensinados a tomarem decisões com alta precisão baseados nas características dos dados (GOODFELLOW et al., 2016). Para a construção desses modelos pode-se utilizar diversas arquiteturas de redes. Na tese, utiliza-se uma rede neural recursiva conhecida como LSTM. O uso do aprendizado profundo remonta da década de 1960 e desde então uma abundância de pesquisas, em diversas áreas do conhecimento, estão sendo desenvolvidas utilizando esse método. Nos últimos anos, a melhoria no desempenho de processamento dos computadores possibilitou o aumento da utilização e a melhoria na precisão dos resultados dos cálculos utilizando aprendizado profundo. Assim, verifica-se que uma importante característica do modelo é o processamento de grandes volumes de dados (INSS CONFERENCE ON BIG DATA, 2015). O aprendizado profundo abrange diversas áreas de pesquisa dentre elas: modelagem gráfica, otimização, reconhecimento de padrões, processamento de sinais e Processamento da Linguagem Natural(DENG; YU, 2013) (SCHMIDHUBER, 2015).

3.5.2 "*Long short-term memory*" - LSTM

As redes "*Long Short-Term Memory*" são um modelo de rede neural recorrente capaz de aprender dependência de ordem em problemas de predição de sequência (BROWNLEE, 2017). A LSTM consiste em uma célula ("*cell*"), um portão de entrada ("*input gate*"), um portão de saída ("*output gate*") e um portão de esquecimento ("*forget gate*"). A célula lembra valores de intervalos de tempo arbitrários e os três portões regulam o fluxo de informações dentro e fora da célula. Os portões controlam as interações entre as células de memória vizinhas e a própria célula de memória. O portão de entrada adiciona informações úteis ao estado atual da célula. Por outro lado, o portão de saída extraí informações úteis do estado atual da célula para passar para a próxima célula. Além disso, o portão de esquecimento excluir informações que não contribuem para o estado atual da célula. (BAO et al., 2017). A Figura 3.3 apresenta um exemplo dessa arquitetura de rede.

Figura 3.3 - Ilustração de uma célula de memória LSTM.



Fonte: Bao et al. (2017).

Nesta tese, a rede neural LSTM é utilizada para realizar a predição dos eventos extremos sociais e naturais. Esse processo é realizado a partir de uma série temporal de ameaça gerada por intermédio das notícias coletadas nas mídias sociais e processadas pela técnica de análise de sentimento para a extração da emoção de ameaça. Essa série temporal de ameaça é processada pelo algoritmo "*P-Model*" para gerar a série endógena de ameaça. Esses dados endógenos de ameaças são processados pela rede neural LSTM para se realizar a predição da variação de ameaça representada pela série temporal endógena.

A utilização do aprendizado profundo com a rede neural LSTM foi realizado na tese por intermédio da API ("*Application Programming Interface*") de aprendizado

profundo "*Keras TensorFlow*". Na Seção 3.5.2.1 é apresentada as principais características dessa API.

3.5.2.1 API de aprendizado profundo "*Keras TensorFlow*"

A "*Keras*" é uma API "*open source*" que trabalha com redes neurais de alto nível e foi escrita na linguagem de programação "*Python*". O seu propósito é a habilitação de experimentação rápida, ou seja, tem o foco no resultado no menor tempo possível. As principais características de aprendizado profundo dessa API são (CHOLLET, 2015):

- Permite prototipagem fácil e rápida (pela facilidade de utilização, modularidade e extensibilidade).
- Suporta redes convolutivas e redes recorrentes, bem como combinações das duas categorias de rede.
- Funciona perfeitamente em processamento com CPU e com GPU.

A "*Keras*" se baseia nos seguintes princípios guiadores (CHOLLET, 2015):

- **Facilidade de uso:** segue as melhores práticas para reduzir a carga cognitiva: fornece APIs consistentes e simples, minimiza o número de ações do usuário necessárias para casos de uso comum e fornece "*feedback*" claro e rápido sobre o erro do usuário.
- **Modularidade:** na "*Keras*" as camadas neurais, funções de custo, otimizadores, esquemas de inicialização, funções de ativação, esquemas de regularização são todos módulos autônomos que você pode combinar para criar novos modelos.
- **Fácil extensibilidade:** novos módulos são simples de se adicionar (como novas classes e funções) e os módulos existentes fornecem exemplos amplos.
- **Trabalha com a "*Python*":** não há arquivos de configuração de modelos separados em um formato declarativo. Os modelos são descritos no código "*Python*", que é compacto, fácil de depurar e permite a extensibilidade.

As estruturas de dados principais da "*Keras*" são uma forma de organizar as camadas ("*layers*") e um modelo ("*models*"). O modelo mais simples é o modelo sequencial,

que é representado por uma pilha linear de camadas. Para arquiteturas mais complexas, existe API funcional, que permite a construção de gráficos arbitrários de camadas. A biblioteca "*Keras*" também contém implementações de redes neurais de uso comum como pré-processamento ("*pre-processing*") (CHOLLET, 2015). Cada uma dessas implementações serão detalhadas nas próximas seções.

a) **Camadas ("*Layers*")** As camadas da API "*Keras*" permite adicionar funcionalidades aos modelos criados. A "*Keras*" disponibiliza doze "*layers*". Na Tabela 3.1 apresenta-se alguns desses "*layers*" (CHOLLET, 2015).

Tabela 3.1 - "*Layers*" disponíveis na API "*Keras*"

"*Layers*"	**Função**
"Core" "Layers"	permite criar modelos de redes neurais
"Convolutional" "Layers"	gera uma rede neural convolutiva $1D$ ou $2D$
"Recurrent" "Layers"	gera uma rede neural recorrente
"Writing your own Keras" "Layers"	possibilita ao usuário elaborar "*Layers*" que atendam funcionalidades especificas com treinamento com pesos

b) **Modelos ("*Models*")** A API "*Keras*" disponibiliza dois modelos: o modelo sequencial e a classe "*Model*", usada com a API funcional. Esses dois modelos possuem as seguintes características.

- **Modelo Sequencial:** consiste em uma pilha linear de "*layers*". Tal modelo é criado por meio da passagem de uma lista de instâncias de "*layers*" para o seu construtor.
- **"*Model*" - API funcional:** utilizado na definição de modelos complexos, com modelos de várias saídas, gráficos acíclicos direcionados ou modelos com camadas compartilhadas.

Na tese, está se aprofundando os estudos do Modelo Sequencial para a construção da rede neural "*LSTM*". Os "*Models*" possuem as seguintes sintaxes:

```
model = Model(inputs, outputs)
```

Para cada tipo de dados o "*Model*" pode receber os seguintes parâmetros:

- **"*model.layers*":** lista das camadas que compõe o gráfico do modelo.
- **"*model.inputs*":** lista de tensores de entrada.
- **"*model.outputs*":** lista de tensores de saída.

c) **Pré-processamento ("*Preprocessing*")** As funções de pré-processamento da "*Keras*" permitem manipular os dados de entrada para uma melhor utilização nos modelos. A API oferece três tipos de implementação de pré-processamento, como apresentado a seguir:

- **"*Sequence Preprocessing*":** transforma a lista de dados em "*Numpy Array*" ou tuplas.
- **"*Text Preprocessing*":** realiza a tokenização, separação em palavras ou indexização de um texto.
- **"*Image Preprocessing*":** gera lotes de dados das imagens lidas.

Na tese, realizou-se a análise do "*Core Layer*", "*model Sequential*" e "*Text Preprocessing*" utilizando-se dados de mídias sociais como jornais e revistas (REUTERS, 2019), (BUTLER, 1999), (GUARDIAN, 2020), (FOLHA, 1996), etc.

d) **Funções de métricas e perdas**

A API "*Keras*" disponibiliza um conjunto de funções de métricas e perdas para realizar o julgamento do desempenho do modelo utilizado e o cálculo da quantidade de otimização durante o treinamento, respectivamente. As funções de métricas e suas classes da API são subdivididas em:

- **Métricas de precisão:** "*Accuracy*", "*BinaryAccuracy*", "*CategoricalAccuracy*", "*TopKCategoricalAccuracy*" e "*SparseTopKCategoricalAccuracy*".
- **Métricas probabilísticas:** "*BinaryCrossentropy*", "*CategoricalCrossentropy*", "*SparseCategoricalCrossentropy*", "*KLDivergence*" e "*Poisson*".
- **Métricas de regressão:** "*MeanSquaredError*", "*RootMeanSquaredError*", "*MeanAbsoluteError*", "*MeanAbsolutePercentageError*", "*MeanSquaredLogarithmicError*", "*CosineSimilarity*" e "*LogCoshError*".
- **Métricas de classificação baseadas em verdadeiros / falsos positivos e negativos:** "*AUC*", "*Precisio*", "*Recall*", "*TruePositives*",

"*TrueNegatives*", "*FalsePositives*", "*FalseNegatives*", "*PrecisionAtRecall*", "*SensitivityAtSpecificity*" e "*SpecificityAtSensitivity*".

- **Métricas de segmentação de imagem:** "*MeanIoU*".
- **Métricas de dobradiça para classificação de "margem máxima":** "*Hinge*", "*SquaredHinge*" e "*CategoricalHinge*".

Já as funções de perdas da API são subdivididas em:

- **Perdas Probabilísticas:** "*BinaryCrossentropy*", "*CategoricalCrossentropy*", "*SparseCategoricalCrossentropy*", "*Poisson*", "*binary_crossentropy*", "*categorical_crossentropy*", "*sparse_categorical_crossentropy*", "*poisson*", "*KLDivergence*" e "*kl_divergence*".
- **Perdas de Regressão:** "*MeanSquaredError*", "*MeanAbsoluteError*", "*MeanAbsolutePercentageError*", "*MeanSquaredLogarithmicError*", "*CosineSimilarity*", "*mean_squared_error*", "*mean_absolute_error*", "*mean_absolute_percentage_error*", "*mean_squared_logarithmic_error*", "*cosine_similarity*", "*Huber class*", "*huber function*", "*LogCosh class*" e "*log_cosh function*".
- **Perdas de Dobradiça para Classificação de "margem máxima":** "*Hinge class*", "*SquaredHinge*", "*CategoricalHinge*", "*hinge function*", "*squared_hinge*" e "*categorical_hinge*".

e) **Função de Otimização**

As funções de otimização ou os otimizadores da "*Keras*" são responsáveis por definir como os pesos das redes neurais são atualizados. A "*Keras*" disponibiliza os seguintes otimizadores: "*SGD*", "*RMSprop*", "*Adam*", "*Adadelta*", "*Adagrad*", "*Adamax*", "*Nadam*" e "*Ftrl*".

As funções de métricas, perdas e otimização são aplicadas na API "*Keras*" por meio do comando "*compile*", que possue a seguinte sintaxe.

```
model.compile(optimizer='adam', loss='mse',
metrics=[tf.keras.metrics.RootMeanSquaredError()])
```

Na tese, utiliza-se a métrica de regressão com a classe "*RootMeanSquaredError*", a função de perda de regressão com a classe "*MeanSquaredError*" e o otimizador "*Adam*". Essas classes e funções foram selecionadas devido a necessidade de extração de caracterísiticas pela rede neural utilizada dos dados coletados das mídias sociais. A função de "*Loss*" apresenta a informação de quão bom os rótulos de classe preditos concordam com os rótulos de verdade considerada para o treinamento. Assim, quanto maior o nível de concordância entre esses dois conjuntos de rótulos, menor será a perda e maior será a precisão da predição para os dados de treinamento. Já a informação de "*Val_loss*" apresenta a mesma informação para os dados de validação (CHOLLET, 2015). A métrica RMSE ou Raiz Quadrada do Erro Quadrático apresenta a informação da diferença entre o valor que foi previsto pelo modelo criado e o valor real que foi observado (CHOLLET, 2015). O otimizador "*Adam*" é um método gradiente estocástico de descida que apresenta uma boa resposta para uma grande quantidade de dados (CHOLLET, 2015).

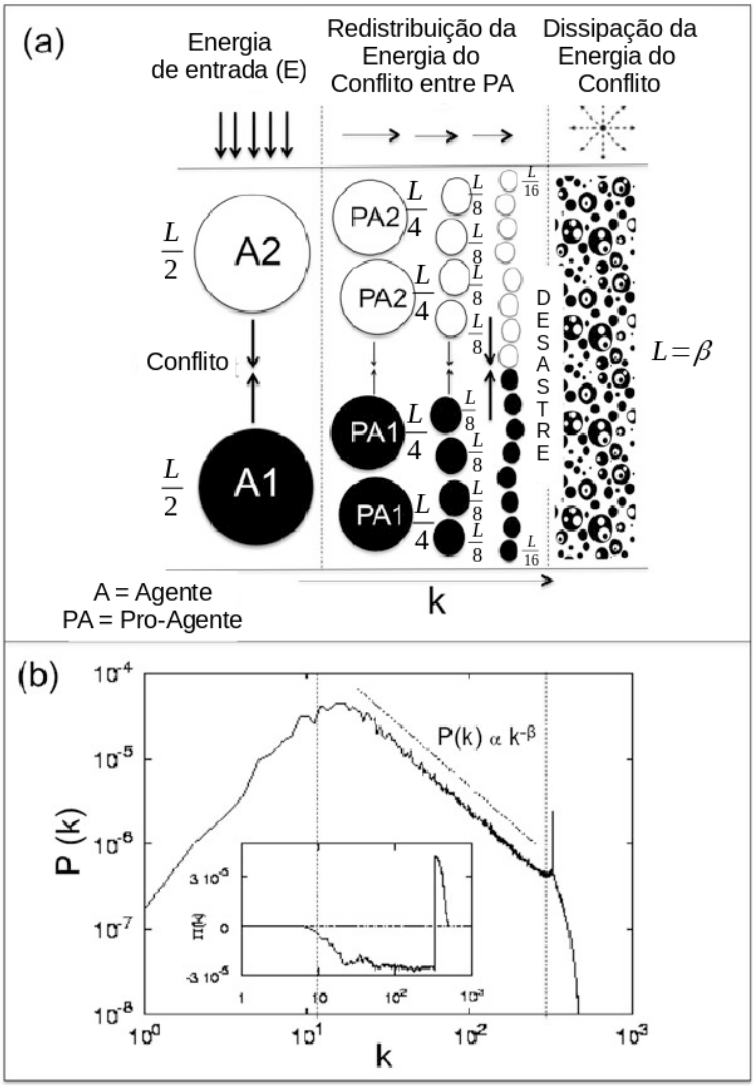
3.6 Série temporal

As informações coletadas das mídias sociais são processadas pela técnica de análise de sentimento para a identificação da emoção e grau de ameaça para cada notícia analisada de cada assunto abordado. Posteriormente, esses graus de ameaças de cada grupo de notícias são processados pelo algoritmo "*P-Model*" e transformados em uma série temporal de ameaça com características endógenas. Posteriormente, essa série é utilizada para a realização da predição da variação de ameaça dos eventos extremos sociais e naturais. O algoritmo "*P-Model*" utilizado na tese, é apresentado com mais detalhes na Seção 3.6.1.

3.6.1 Algoritmo "*P-Model*"

O "*P-Model*" foi proposto por (MENEVEAU; SREENIVASAN, 1987) como uma cascata turbulenta não-homogênea. Esse algoritmo foi utilizado por (ROSA et al., 2019) para criar uma cascata não-homogênea que é compatível com as flutuações observadas nas séries temporais estocásticas. Segundo (ROSA et al., 2019), essa cascata não homogênea é compatível com a energia dissipada por eventos extremos até o momento do seu ápice, no qual se tem a máxima dissipação de energia. A Figura 3.4 apresenta um exemplo dessa dissipação de energia para eventos extremos referente a conflitos armados.

Figura 3.4 - Cascata multiplicativa para conflitos armados. (a) Hierarquia de escala para conflitos armados; (b) Respectivo padrão de densidade de espectro de energia esperada mostrando a transição da faixa inercial para o evento extremo que é a resposta ao alto regime dissipativo.



Fonte: Adaptado de Rosa et al. (2019).

A cascata multiplicativa do "*P-Model*" é representada pelas equações 3.1 e 3.2, sendo definidas em Halsey et al. (1987).

$$\alpha = \frac{\log_2 p_1 + (w-1)\log_2 p_2}{\log_2 l_1 + (w-1)\log_2 l_2} \tag{3.1}$$

e

$$f(\alpha) = \frac{(w-1)\log_2(w-1) - w\log_2 w}{\log_2 l_1 + (w-1)\log_2 l_2} \tag{3.2}$$

onde,

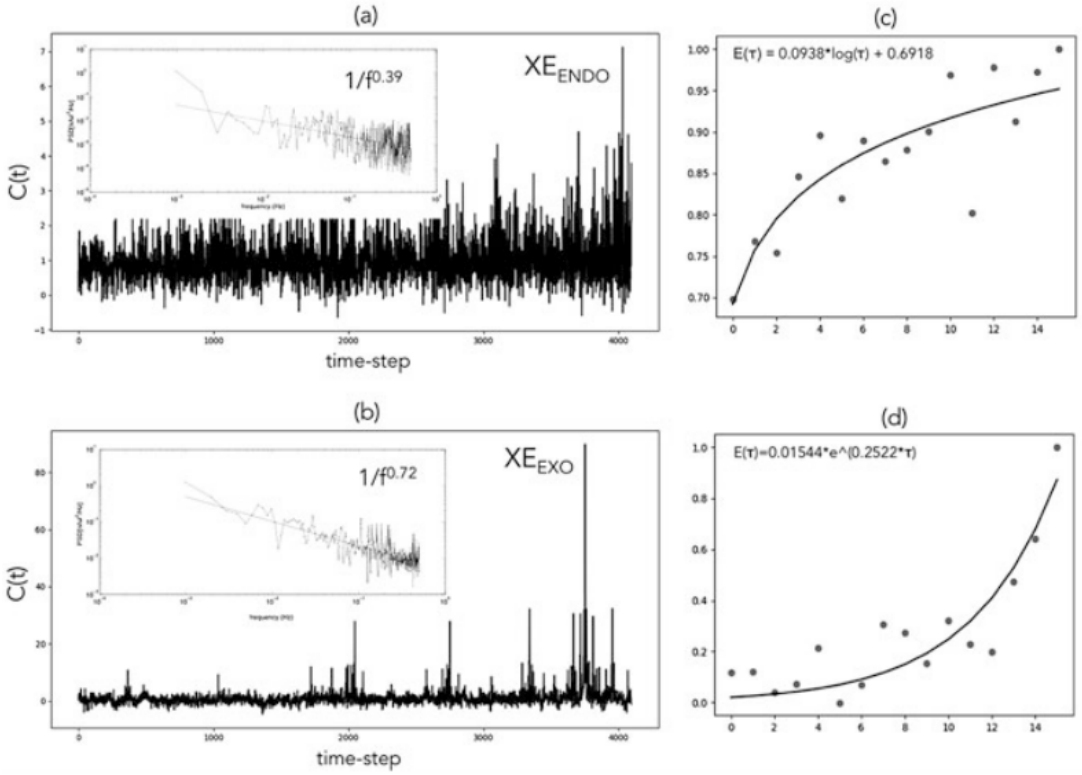
α - força da singularidade
p_i - probabilidade de que algum evento ocorrer na $i-ésima$ fração, para $i = 1, 2$
w - peso mutiplicativo dado por $1 - (1 - 2p)$
l_i - $i-ésima$ fração de um "*eddy*" de tamanho L, para $i = 1, 2$
$f(\alpha)$ - descreve como as singularidades são densamente distribuídas

Segundo a utilização em (ROSA et al., 2019), de forma generalizada, o "*P-Model*" apresenta uma visão clássica da cascata de "*eddy*" antes do intervalo inercial da turbulência totalmente desenvolvida, na qual o fluxo de energia (EK), como apresentado na Figura 3.4 (a), dissipa-se no comprimento da escala β de Kolmogorov (KEYLOCK, 2017) em "*eddies*" de tamanho L. Posteriormente, cada "*eddy*" de tamanho L é dividida em duas partes iguais, sendo representadas como $L/2$, l_1 e l_2. Dessa forma, em cada etapa da cascata, o fluxo de energia é distribuído, como uma probabilidade, desigualmente em fração de p_1 e $p_2 = 1-p_1$, no qual $p_1+p_2 = 1$. Esse processo é iterado sobre p_1 fixo até que cada "*eddy*" atinja a escala β de Kolmogorov (KEYLOCK, 2017). Começando com uma distribuição de energia não-homogênea, uma fração $f(\alpha)$ da massa multifractal é transferida de uma metade para a outra na direção escolhida aleatoriamente. Isso é equivalente a multiplicar o campo de densidade originalmente uniforme em ambos os lados por fatores. O mesmo procedimento é repetido M vezes, recursivamente em escalas cada vez menores, usando frações que variam de α em segmentos de comprimento $\frac{L}{2n}$, no qual o peso multiplicativo w é parametrizado como $1 - (1 - 2p)$, resultando a matriz discreta $C(m)$ onde m conta como passos de tempo. Esse procedimento do algoritmo "*P-Model*", dado por Venema (BONN, 2019), pode produzir séries temporais nas quais a variância é finita se você extrapolar seu espectro de potência para escalas infinitamente grandes (ROSA et al., 2019).

A série temporal, $C(m = t)$, com $M = 211$, representando o evento extremo turbulento não homogêneo é gerada usando o algoritmo de Venema (BONN, 2019), nas quais as entradas são: o tamanho da série temporal em número de pontos (M), o espectro de potência PSD (β_{PSD}), e o valor de p, sendo a distribuição fracionária de

energia em cascata semelhante à turbulência não homogênea (ROSA et al., 2019). O processo dissipativo homogêneo próximo ao equilíbrio termodinâmico é recuperado quando $(\beta_{PSD}, p) = (-5/3, 0.5)$. Os desvios da cascata homogênea são agravados por mudanças abruptas na frequência e magnitude do conflito social. Essas alterações são chamadas de eventos extremos (XE) e sua causa pode ser devida a fatores mais internos do que externos. Quando o nível de conflito aumenta significativamente devido a interações internas, o evento extremo é denominado endógeno ($XEendo$). Quando a transferência de energia externa ou dissipação abrupta é a principal causa de XE, é chamado de evento extremo exógeno ($XEexo$). No domínio da lei de potência, eventos do tipo $XEendo$ e $XEexo$ pertencem a diferentes classes de universalidade (ROSA et al., 2019). Os processos endógenos e exógenos típicos, combinando o algoritmo "*P-Model*" e SDGA ("*Sornette-Deschtres-Gilbert-Ageo*") (SORNETTE, 2006), são obtidos para $(\beta_{PSD} \approx -0.4, p = 0.25)$ e $(\beta_{PSD} \approx -0.7, p = 0.25)$, respectivamente. A Figura 3.5 mostra a série temporal XE para diferentes combinações de p e β_{PSD}. A energia cumulativa do processo no domínio do tempo é definida como média normalizada $< C(\tau) >$ onde τ é um intervalo de tempo da janela escolhido ao longo do sinal. De acordo com Rosa et al. (2019), as tendência de energia cumulativa típicas não são lineares, sendo logarítmicas para $XEendo$ (Figura 3.5 (c)) e exponenciais para $XEexo$ Figura (3.5 (d)).

Figura 3.5 - Exemplo de duas séries temporais típicas timuladas a partir do algoritmo "*P-Model*" fixando $p = 0.25$ e variando o valor de β. (a) Um padrão $XEendo$ de $\beta = -0.39$. (b) Um padrão $XEexo$ de $\beta = -0.72$. Os respectivos PSDs são mostrados usando imagens menores incorporadas. As Figuras (c) e (d) mostram a respectiva energia cumulativa que expressa os padrões funcionais típicos para endogenia (log) e exogenia (exponencial).



Fonte: Adaptado de Rosa et al. (2019).

Na tese, o "*P-Model*" é utilizado como um gerador de séries temporais endógenas, considerando os valores de entrada: $M = 2^{14}$, $p = -0.60$ e β, denominado "*slope*" e calculado utilizando o conceito de DFA ("*Detrended Fluctuation Analysis*"), sendo apresentado na Seção 3.6.1.1.

3.6.1.1 "*Detrended fluctuation analysis*" - DFA

O método de DFA foi projetado por (C.-K.PENG et al., 1994) para investigar a correlação de longo alcance em séries não-estacionárias (MORARIU et al., 2007). O espectro via DFA tem sido utilizado para o estudo e caracterização de processos turbulentos em fluidos neutros (VELHO et al., 2001) (RAMOS et al., 2004) e plasmas (NEELAKSHI et al., 2019). No algoritmo "*P-model*", o índice espectral β, via PSD ("*Power-Sppectral Density*") está associado ao índice espectral α obtido via DFA. O índice α pode ser obtido analiticamente a partir do índice β, o que permite a comparação de desempe-

nho entre ambos (VERONESE et al., 2011). O DFA, segundo (HARVARD, 2019), pode ser aplicado considerando os seguintes passos.

- a série temporal, com N amostras, a ser analisada, é inicialmente integrada
- a série temporal integrada é dividida em caixas de igual comprimento, n
- em cada caixa de comprimento n, uma linha de quadrados mínimos é ajustada aos dados (representando a tendência nessa caixa)
- a coordenada y dos segmentos de linha reta é denotada por $y_n(k)$
- diminuí-se a série temporal integrada, $y(k)$, subtraindo a tendência local, $y_n(k)$, em cada caixa
- calcula-se a flutuação da raiz quadrada média desta série temporal integrada e sem tendência utilizando a Equação 3.3

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[y(k) - y_n(k)]^2} \tag{3.3}$$

Esse cálculo é repetido em todas as escalas de tempo (tamanhos de caixa) para caracterizar a relação entre $F(n)$, a flutuação média, em função do tamanho da caixa. Normalmente, $F(n)$ aumentará com o tamanho da caixa n. Uma relação linear em um gráfico log-log indica a presença de escala de lei de potência (fractal). Sob tais condições, as flutuações podem ser caracterizadas por um expoente de escala, a inclinação da linha que relaciona $log(F(n))$ a $log(n)$ (HARVARD, 2019).

Na tese, o cálculo de DFA é realizado utilizando a função "*DFA()*" da biblioteca "*nops*" da linguagem de programação "*Python 3.7*". A entrada da função é a série de ameaça, gerada pela aplicação da técnica de análise de sentimento nos dados coletados das mídias sociais. O valor calculado é utilizado como entrada do algoritmo "*P-Model*", para a geração da série endógena de ameaça, como apresentada na Seção 4.4.

3.6.2 Média aritmética

Na tese, a média aritmética é aplicada para verificação das séries temporais endógenas na Seção 5.3.1.1 e aplicada por meio da função "*mean*" da biblioteca "*scipy*"

encontradas na linguagem *Python 3.7*. A média aritmética utilizada pela função "*mean*" é representada pela Equação 3.4.

$$\bar{X} = \sum \frac{X_i}{n} \tag{3.4}$$

onde,

$n-$ tamanho do conjunto de dados
X_i- iésimo valor de X

3.6.3 Variância

Na tese, a variância também foi utilizada para a verificação das séries temporais endógenas na Seção 5.3.1.1 e aplicada por meio da função "*variance*" da biblioteca "*statistics*" encontradas na linguagem *Python 3.7*. A variância utilizada pela função "*variance*" é representado pela Equação 3.5.

$$v = \sum \frac{(X_i - \bar{X})^2}{n-1} \tag{3.5}$$

onde,

$n-$ tamanho do conjunto de dados
X_i- iésimo valor de X
$\bar{X}-$ média dos valores de X

3.6.4 Desvio padrão

O desvio padrão, s, utilizado na Equação 3.7 de "*Skewness*" é representado pela Equação 3.6.

$$s = \sqrt{v} \tag{3.6}$$

onde,

$v-$ é a variância do conjunto de dados.

3.6.5 Assimetria

A assimetria é uma medida da simetria de uma distribuição. A assimetria também é definido como a medida relativa em duas caudas ("*tails*). Dessa forma, quanto mais o valor de assimetria' se aproxima de zero mais se considera o conjunto da dados simétrico (WHEELER, 2011). A assimetria é representado pela Equação 3.7.

$$a_3 = \sum \frac{(X_i - \bar{X})^3}{ns^3} \tag{3.7}$$

onde,

$n-$ tamanho do conjunto de dados
X_i- iésimo valor de X
$\bar{X}-$ média dos valores de X
$s-$ desvio padrão do conjunto de dados

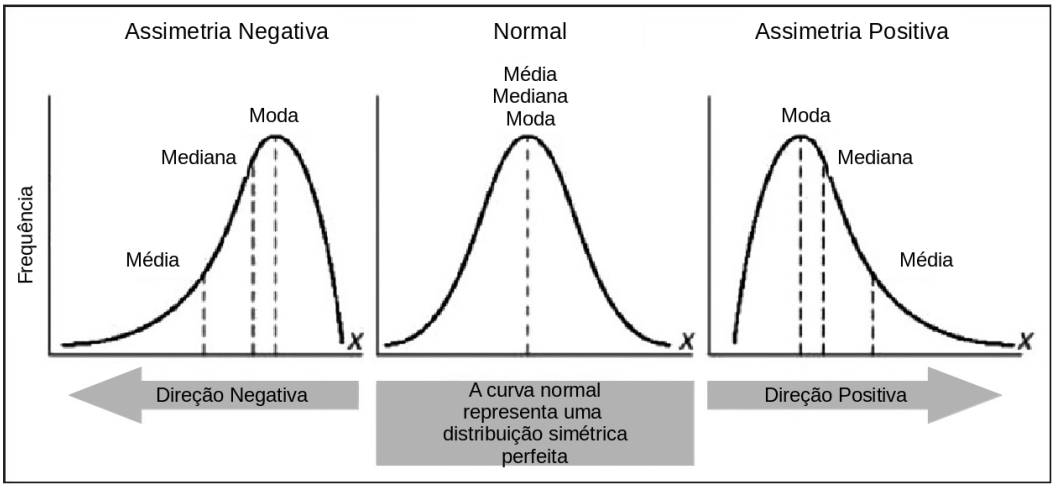
O parâmetro de assimetria também é referenciado como terceiro momento central padronizado para o modelo de probabilidade, justificando o expoente 3 do somatório. De acordo com Groeneveld (1991), a assimetria pode ser classificada em:

- $a_3 > 0$ - a calda da distribuição aponta para a direita e o centro de massa da distribuição se localiza à esquerda.
- $a_3 < 0$ - a calda da distribuição é deslocada para a esquerda e o centro de massa da distribuição de localiza à direita.
- $a_3 = 0$ - distribuição é aproximadamente simétrica (na terceira potência do desvio em relação à média).

A Figura 3.6 apresenta um exemplo do posicionamento da calda para cada classificação de assimetria.

Na tese, a assimetria é aplicada para verificação da variação de ameaça nas séries temporais endógenas na Seção 5.3.1.1. Considera-se que a assimetria negativa possa

Figura 3.6 - Exemplo de assimetria negativa, neutra e positiva



Fonte: Adaptado de UNIVERSITY OF SOUTH ALABAMA (2014).

representar um possível aumento nas hipóteses de ocorrerem algum evento extremo devido à característica da distribuição de cauda à esquerda, que está presente nos dados das séries temporais. A assimetria é aplicada por meio da função "*skew*" da biblioteca "*scipy*" encontradas na linguagem *Python 3.7*.

3.6.6 Curtose

O parâmetro de estatística curtose é definido como uma medida que combina os pesos das caudas ("*tails*) em relação ao restante da distribuição (WHEELER, 2011). O curtose pode ser definido pela Equação 3.8 (WHEELER, 2011).

$$a_4 = \sum \frac{(X_i - \bar{X})^4}{ns^4} \tag{3.8}$$

onde,

$n-$ tamanho do conjunto de dados
X_i- iésimo valor de X
$\bar{X}-$ média dos valores de X
$s-$ desvio padrão do conjunto de dados

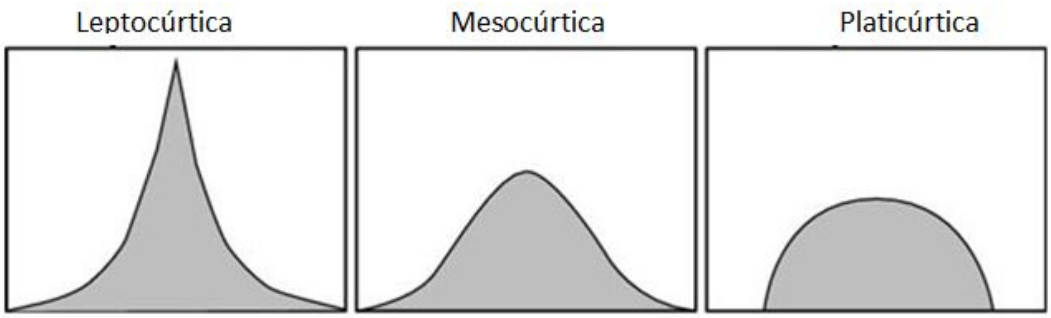
O parâmetro de curtose também é referenciado como quarto momento central padronizado para o modelo de probabilidade, justificando o expoente 4 do somatório.

Segundo (BROWN, 2020), uma distribuição normal, tem uma curtose igual a 3 e o excesso de curtoses é igual a-3. Assim, (HAYES; JAMES, 2021) apresenta a seguinte classificação para o excesso de curtose.

- Uma distribuição **normal** tem curtoses exatamente 3 (excesso de curtose exatamente 0).
- Uma distribuição com curtose ≈ 3 (excesso ≈ 0) é denominada de **mesocúrtica**.
- Uma distribuição com curtose' < 3 (excesso < 0) é denominada de **platicúrtica**. Tal distribuição possui caldas mais curtas e finas e, frequentemente, seu pico central é mais baixo e mais largo, comparados com uma distribuição normal.
- Uma distribuição com curtose > 3 (excesso > 0) é denominada de **leptocúrtica** e possui caldas mais longas e mais grossas, e frequentemente seu pico central é mais alto e mais nítido, comparados com uma distribuição normal, resultando em uma maior oportunidade de eventos positivos ou negativos extremos.

A Figura 3.7 apresenta um exemplo dessas distribuições da curtoses para a verificação do posicionamento dos picos e das caldas.

Figura 3.7 - Exemplo das distribuições leptocúrtica, platicúrtica e mesocúrtica.



Fonte: Adaptado de Vieira (2018).

Na tese, utilizam-se os conceitos de curtose leptocúrtica como parâmetro de verificação das séries temporais endógenas de ameaças devido às características de seus picos elevados e de suas caldas representarem uma probabilidade maior de ocorrência de um evento extremo considerando a variação de ameaças encontrada nos dados

das séries temporais, como apresentado Seção 5.3.1.1. A curtose é aplicada por meio da função "*kurtosis*" da biblioteca "*scipy*" encontradas na linguagem *Python 3.7*.

3.6.7 "*Dynamic time warping*" - DTW

Nesta tese utiliza-se o algoritmo "*Dynamic time warping*"para comparar o resultado de predição de parte da série temporal endógena com a parte original da série endógena, apresentado na Seção 5.3 de resultados da tese. O DTW é utilizado para o cálculo da dissimilaridade entre duas séries temporais de mesma dimensão (GIUSTI; BATISTA, 2013) ou seja na tese o valor de DTW representa o grau de diferença entre as duas partes de séries. Na tese, o DTW foi utilizado por meio do pacote "*tslearn.metrics.dtw*" (TAVENARD et al., 2020) na linguagem "*Python 3.7*". A fórmula que representa o cálculo do DTW pode ser visualizada na Equação 3.9

$$DTW(x, y) = \sqrt{\sum_{(i,j\epsilon\pi)} (X_i - Y_j)^2} \tag{3.9}$$

onde,

X e $Y-$ representam as séries a serem comparadas
i e $j-$ representam as posições i,j em uma matriz de cada elemento das séries
$\pi-$ representa o conjunto de caminhos entre as séries

Nesta seção, apresentaram-se conceitos fundamentais de todas as técnicas utilizadas no desenvolvimento da tese. A aplicação dessas técnicas é descrita e detalhada na Seção 4, que apresenta a Metodologia de desenvolvimento da tese.

4 METODOLOGIA

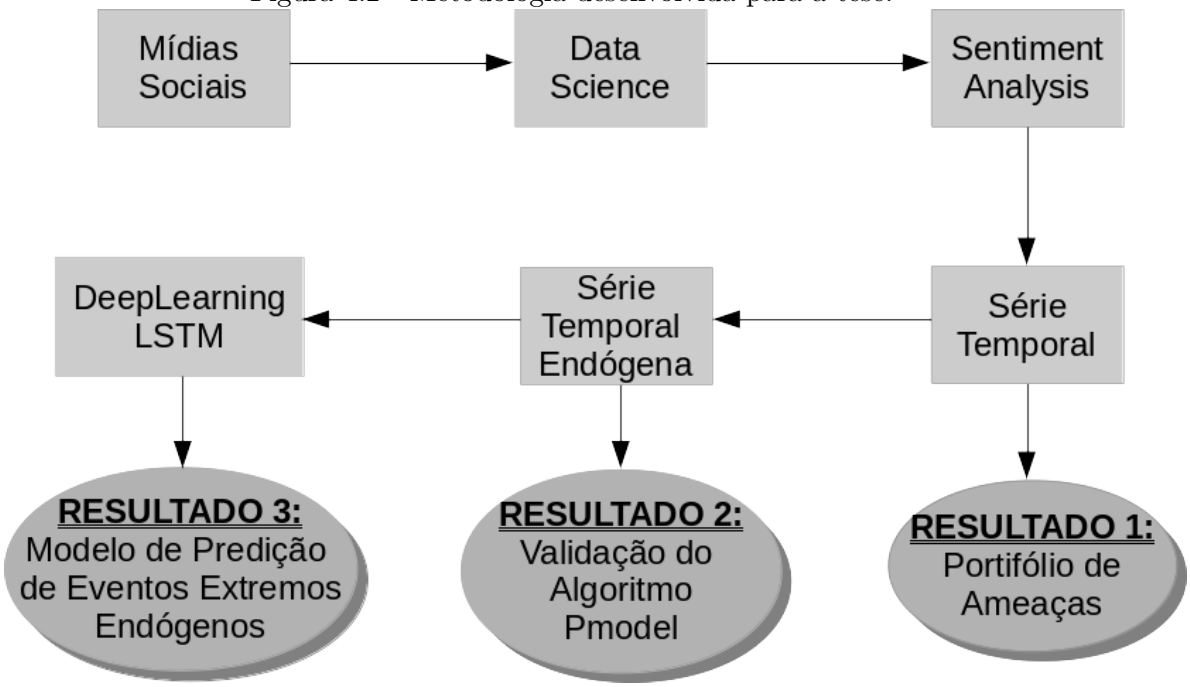
A grande popularização da internet a partir do fim do século XX fez com que uma imensa quantidade de informações fossem divulgadas e compartilhadas pela sociedade na "*World Wide Web*". Em consequência a isso, viu-se uma migração em massa das mídias sociais em papel impresso para a tela dos computadores, "*tablets*" e "*smartphones*". Essa migração da forma de divulgação da informação aconteceu para os dados em formato de textos passados e presentes, gerando uma fonte ilimitada de dados com informações das mais diversas áreas (CASTELLS, 2003). O acesso público e fácil a essa quantidade de conhecimento armazenada, despertou o interesse de variadas áreas de organização da sociedade, que perceberam a importância e o ganho em se analisar e entender de forma rápida e precisa esses dados (CEDERMAN; WEIDMANN, 2017). Um desses domínios de conhecimento que merece a atenção para um estudo mais detalhado para uma melhor identificação das suas causas e prevenção de consequências são os eventos extremos. Os eventos extremos fazem parte da evolução da sociedade e podem designar eventos sociais, políticos e naturais de forma endógena ou exógena (ROSA et al., 2019). Assim, nesta tese considera-se como dados de análise as notícias de mídias sociais, como revista e jornais, relacionadas a eventos extremos sociais endógenos, como os conflitos armados e eventos extremos naturais endógenos como o desmatamento.

Na tese, realizou-se a análise de notícias utilizando os conceitos de análise de sentimento que, com técnicas de séries temporais e aprendizado de máquina, gerou-se um preditor de eventos extremos endógenos. Para a realização desse processo inicialmente coleta-se notícias que apresentem indícios de ameaças referentes aos eventos extremos analisados e que essa ameaça possa levar a uma sequência de eventos menores até se ter como consequência um evento extremo endógeno. A Figura 4.1 ilustra o processo descrito.

Figura 4.1 - Reflexão para a coleta das notícias das mídias sociais.

A metodologia desenvolvida, na tese, para realizar a predição de eventos extremos endógenos, com aplicação em eventos sociais e eventos naturais, pode ser visualizada na Figura 4.2

Figura 4.2 - Metodologia desenvolvida para a tese.



4.1 Mídias sociais

No desenvolvimento da tese, consideraram-se notícias de mídias sociais relacionadas com eventos sociais e naturais extremos como dados de entrada. Os modelos de eventos selecionados para a pesquisa foram conflitos armados e desmatamento, respectivamente, devido à importância das suas consequências para a vida da sociedade.

Para os dados de eventos sociais extremos consideram-se notícias coletadas de mídias sociais de grande circulação mundial, como (REUTERS, 2019), (CNN, 2020), (GUARDIAN, 2020), etc e relacionadas a alguma forma de ameaças ou ofensas verbais trocadas entre chefes de estados de países que possam sugerir o surgimento de um conflito armado. Como estudo de caso, coletaram-se notícias sobre os conflitos da Síria e USA no período de 1 de janeiro de 2016 até 5 de abril de 2017. A data final coincide com o dia que antecede o lançamento dos mísseis "*Tomahawk*", considerados de médio alcance e invisíveis a radares, por dois navios norte-americanos no Mar Mediterrâneo para a base aérea do regime de Bashar Al-Assad (UOL, 2017). Esses dados foram coletados de forma empírica manual por busca na internet e nas mídias sociais citadas considerando a data do período analisado do conflito. Ainda com a intenção de se aprofundar na análise dos conflitos armados, realiza-se continuamente a coleta de notícias das mídias sociais relacionadas a alguma ameaça entre chefes

de estados que possam levar a um conflito armado entre Irã e USA e também um conflito de ordem mundial relacionado com os países, China, Índia, Paquistão, Reino Unido, Japão, USA, Coreia do Norte, Coreia do Sul, Taiwan e Indonésia. A ideia da análise dos conflitos armados, considerando a reflexão apresentada na Figura 4.1, é caracterizada como apresentado na Tabela 4.1.

Tabela 4.1 - Caracterização dos termos ameaça, evento e evento extremo para os eventos extremos sociais.

Termo	**Caracterização**
ameaça	alguma forma de ofensa verbal
evento	ataque ocasionado pela ameaça
evento extremo	conflito armado gerado pela ofensa verbal e pelo ataque

Para ambas as coletas relacionadas aos conflitos armados, consideraram-se referentes ao tema como as apresentadas na Tabela 4.2.

Tabela 4.2 - Amostra de algumas palavras sobre conflitos armados usadas na busca de notícias de forma manual e pelo "*chatbot Newsbot*".

arms	attack	force
threat	weapons	tanks
conflict	armed	war

Para os dados de eventos extremos naturais, consideram-se notícias coletadas de mídias sociais de grande circulação nacional e também de mídias sociais específicas da área ambiental, como (FOLHA, 1996), (GLOBO, 2006), (BUTLER, 1999), (LIMA, 1994), etc. Essas notícias foram coletadas, de forma empírica manual, desde janeiro de 2015 até outubro de 2020 e estavam relacionadas com a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica. Nesse contexto, a ideia, considerando a reflexão apresentada na Figura 4.1, é caracterizada como apresentado na Tabela 4.3.

Tabela 4.3 - Caracterização dos termos ameaça, evento e evento extremo para os eventos extremos naturais.

Termo	**Caracterização**
ameaça	menção a possibilidade de seca
evento	queimada ocasionada pela ameaça
evento extremo	desmatamento gerado pela ameaça e pelo ataque

A coleta de notícias, referentes ao evento extremo natural, considerou palavras relacionadas ao tema, como as apresentadas na Tabela 4.4.

Tabela 4.4 - Amostra de algumas palavras sobre ocorrência de seca, queimadas e desmatamento usadas na busca de notícias de forma manual e pelo "*chatbot Newsbot*".

seca	queimada	baixa
pluviosidade	pouca	chuva
desmatamento	amazonia	queima

As notícias coletadas foram armazenadas em um arquivo "*.csv*" com as informações de data, URL, Grau de Similaridade de Ameaça (GSA). Na Seção 4.2, apresentam-se como os dados coletados nas mídias sociais foram organizados no arquivo "*.csv*".

4.2 Ciência de dados

A aplicação dos conceitos de ciência de dados na tese, iniciou-se com a seleção de dados e a definição do dado alvo ou definição dos temas dos eventos extremos apresentados na Seção 4.1, considerando o processo de KDD apresentado na Figura 3.1. O pré-processamento dos dados, iniciou-se com o armazenamento das notícias coletadas realizado por meio do desenvolvimento de uma estrutura de tabela que possui como colunas Data, URL e o Grau de Similaridade de Ameaça (GSA). Os dados Data e URL foram preenchidos com o processo de coleta nas mídias sociais. O dado Grau de Similaridade de Ameaça foi calculado na etapa de análise de sentimento da metodologia apresentada na Figura 4.2.

Assim, para cada tema analisado, realizou-se a coleta de pelo menos 1 notícias por mês de cada ano do período considerado na análise. Na etapa de transformação dos dados do processo de KDD, as notícias foram organizadas em ordem crescente de data, ou seja, considerando dia, mês e ano, com a notícia mais antiga no topo da tabela. Essa tabela é armazenada em arquivo "*.csv*" e ilustrada pelas Tabelas 4.5 e

4.6.

Tabela 4.5 - Exemplo da organização das notícias sobre eventos extremos sociais armazenadas no arquivo "*.csv*".

Data	URL
06/03/17	cnn.com/trump-travel-ban
07/03/17	cnn.com/save-the-children
09/03/17	cnn.com/jon-huntsman
10/03/17	cnn.com/russia-turk
10/03/17	cnn.com/syria-aleppo

Tabela 4.6 - Exemplo da organização das notícias sobre eventos extremos naturais (ocorrência de seca, queimadas e desmatamentos na Amazônia) armazenadas no arquivo "*.csv*".

Data	URL
02/03/17	https://jornal.usp.br/desmatamentochuvasamazonia/
13/03/17	https://www.dw.com/amazoniaciclodesmatamentoseca/
18/03/17	http://crisalida.eco.br/efeitossecaperdaamazonia/
20/03/17	https://www.pirelli.com/mudancaclimatica/
29/03/17	https://www.ecodebate.com.br/amazoniasubstituidagramineas/

As próximas etapas do KDD, mineração de dados, interpretação ou avaliação, foram desenvolvidas, aplicando-se os conceitos de processamento da linguagem natural e análise de sentimento. A aplicação desses conceitos é detalhada na Seção 4.3.

4.3 Análise de sentimento

Depois da criação do arquivo "*.csv*" com as notícias coletados e organizadas, inicia-se o processo de tratamento e análise dessas notícias. A informação de URL contida no arquivo "*.csv*" é utilizada para realizar a leitura em tempo real de cada notícia. Como essa leitura é realizada via "*Web*" utilizando a biblioteca "*BS4*" da linguagem de programação "*Python 3.7*", tem-se acesso a um arquivo "*.html (Hypertext Markup Language)*", no qual é realizado somente a leitura das informações contidas dentro das "*tags*" de parágrafos $< p >< /p >$. O resultado desse processo é um texto contendo as informações da notícia a ser analisada.

O tratamento das informações, contidas no texto, é realizado utilizando os conceitos do processamento da linguagem natural por meio da utilização da etapa de toke-

nização, descrita na Seção 3.3. Nessa etapa, realiza-se a eliminação de símbolos e caracteres que não tenham representação de sentido para análise, como !, ?, $, &, etc. A tokenização foi aplicada utilizando as APIs "*NLTK* (*Natural Language toolKit*)" (BIRD et al., 2009) na Linguagem "*Python*" e "*Embedded*" do "*Keras Tensorflow*" (CHOLLET, 2015). A NLTK é uma plataforma de código aberto, que permite a criação de programas "*Python*" para se trabalhar com dados de linguagem humana. A plataforma tem uma interface fácil de usar, com a biblioteca de processamento de texto definida para classificação, tokenização, derivação, marcação, raciocínio semântico, análise e bibliotecas relacionadas ao processamento da linguagem natural (BIRD et al., 2009). A "*Keras*" é uma API de código aberto, que funciona com redes neurais de alto nível, escrita em "*Python*" e permite desenvolver aplicativos em conjunto com a "*TensorFlow API*" (uma plataforma de aprendizado de máquina de código aberto de ponta a ponta) (GOOGLE, 2015). Seu objetivo é permitir uma experimentação rápida, ou seja, focar no resultado no menor tempo possível (CHOLLET, 2015).

Depois da tokenização das informações, inicia-se a fase de análise das notícias aplicando-se os conceitos de análise de sentimento por meio da biblioteca "*SpaCy* (*Industrial-Strength Natural Language*)". A "*SpaCy*" é uma biblioteca de código aberto gratuita para processamento da linguagem natural avançada em "*Python*". A "*SpaCy*" foi projetada especificamente para uso em produção e ajuda a criar aplicativos que processam e entendem grandes volumes de texto. Ela pode ser usada para construir sistemas de extração de informações ou de compreensão de linguagem natural, ou para pré-processar texto para aprendizado profundo. Na tese utiliza-se a análise de sentimento com base em tópicos ou recursos para a extração de informação das notícias selecionadas. Nesse processo de extração de informação, inicialmente, define-se um texto base, no qual se define o conceito da emoção da ameaça que gostaria de se identificar nos textos analisados. Assim, considera-se que essa definição representa 100% essa emoção de ameaças. Esse texto base é definido de forma empírica, considerando o conhecimento das pessoas da tese sobre o assunto abordado. A seguir são apresentados trechos dos textos bases para o conflito armado ente a Síria e USA na língua inglesa e sobre secas, queimadas e desmatamento na língua portuguesa. Os textos bases completos, utilizados na fase de experimentação da tese, são apresentados nas Seções B.0.1 e B.0.2.

Trecho do texto base para o tema conflitos armados entre a Síria e USA

`The jihadist organisation Jabhat al-Nusra (aka the al-Nusra Front),`

involved in Síria civil war since 2012, announced in a video that it is formally detaching itself from al-Qaeda. The move, which had been negotiated for well over a year, was confirmed on July 28. In the video, Nusra leader Abu Mohammad al-Joulani said the detachment was in the necessity of the continuity of the jihad of al-Sham [Síria]. Renamed as Jabhat Fatah al-Sham (the Front for the Conquest of Síria), Nusra will try to bring other rebel factions into a unified body, which al-Joulani says will represent the masses of the people, liberating their lands, giving victory to their faith, and upholding their testimony of faith. Some US-based observers quickly reacted with suspicion and derision. Thomas Joscelyn in the Long War Journal dismissed the statement as propaganda with no substance: Even if Joulani did say that his group had really split from al-Qaeda (which he didnt), there is no good reason to believe him.White House spokesman Josh Earnest commented: There continues to be increasing concern about Nusra Fronts growing capacity for external operations that could threaten both the United States and Europe. Earnests State Department counterpart John Kirby was more dismissive: They are still considered a foreign terrorist organisation. We judge a group by what they do, not by what they call themselves. These reactions were all too predictable. They were also misguided, and the thinking behind them is damaging.

Trecho do texto base para o evento extremo natural - ocorrência de secas, queimadas e desmatamento

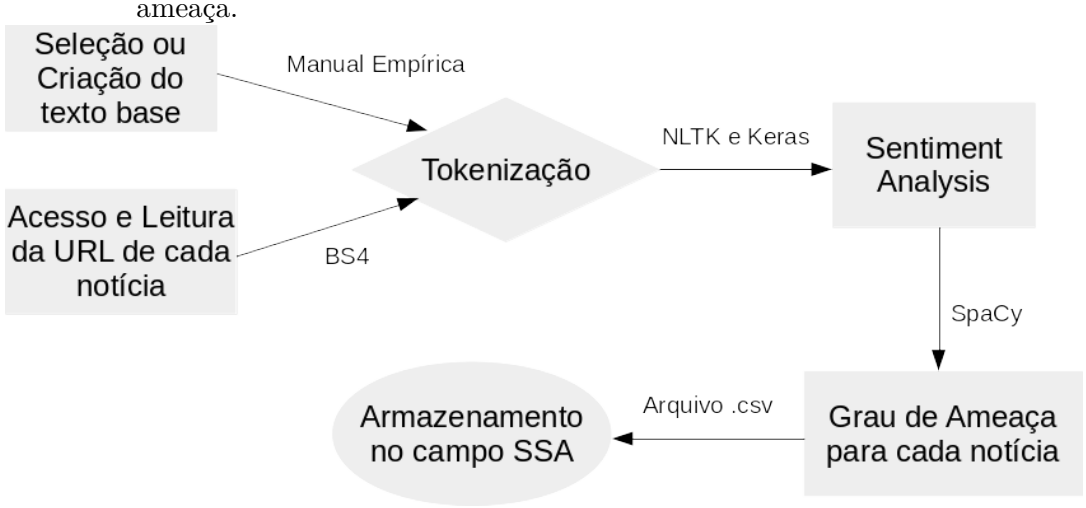
---

A temporada de incêndios de 2020 na floresta amazônica pode ser muito mais rigorosa do que em 2019 e um dos motivos dessa piora são as mesmas condições climáticas que intensificam a temporada de furacões no hemisfério norte segundo pesquisadores Em agosto passado uma série de grandes incêndios provocados pelo homem na Amazônia lançou nuvens de fumaça sobre a cidade de São Paulo transformando o dia em noite e gerou protestos internacionais Mas embora esses incêndios tenham sido incomuns e alarmantes a situação poderia ter sido ainda pior se a Amazônia estivesse em época de estiagem Mas neste ano

```
infelizmente  condições mais secas do que a média são exatamente o
que está previsto para o sul da Amazônia e um dos motivos é o
aumento extraordinário de calor no Atlântico Tropical Norte  a
milhares de quilômetros de distância  Esse calor oceânico também
fez com que a temporada de furacões no Atlântico batesse recordes
logo de início um prenúncio das previsões para  uma temporada
extraordinariamente tumultuada  Algumas pesquisas sugerem existir
uma relação causal entre os próprios furacões e os piores anos de
incêndios na Amazônia embora isso seja assunto de maior debate.
```

Depois do texto base gerado, identifica-se a porcentagem de similaridade da nova notícia processando-a com o texto base por meio da biblioteca "*SpaCy*". O resultado desse processo é o grau de similaridade de ameaça (GSA) que cada notícia representa em relação ao evento extremo analisado. Esse grau de ameaça é calculado e armazenado para cada notícia acessada e armazenado no campo GSA do arquivo "*.csv*". O valor de GSA é apresentado com 14 casas decimais para uma melhor precisão do resultado. O processo descrito é ilustrado na Figura 4.3.

Figura 4.3 - Aplicação da análise de sentimento para o cálculo do grau de similaridade de ameaça.



A seguir, apresenta-se um pseudocódigo do processo de cálculo do GSA, que mostra aplicação das bibliotecas mencionadas no desenvolvimento da metologia da tese.

```
Pseudocodigo Cálculo_GSA():
    /*Leitura dos dados de URL do arquivo .csv file*/
    data_url = read_csv_file()

    /*Loop para a leitura de cada notícia com URL no arquivo .csv*/
    Para i até tamanho(data_url):
       /*Acesso a URL da notícia*/
       open_html_file = request.get(data_url[i])

       /*Leitura do conteúdo do arquivo .html com a API BS4*/
       html_file = BS4(open_html_file.content)

       /*Seleção dos dados entre a tag <p><p/>*/
       text = found_all('<p><p/>')

       /*Eliminação dos caracteres sem significado utilizando
       o processo de tokenização*/
       data = NLTK_tokenização(text)

       /*Leitura do texto base de notícias */
       news_base = read_news_base()

       /*Eliminação dos carcteres sem significado utilizando
       o processo de tokenização*/
       news_base_data = NLTK_tokenização(news_base)

       /*Transforma o texto em um vetor de palavras */
       nbd_nlp = nlp(news_base_data)
       data_nlp = nlp(data)

       /*Compara os vetores de palavras para o cálculo da
       similaridade entre os textos*/
       GSA = nbd_nlp.similarity(data_nlp)

       /*Armazena o valor de GSA no arquivo .csv*/
       write_csv_file(GSA)
```

```
    Fim Para
Fim Pseudocodigo
```

As Tabelas 4.7 e 4.8, a seguir, apresentam exemplos das notícias coletadas com o valor do grau de similaridade de ameaça (GSA) para cada notícia.

Tabela 4.7 - Notícias sobre o conflito entre a Síria e USA com valor GSA, que representa o grau de ameaça.

Data	URL	GSA
06/03/17	cnn.com/trump-travel-ban	0,92
07/03/17	cnn.com/save-the-children	0,73
09/03/17	cnn.com/jon-huntsman	0,83
10/03/17	cnn.com/russia-turk	0,78
10/03/17	cnn.com/syria-aleppo	0,87

Tabela 4.8 - Notícias sobre a ocorrência de seca, queimadas e desmatamentos na Amazônia com valor GSA que representa o grau de ameaça.

Data	URL	GSA
02/03/17	jornal.usp.br/desmatamentochuvasamazonia/	0,91
13/03/17	dw.com/amazoniaciclodesmatamentoseca/	0,85
18/03/17	crisalida.eco.br/efeitossecaperdaamazonia/	0,89
20/03/17	pirelli.com/mudancaclimatica/	0,89
29/03/17	ecodebate.com.br/amazoniasubstituidagramineas/	0,91

4.4 Série temporal

O grau de similaridade de ameaça, descrito na Seção 4.3, é calculado e armazenado para cada notícia coletada durante o período de análise. Esses valores de GSA, organizados em ordem crescente de data, formam a série temporal utilizada na tese. A Tabela 4.9 apresenta um trecho da série temporal de ameaças, gerada na tese, para o conflito armado entre a Síria e USA e a Tabela 4.10 apresenta um trecho da série temporal de ameaças gerada na tese, com os dados de ocorrência de seca, queimadas e desmatamentos para o ano de 2017.

Tabela 4.9 - Exemplo da série temporal de ameaças para o conflito entre a Síria e USA.

Data	**GSA**
06/03/17	0,92
07/03/17	0,73
09/03/17	0,83
10/03/17	0,78
10/03/17	0,87

Tabela 4.10 - Exemplo da série temporal de ameaças dos dados de ocorrência de seca, queimadas e desmatamentos no ano de 2017.

Data	**GSA**
02/03/17	0,91
13/03/17	0,85
18/03/17	0,89
20/03/17	0,89
29/03/17	0,91

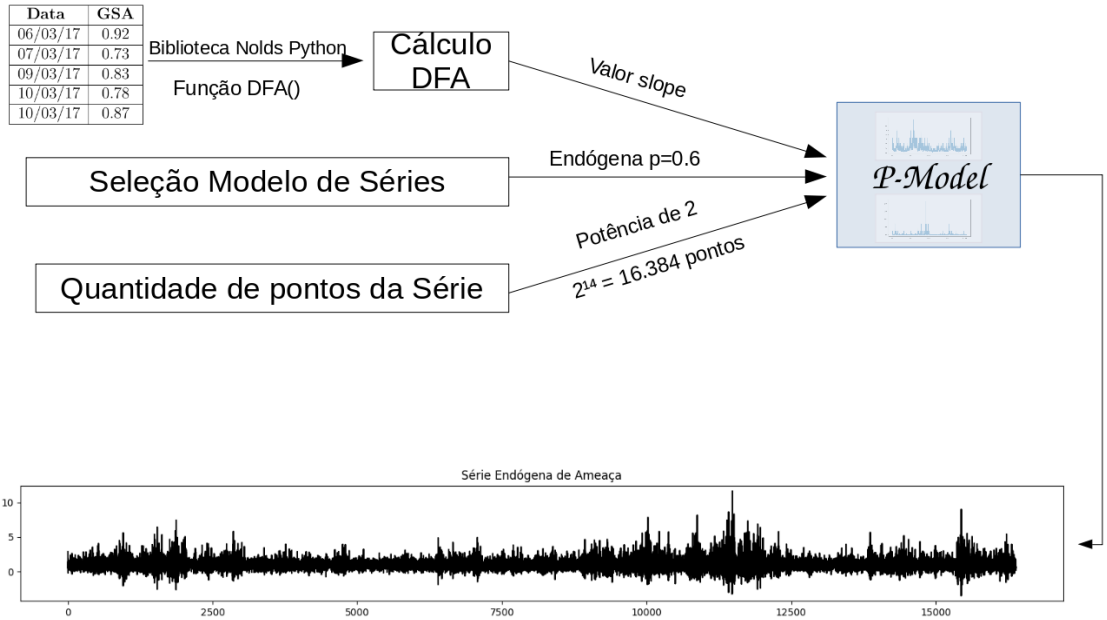
O processo de cálculo da série temporal gerou como resultado, além da série temporal, um portfólio de ameaças que contém as informações de data, URL, grau de ameaça para cada modelo de evento analisado. A intenção do portfólio é gerar uma fonte de informação pública para a análise dos eventos extremos estudados e uma iniciativa para a construção de portfólios de ameaças para outros temas de interesse da sociedade. Os portfólios de ameaças de eventos extremos sociais e naturais são apresentados na Seção 5.1.

4.5 Série temporal endógena

De acordo com Sornette (2006), os eventos extremos sociais e naturais são considerados como sistemas endógenos devido suas características de auto-organização. Dessa forma, utiliza-se na tese o algoritmo "*P-Model*", para gerar as séries temporais com características endógenas. Essas séries endógenas são geradas a partir da série temporal de ameaças de eventos extremos sociais e naturais, sendo obtida pela utilização da técnica de análise de sentimento. Como apresentado na Seção 3.6.1, o "*P-Model*" permite gerar séries temporais tanto endógenas quanto exógenas e utiliza três parâmetros de entrada: a quantidade de elementos da série temporal, p que representa a categoria de série conforme o valor passado e o valor de β ("*slope*"), que na tese é obtido pelo cálculo do DFA da série temporal de ameaça gerada na Seção 4.4. Para o cálculo do DFA, utilizou-se a biblioteca "*nolds*" da "*Python*" 3.7 e para

a aplicação do "*P-Model*" empregou-se o algoritmo também na linguagem "*Python*". A seguir apresentam-se, na Figura 4.4 uma esquematização desse processo e posteriormente o pseudocódigo que demonstra como foram aplicadas as bibliotecas e o algoritmo "*P-Model*", no desenvolvimento da metodologia.

Figura 4.4 - Esquematização do processo para gerar a série endógena de ameaças utilizando o algoritmo "*P-Model*".




```
Pseudocodigo Serie_Endogena():
    /*Declaração variáveis de entrada do pmodel*/
    var slope: decimal
    var p: decimal
    var dados_serie: string
    var dados_pmodel: string
    var tamanho_serie: inteiro

    /*Lendo dados da série temporal de ameaça do
    arquivo .csv*/
    dados_serie = decimal(read_csv_file())

    /*Calculando o valor do slope por meio do DFA
    da série temporal de ameaça*/
```

```
    slope = dfa(dados_serie)

    /*Definição do parâmetro p para série temporal
    endógena*/
    p = 0,60

    /*Gerando a série temporal endógena utilizando
    o algoritmo P-Model*/
    dados_pmodel = PModel(tamanho_serie, p, slope)

    /*Armazenando a série endógena em um arquivo
    .txt*/
    write_file(dados_pmodel)
Fim Pseudocodigo
```

A utilização do "*P-Model*" gerou o segundo resultado da tese como apresentado na metodologia da Figura 4.2. Esse resultado refere-se a validação do algoritmo "*P-Model*" na criação de séries temporais endógenas para os eventos extremos. Esse resultado é apresentado na Seção 5.2.

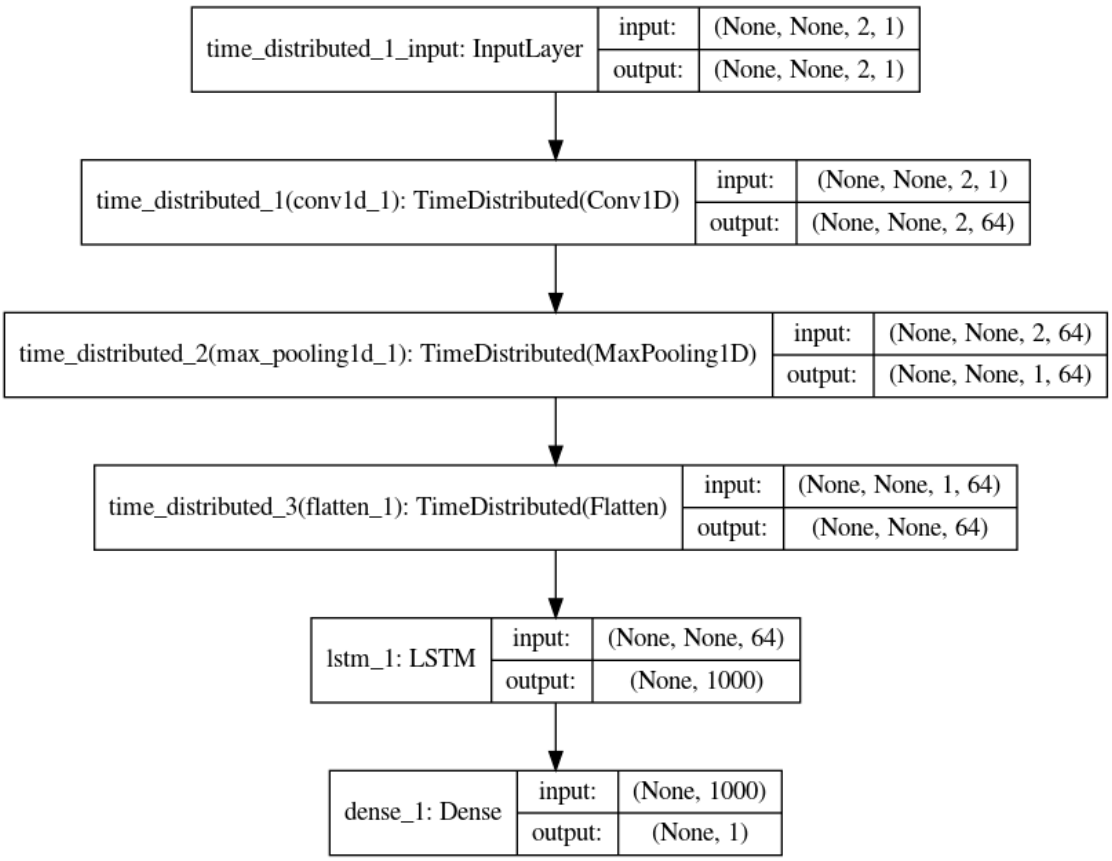
4.6 Arquitetura de apresendizado profundo com rede neural LSTM

A técnica de aprendizado profundo, caracteriza-se por ter bom desempenho na extração de características de grandes volumes de dados (EMMERT-STREIB et al., 2018). Assim, as séries temporais endógenas geradas pelo algoritmo "*P-Model*" possuem 16384 pontos e são utilizadas para dar origem ao modelo de predição dos eventos extremos. Para a elaboração desse modelo, utiliza-se a API de aprendizado profundo "*Keras*" (CHOLLET, 2015) na linguagem "*Python 3.7*". Na aplicação da técnica de aprendizado profundo, utiliza-se uma estrutura de rede neural que possibilita uma boa eficiência na extração e aprendizagem automática de dados de sequência unidimensional (BROWNLEE, 2017) (BARINO; SANTOS, 2021). A rede LSTM, utilizada nessa estrutura, é uma arquitetura de rede neural artificial recorrente (RNN), que também pode ser usada no campo da aprendizagem profunda e diferentemente das redes neurais padrões, possui conexões de "*feedback*" (BROWNLEE, 2017). A seguir é apresentado o pseudocódigo da estrutura da rede, que obteve o melhor resultado na fase de experimentação da tese.

```
model = Sequential()
model.add(TimeDistributed(Conv1D(filters=64, kernel_size=1,
activation='relu'), input_shape=(None, n_steps, n_features)))
model.add(TimeDistributed(MaxPooling1D(pool_size=2)))
model.add(TimeDistributed(Flatten()))
model.add(LSTM(1000, activation='relu'))
model.add(Dense(1))
model.compile(optimizer=Adam(learning_rate=0.001, beta_1=0.9,
beta_2=0.999, epsilon=1e-03, amsgrad=False), loss='mse',
metrics=[RootMeanSquaredError()])
```

A representação de cada camada da arquitetura de rede neural construída com a API de aprendizado profundo "*Keras*", é apresentada na Figura 4.5.

Figura 4.5 - Representação das camadas da arquitetura de aprendizado profundo e rede neural utilizada na tese.



Os gráficos de perda e RMSE foram gerados para os dados de treinamento e validação para a comprovação da generalização do modelo de aprendizado de máquina, apresentado na Figura 4.5. Na Figura 4.6, apresenta-se o gráfico de perda e na Figura 4.7, apresenta-se o gráfico de RMSE para os dados de treinamento e de validação, da predição realizada para a série temporal endógena de ameaças do conflito entre a Síria e USA, considerado o principal caso de estudo da tese.

Figura 4.6 - Variação da perda para os dados de treinamento e validação da predição do conflito entre a Síria e USA.

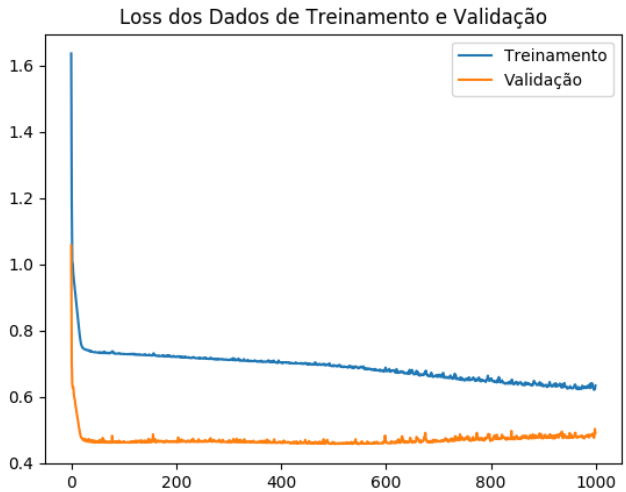
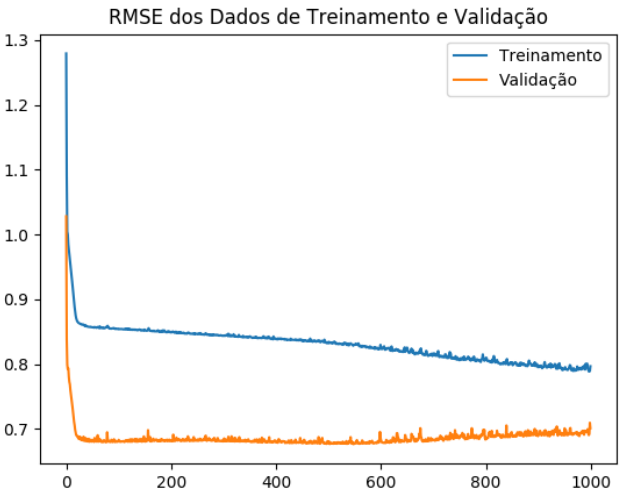


Figura 4.7 - Variação do RMSE para os dados de treinamento e validação da predição do conflito entre a Síria e USA.



Observa-se das informações apresentadas nos gráficos de perda e RMSE, que os valores de treinamento e validação estão se aproximando, com pequenas variações nos valores, ao longo que as épocas continuam. Verifica-se que, entre as épocas 70 e 500, essa aproximação é um pouco mais intensa do que para as épocas a partir de 500. Isso sugere que o modelo está se ajustando bem e consequentemente, terá uma boa resposta para a predição.

Essa arquitetura utilizando a rede LSTM foi aplicada para realizar a predição dos eventos extremos sociais e naturais analisados com as técnicas de análise de sentimento e "*P-Model*". Com essa predição, obtém-se o terceiro resultado da tese, como apresentado na metodologia da Figura 4.2. Os detalhes desse terceiro resultado são apresentados na Seção 5.3.

5 ANÁLISE E RESULTADOS

A área de análise de sentimento está em grande ascensão nos últimos anos (MÄNTYLÄ et al., 2018) devido a grande popularização das mídias e redes sociais. Apesar do seu grande crescimento, ainda se considera uma área com grande potencial de estudo e aplicação nas mais diversas áreas da sociedade (HOLLANDER et al., 2016). Dessa forma, uma das áreas que ainda podem se beneficiar das análises da técnica de análise de sentimento é a de eventos extremos naturais e sociais (DAVIS, 2019)(DAHAL et al., 2019). Neste capítulo, são apresentados os resultados alcançados com a aplicação da metodologia desenvolvida na tese e apresentada na Seção 4. Durante o desenvolvimento da tese, obtiveram-se três importantes resultados: o portfólio de ameaças, que utiliza as técnicas de análise de sentimento, a validação do "*P-Model*", que utiliza as informações do portfólio de ameaças e a predição dos eventos extremos naturais e sociais abordados. Nas Seções 5.1, 5.2 e 5.3, detalham-se esses resultados alcançados na tese.

5.1 Resultado 1: portifólio de ameaças

O portfólio de ameaças foi elaborado com a iniciativa inovadora de gerar uma fonte de informação pública sobre os eventos extremos coletadas das mídias sociais. Como a informação da mídia social é seu endereço na internet, os dados das notícias não sofrem nenhuma intervenção humana e permanecem no seu estado original. Somando-se a informação do URL, tem-se a data e o grau de ameaça de cada notícia. Essa forma de organização difere das fontes de dados relacionados ao tema ((UPPSALA CONFLICT DATA PROGRAM - UCDP, 1980) (RALEIGH, 2010) (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE, 2019)), que apresentam as informações após serem processadas. Para cada modelo de evento extremo estudado se elaborou um portfólio de ameaças, com as informações padronizadas, como são apresentados nas Seções 5.1.1 e 5.1.2.

5.1.1 Eventos extremos sociais - conflitos armados

Assim, para o evento extremo social com abordagem em conflitos armados se obteve três portfólios: conflito entre a Síria e USA, conflito do Irã e USA e um possível conflito global apresentados nas Seções 5.1.1.1, 5.1.1.2 e 5.1.1.3, respectivamente.

5.1.1.1 Conflito armado entre a Síria e USA

O portfólio do conflito entre a Síria e USA possui 47 notícias, com as datas no período de janeiro de 2016 e abril de 2017 como já mencionado, que contenham em seu texto

alguma ação e/ou coação referentes à soberania de um dos dois países. A Figura 5.1 apresenta o portfólio com informações de Data, URL e Grau de Similaridade com Ameaça.

Figura 5.1 - Visão geral do portfólio de ameaças para o conflito armado entre a Síria e USA no período de janeiro de 2016 e abril de 2017.

Data	URL	Grau de Similaridade com Ameaça (GSA)
12/01/2016	https://time.com/4180526/what-obama-gets-wrong-about-conflict-in-the-middle-east/	0,941640414401327
13/01/2016	https://gpinvestigations.pri.org/the-us-is-killing-more-civilians-in-iraq-and-syria-than-it-acknowledges-9d	0,804636590055959
25/01/2016	https://www.osw.waw.pl/en/publikacje/osw-commentary/2016-01-25/germanys-engagement-resolution-sy	0,909058032781724
04/02/2016	https://cy.usembassy.gov/initiatives-united-states-toward-conflict-resolution-syria-eastern-mediterranean-l	0,91735779730814
18/02/2016	https://qz.com/625389/us-backed-rebel-groups-are-ready-to-turn-on-each-other-in-syria/	0,895609868593785
22/02/2016	https://sy.usembassy.gov/joint-statement-united-states-russian-federation-co-chairs-issg-cessation-hostili	0,912941970243725
22/02/2016	https://www.msf.org/syria-%E2%80%9Cour-bombs-are-smarter-yours%E2%80%9D	0,934013655159411
29/02/2016	https://qz.com/625389/us-backed-rebel-groups-are-ready-to-turn-on-each-other-in-syria/	0,895609868593785
15/03/2016	https://webdoc.france24.com/syria-conflict-fighters-alliances-kurds-russia-turkey/	0,950106316951287
15/03/2016	https://www.abc.net.au/news/2016-03-15/syrian-conflict-six-years-on/7231680	0,910742565794612
17/03/2016	https://www.brookings.edu/blog/markaz/2016/03/17/why-the-united-states-hasnt-intervened-in-syria/	0,952455768177663
08/04/2016	https://www.ejiltalk.org/the-united-states-is-at-war-with-syria-according-to-the-icrcs-new-geneva-convent	0,936558939916675
24/04/2016	https://www.reuters.com/article/us-mideast-crisis-usa-syria-idUSKCN0XL0ZE	0,895917656226195
04/05/2016	https://mepc.org/commentary/who-blame-syria	0,946603316792111
03/06/2016	https://www.washingtonpost.com/graphics/national/obama-legacy/intervention-libya-and-syrian-crisis.html	0,952931143190229
08/06/2016	https://www.bbc.com/news/world-middle-east-36479415?SThisFB	0,860796892703854
16/06/2016	https://www.nytimes.com/2016/06/17/world/middleeast/syria-assad-obama-airstrikes-diplomats-memo.htm	0,933615947000997
19/06/2016	https://tcf.org/content/report/the-case-for-a-more-robust-intervention-in-syria/?agreed=1	0,955276440320888
13/07/2016	https://www.reuters.com/article/us-usa-russia-idUSKCN0ZT2QZ	0,878055360876125
21/07/2016	https://www.bbc.com/news/world-middle-east-36843990	0,857245684375056
09/08/2016	https://theconversation.com/american-obsession-with-al-qaeda-in-syria-is-making-the-conflict-there-wors	0,959503381241242
15/08/2016	https://www.csis.org/analysis/us-wars-iraq-syria-libya-and-yemen-what-are-endstates	0,938673037176274
30/08/2016	https://www.sandersinstitute.com/blog/americas-true-role-in-syria	0,880273166567534
25/08/2016	https://www.businessinsider.com/us-f-22-pilots-conflict-with-syrian-jets-2016-8	0,840487879743308
12/09/2016	https://www.usatoday.com/story/news/world/2016/09/12/americans-syria-cease-fire-civil-war/90259128/	0,902215841096442
18/09/2016	https://www.bbc.com/news/world-middle-east-37398721	0,901791357404282
21/09/2016	https://www.bbc.com/news/world-middle-east-37424871	0,878388359787911
25/09/2016	https://www.reuters.com/article/us-mideast-crisis-syria-options-idUSKCN11Y37J	0,900803151979752
07/10/2016	https://www.euractiv.com/section/global-europe/interview/middle-east-expert-syrian-war-is-only-going-to-g	0,869353843401692
16/10/2016	https://www.reuters.com/article/us-mideast-crisis-syria-obama-idUSKBN1452KX	0,888539850707904
25/10/2016	https://www.theguardian.com/world/2016/oct/25/hillary-clinton-syria-no-fly-zones-russia-us-war	0,922573046441244
26/10/2016	https://www.voanews.com/usa/us-politics/trump-clintons-syria-policy-would-bring-world-war-3	0,932207317009864
16/11/2016	https://www.dw.com/en/donald-trumps-vision-for-syria/a-36412242	0,939145174720283
13/12/2016	https://time.com/4611414/donald-trump-middle-east-policy/	0,947719027093682
21/12/2016	https://time.com/4611414/donald-trump-middle-east-policy/	0,947719027093682
28/12/2016	https://www.aljazeera.com/news/2016/12/28/russia-turkey-us-supporting-syria-terrorist-groups	0,938764433765508
29/12/2016	https://foreignpolicy.com/2016/12/29/obama-never-understood-how-history-works/	0,946561674313868
18/01/2017	https://www.npr.org/sections/parallels/2017/01/18/510447582/after-8-years-of-unbroken-war-obama-hano	0,940619275075397
25/01/2017	https://blog.cei.iscte-iul.pt/obamas-military-legacy-iii-the-wars-in-libya-and-syria/	0,956159081557196
21/02/2017	https://www.forbes.com/sites/dougbandow/2017/02/21/u-s-troops-dont-belong-in-syria-america-should-st	0,939912987021703
12/03/2017	https://edition.cnn.com/2017/03/11/middleeast/syria-conflict-isis-us-military/index.html	0,746103039365248
12/03/2017	https://time.com/4699178/us-troop-increase-presence-syria-raqqa-isis/	0,947545479427441
13/03/2017	https://www.csmonitor.com/USA/Foreign-Policy/2017/0313/Marines-sent-to-Syria.-Can-US-withstand-pull	0,952945978006924
14/03/2017	https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(17)30758-4/fulltext	0,750967102048839
04/04/2017	https://www.npr.org/2017/04/04/522627064/on-the-record-president-trump-vs-private-citizen-trump-on-syr	0,930656985157977
05/04/2017	https://www.npr.org/2017/04/04/522627064/on-the-record-president-trump-vs-private-citizen-trump-on-syr	0,930656985157977
05/04/2017	https://www.vox.com/world/2017/4/5/15189820/trump-response-atrocity-syria-blame-obama	0,935767353592602

5.1.1.2 Conflito armado entre Irã e USA

O portfólio do conflito entre Irã e USA possui 58 notícias, com as datas no período de janeiro de 2019 e de janeiro de 2020, que contenham em seu texto alguma ação e/ou coação referentes à soberania de um dos dois países. A Figura 5.2 apresenta o portfólio com informações de Data, URL e Grau de Similaridade com Ameaça.

Figura 5.2 - Visão geral do portfólio de ameaças para o conflito armado entre Irã e USA no período de janeiro de 2019 e janeiro de 2020.

Data	URL	Grau de Similaridade com Ameaça (GSA)
15/01/2019	https://smallwarsjournal.com/jrnl/art/americas-strategic-options-middle-east	0,929298929393456
21/01/2019	https://www.jpost.com/arab-israeli-conflict/defense-site-iran-has-a-stronger-mil	0,715547280045349
29/01/2019	https://iranprimer.usip.org/blog/2019/jan/29/us-intelligence-community-iran	0,961092571110417
04/02/2019	https://edition.cnn.com/interactive/2019/02/middleeast/yemen-lost-us-arms/	0,959931389739587
04/02/2019	https://www.aljazeera.com/news/2019/2/4/us-needs-military-base-in-iraq-to-wa	0,942323327945551
05/02/2019	https://www.trtworld.com/opinion/iraq-is-more-than-just-a-watchtower-for-the-u	0,970754267924304
17/02/2019	https://www.express.co.uk/news/world/1088405/iran-news-world-war-3-iran-isr	0,918965907716134
26/02/2019	https://www.atlanticcouncil.org/blogs/iransource/is-saudi-arabia-pulling-pakistan	0,891630266914655
07/03/2019	https://www.usatoday.com/story/opinion/2019/03/07/john-bolton-neoconservati	0,927438396166111
11/03/2019	https://www.csis.org/war-by-proxy	0,941698525543711
11/03/2019	https://www.dw.com/en/irans-president-visits-iraq-to-strengthen-ties-amid-us-	0,899528929842172
28/03/2019	https://www.brookings.edu/research/constraining-irans-missile-capabilities/	0,955170541222086
04/04/2019	https://www.militarytimes.com/news/your-military/2019/04/04/iran-killed-more-u	0,892115120297488
14/04/2019	https://abcnews.go.com/Politics/concerns-trump-administration-laying-legal-gro	0,930079105632493
25/04/2019	https://www.ynetnews.com/articles/0.7340.L-5499593.00.html	0,695579068718755
30/04/2019	https://www.aljazeera.com/opinions/2019/4/30/arrogance-fanaticism-and-the-pr	0,966665165700468
14/05/2019	https://www.washingtonpost.com/world/2019/05/14/conflict-with-iran-would-not	0,933431543223389
17/05/2019	https://www.bbc.com/news/world-middle-east-48298517	0,908663400489168
17/05/2019	https://theintercept.com/2019/05/17/us-media-journalists-iran-coverage/	0,962910910714543
19/05/2019	https://www.reuters.com/article/us-usa-iran-guards/iran-is-not-pursuing-war-	0,758425268008993
21/05/2019	https://foreignpolicy.com/2019/05/21/how-to-prevent-an-accidental-war-with-ira	0,985240485373058
22/05/2019	https://www.npr.org/2019/05/22/725428447/what-to-know-about-the-sudden-ta	0,968519763537639
02/06/2019	https://www.reuters.com/article/us-usa-iran-switzerland/u-s-prepared-to-talk-	0,889041721960372
21/06/2019	https://www.reuters.com/article/us-usa-iran/america-is-trying-to-create-iran	0,771998266291761
25/06/2019	https://www.reuters.com/article/us-mideast-iran-rouhani/iran-is-not-looking-f	0,787029213681551
27/06/2019	https://www.reuters.com/article/us-mideast-iran-usa-zarif/irans-zarif-says-sho	0,812463464797735
15/07/2019	https://www.reuters.com/article/us-mideast-iran-nuclear/iran-threatens-to-ret	0,795822938121938
23/07/2019	https://www.reuters.com/article/mideast-iran-navy/iran-observes-all-u-s-ships	0,831876756376396
24/07/2019	https://www.reuters.com/article/us-mideast-iran-president/iran-is-ready-to-neg	0,838044498956127
29/07/2019	https://www.reuters.com/article/us-mideast-iran-jahangiri/irans-foreign-polic	0,833893553906019
01/08/2019	https://www.reuters.com/article/us-mideast-iran-tanker-raab/britain-tells-ira	0,847133532561695
06/08/2019	https://www.reuters.com/article/mideast-iran-rouhani/war-with-iran-is-the-moth	0,756517726561894
24/08/2019	https://www.reuters.com/article/us-mideast-iran-missiles/iran-test-fires-new	0,846200788970415
24/08/2019	https://www.reuters.com/article/mideast-iran-missiles/iran-has-test-fired-a-n	0,80809500162748
17/09/2019	https://www.reuters.com/article/us-iran-usa-khamenei/irans-khamenei-rejects-	0,85561994013579
20/09/2019	https://www.reuters.com/article/us-saudi-aramco-iran-zarif/irans-zarif-says-sa	0,828266895469164
21/09/2019	https://www.reuters.com/article/us-saudi-aramco-iran/iran-says-it-will-dest	0,874853200702361
22/09/2019	https://www.reuters.com/article/us-saudi-aramco-iran-navy/iran-navy-comman	0,808570309938939
30/09/2019	https://www.reuters.com/article/us-saudi-aramco-iran-messages/saudi-arabi	0,831190738941602
03/10/2019	https://edition.cnn.com/2019/10/03/middleeast/iran-war-games-kiley-intl/index.h	0,800139414199633
11/10/2019	https://www.reuters.com/article/us-mideast-iran-tanker-data/iran-tanker-report	0,772043068725765
11/10/2019	https://www.reuters.com/article/mideast-iran-tanker-data/iran-tanker-reported-	0,771077758618327
12/10/2019	https://www.bostonglobe.com/opinion/editorials/2019/10/12/the-shouldn-take-m	0,960352149522541
25/10/2019	https://www.aljazeera.com/news/2019/10/25/dangerous-game-us-europe-and-	0,964912177720271
07/11/2019	https://www.theguardian.com/world/2019/nov/07/iran-has-military-advantage-ov	0,957906063869958
11/11/2019	https://www.timesofisrael.com/hezbollah-chief-chance-of-us-iran-war-down-by	0,937462302884653
18/11/2019	https://www.timesofisrael.com/hezbollah-chief-chance-of-us-iran-war-down-by	0,937462302884653
21/11/2019	https://www.nbcnews.com/news/world/counter-iran-military-may-be-forced-op	0,974815700966214
23/11/2019	https://foreignpolicy.com/2019/11/23/very-possible-iran-will-attack-again-top-us	0,958417476889442
05/12/2019	https://www.militarytimes.com/flashpoints/2019/12/05/how-force-buildups-on-bo	0,92117458019294
12/12/2019	https://www.npr.org/2019/12/12/787377987/u-s-military-official-warns-of-dange	0,976359266886652
30/12/2020	https://www.reuters.com/article/us-iraq-security-usa-idUSKBN1YY0II	0,825746167303319
31/12/2019	https://www.bloomberg.com/news/articles/2019-12-31/tracking-the-escalating-c	0,799298709954905
05/01/2019	https://qz.com/1779727/attack-on-irans-soleimani-fits-a-long-history-of-us-milit	0,941168922184353
05/01/2020	https://www.reuters.com/article/us-iraq-security-iran-usa/u-s-lacks-courage-fo	0,781198346158911
06/01/2020	https://edition.cnn.com/2020/01/05/opinions/us-iran-history-ware-intl/index.html	0,789359923407413
06/01/2020	https://www.marketwatch.com/story/what-stock-market-investors-need-to-kno	0,948445074485563
07/01/2020	https://www.reuters.com/article/us-iraq-security/trump-softens-rhetoric-after	0,852567887816475

5.1.1.3 Conflito armado global

O portfólio do conflito armado global possui 46 notícias, com as datas no período de julho de 2020 e dezembro de 2020. O conflito armado Global está relacionado com um grupo de eventos que estão acontecendo no Mar da China e no Mar do Japão. Esses eventos estão relacionados com o processo de fortalecimento militar da China e a ameça que isto representa para Taiwan, Coreia do Sul, Japão, Filipinas, Malásia e Indonésia. Além do envolvimento da Índia, EUA e o Reino Unido. É preciso notar que a rota comercial marítima mais cobiçada do mundo passa exatamente por estes dois Mares. Com o exercício militar chinês todos os países da região se sentem

ameaçados. A Figura 5.3 apresenta o portfólio com informações de Data, URL e Grau de Similaridade com Ameaça.

Figura 5.3 - Visão geral do portfólio de ameaças para o conflito armado global no período de julho de 2020 e dezembro de 2020.

Data	URL	Grau de Similaridade com Ameaça (GSA)
07/07/2020	https://www.newframe.com/india-and-china-on-the-brink-of-war/	0,926796337566917
10/07/2020	https://mg.co.za/world/2020-07-10-india-and-china-border-conflict-intensifies/	0,924460704266181
10/07/2020	https://www.lowyinstitute.org/the-interpreter/india-china-conflict-move-himalayas-high-seas	0,930904660384366
16/07/2020	https://www.ispionline.it/en/pubblicazione/geopolitics-china-india-conflict-26957	0,899322647056046
23/07/2020	https://thediplomat.com/2020/07/this-time-the-us-is-taking-indias-side-against-china/	0,90621541458538
30/07/2020	https://www.bbc.com/news/world-asia-53171124	0,953267315838351
30/07/2020	https://www.lowyinstitute.org/the-interpreter/which-side-would-us-public-choose-india-china-conflict	0,924592948331135
12/08/2020	https://thediplomat.com/2020/08/indian-frustration-with-china-grows/	0,912285547064557
19/08/2020	https://carnegieendowment.org/2020/08/19/at-crossroads-china-india-nuclear-relations-after-border-clash-pub-82489	0,901187255801879
31/08/2020	https://www.bbc.com/news/world-asia-india-53971397	0,86781243433253 8
31/08/2020	https://thediplomat.com/2020/08/china-attempts-to-shift-its-boundary-with-india-in-ladakh-again/	0,929652677685964
05/09/2020	https://thediplomat.com/2020/09/india-china-tensions-spike-in-the-himalayas/	0,907812878481198
09/09/2020	https://www.thequint.com/voices/opinion/india-china-border-conflict-pangong-tso-ladakh-indian-army-chinese-army-strategy	0,914690324742354
09/09/2020	https://www.dw.com/en/india-china-border-row-raises-tensions/a-54870600	0,797545122623281
13/09/2020	https://www.ft.com/content/ecd85303-9818-45c7-9e9d-753e88f7b7f5	0,662223980826099
15/09/2020	https://www.colorado.edu/polisci/2020/09/15/china-india-border-conflict-great-game-and-sino-indian-split	0,919850427589507
18/09/2020	https://www.indiatoday.in/news-analysis/story/china-avoid-conflict-secretly-planning-counterattack-india-ladakh-standoff-172	0,919870018407835
24/09/2020	https://thediplomat.com/2020/09/the-india-china-ladakh-crisis-why-so-silent-world/	0,940198964118209
07/10/2020	https://www.voanews.com/south-central-asia/india-races-building-roads-tense-india-china-border	0,937598946642801
10/10/2020	https://thewire.in/security/india-china-lac-border-dispute	0,949104326195341
12/10/2020	https://www.aljazeera.com/news/2020/10/12/india-china-hold-fresh-military-level-talks-to-end-border-row	0,979615562268959
15/10/2020	https://www.e-ir.info/2020/10/15/the-india-china-conflict-in-the-himalaya-should-break-open-ir-theory/	0,937976809336539
16/10/2020	https://swarajyamag.com/defence/majors-squad-the-game-of-alliances-in-a-possible-india-china-conflict	0,927003731963016
17/10/2020	https://www.businesstoday.in/current/economy-politics/india-china-clashes-border-conflict-left-relationship-profoundly-disturb	0,921729570045455
22/10/2020	https://www.eastasiaforum.org/2020/10/23/how-russia-emerged-as-key-mediator-in-the-china-india-dispute/	0,907283645521326
29/10/2020	https://carnegieendowment.org/2020/10/29/after-border-clash-will-china-india-competition-go-nuclear-pub-83072	0,946913762240177
06/11/2020	https://www.bloomberg.com/news/articles/2020-11-06/india-defense-chief-says-china-clash-may-lead-to-wider-conflict	0,676522232781024
06/11/2020	https://www.businessinsider.com/india-general-warns-china-border-dispute-could-escalate-to-war-2020-11	0,945225009474206
06/11/2020	https://www.aljazeera.com/news/2020/11/6/indias-top-general-says-face-off-with-china-could-spark-bigger	0,974796411867952
11/11/2020	https://www.dw.com/en/india-china-military-standoff-continues/a-55553845	0,866493530106006
11/11/2020	https://www.indiatvnews.com/news/india/india-china-border-dispute-resolution-3-step-disengagement-plan-pangong-lake-lada	0,849980288412218
10/11/2020	https://www.dw.com/en/india-china-military-standoff-continues/a-55553845	0,866493530106006
15/11/2020	https://www.ispionline.it/en/pubblicazione/sino-indian-border-dispute-has-india-usable-past-28255	0,928250503906886
16/11/2020	https://www.thequint.com/voices/opinion/united-states-allies-india-china-border-standoff-throwback-to-cold-war-era-new-amer	0,942501638774143
20/11/2020	https://indianexpress.com/article/opinion/chinas-endgame-in-ladakh-7056037/	0,941553749187816
22/11/2020	https://eurasiantimes.com/china-not-looking-for-solutions-wants-the-keep-the-indian-border-inflamed-swedish-author/	0,91953947045707
23/11/2020	https://www.lawfareblog.com/india-and-china-border-briefer-shadow-article-370s-revocation	0,781939750975949
01/12/2020	https://www.indiatoday.in/india/story/china-rotating-troops-daily-ladakh-standoff-sites-winters-1745721-2020-12-01	0,811670951750415
03/10/2020	https://carnegieindia.org/2020/12/03/china-should-beware-of-economic-consequences-of-fighting-with-india-pub-83375	0,939104977222882
03/12/2020	https://www.express.co.uk/news/world/1367686/India-china-news-Galway-valley-us-world-war-3-USCC-report-world-war-3-late	0,912533709188546
05/12/2020	https://indianexpress.com/article/explained/how-the-indian-army-stays-fighting-fit-on-the-lac-in-harsh-winters-7074943/	0,941755788395269
06/12/2020	https://www.theguardian.com/politics/2020/dec/06/ministers-face-double-defeat-in-lords-over-china-trading-links	0,935593665385658
06/12/2020	https://thewire.in/external-affairs/what-changed-india-china-ties-2020-result-rising-tensions	0,939567651719013
07/12/2020	https://www.scmp.com/topics/china-india-border-dispute	0,860657832004403
07/12/2020	https://www.express.co.uk/news/world/1369170/china-india-news-eastern-ladakh-lac-border-army-military-world-war-3-latest-r	0,906718460940303
07/12/2020	https://www.livemint.com/news/india/india-accuses-china-of-helping-rebel-groups-on-myanmar-border/amp-11607311466176.	0,96100425878223

5.1.2 Eventos extremos naturais - ocorrência de seca, queimadas e desmatamentos

Para o portfólio de eventos naturais sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica coletaram-se notícias de janeiro a dezembro de 2015 a 2020 que contenham informações sobre a ocorrência de seca, queimadas e desmatamentos nessa região. Os portfólios de cada ano são apresentados nas Seções 5.1.2.1, 5.1.2.2, 5.1.2.3, 5.1.2.4, 5.1.2.5 e 5.1.2.6, respectivamente.

5.1.2.1 Ocorrência de seca, queimadas e desmatamentos para o ano de 2015

O portfólio de ameaça de ocorrência de seca, queimadas e desmatamentos para o de 2015 possui 32 notícias, com as datas no período de janeiro de 2015 e dezembro de 2015. A Figura 5.4 apresenta o portfólio com informações de Data, URL e Grau de Similaridade com Ameaça.

Figura 5.4 - Visão geral do portfólio de ameaças para a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para o ano de 2015.

Data	URL	Grau de Similaridade com Ameaça (GSA)
06/01/2015	https://www.socioambiental.org/pt-br/noticias-socioambientais/incendios	0,910838951930764
29/01/2015	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/201	0,91387504551083
08/02/2015	https://www.amazoniasocioambiental.org/pt-br/radar/o-perigo-de-os-rios	0,878073807794322
02/03/2015	https://exame.com/brasil/desmatamento-aumentou-282-na-amazonia-leg	0,87304296051469
03/03/2015	https://mst.org.br/2015/03/03/desmatamento-na-amazonia-ja-afeta-o-cli	0,898870257799185
20/03/2015	https://www.correiobraziliense.com.br/app/noticia/ciencia-e-saude/2015	0,924840659632992
21/03/2015	http://g1.globo.com/jornal-nacional/noticia/2015/03/desmatamento-na-a	0,764229911547158
21/03/2015	https://exame.com/brasil/desmatamento-aumentou-282-na-amazonia-leg	0,87304296051469
10/04/2015	https://amazonia.org.br/2015/04/desmatamento-a-falta-de-agua-comeca-	0,918307432352203
14/04/2015	https://ciclovivo.com.br/planeta/meio-ambiente/greenpeace-associa-desm	0,901987741089078
21/04/2015	https://mauriciotuffani.blogfolha.uol.com.br/2015/04/21/desmatamento-n	0,888122805851676
22/05/2015	https://alfonsin.com.br/desmatamento-na-amaznia-mantm-tendncia-de-d	0,898384095343391
11/06/2015	http://www.upsa.com.br/2015/06/11/entrevista-com-secretario-andre-lima	0,890659399482609
25/06/2015	https://www.oeco.org.br/reportagens/29200-queimadas-em-unidades-de-	0,852760266306481
30/06/2015	https://www.oeco.org.br/reportagens/29206-desmatamento-avanca-na-re	0,879440663892557
20/07/2015	https://www.ecodebate.com.br/2015/07/27/incendios-florestais-e-queima	0,887806529339646
10/08/2015	https://agazetadoacre.com/2015/08/noticias/geral/periodo-seco-associa	0,900342531302654
28/08/2015	https://amazonia.org.br/2015/08/para-fugir-de-monitoramento-por-sateli	0,895288001279877
21/09/2015	https://amazoniareal.com.br/um-verao-tipico-do-el-nino-e-a-ausencia-de-	0,895566539149515
01/10/2015	https://amazoniareal.com.br/inpe-registrou-95-a-mais-de-queimadas-e-in	0,885642645562978
03/10/2015	https://amazoniareal.com.br/nao-adianta-culpar-o-para-fumaca-em-mana	0,878951717009428
13/10/2015	http://g1.globo.com/ma/maranhao/noticia/2015/10/em-26-anos-maranhao	0,854196371425068
22/10/2015	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/20	0,849632142607815
11/11/2015	https://amazonia.org.br/2015/11/por-tras-da-alta-das-queimadas-esta-	0,901538149132405
24/11/2015	https://conexaoplaneta.com.br/blog/amazonia-mais-da-metade-das-espe	0,899069821256005
25/11/2015	https://www.voaportugues.com/a/aquecimento-global-deve-causar-seca	0,907176497125915
26/11/2015	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/ag-estado/	0,910866427431441
26/11/2015	https://epoca.globo.com/colunas-e-blogs/blog-do-planeta/noticia/2015/	0,903921668252564
10/12/2015	https://www.dw.com/pt-br/o-efeito-das-mudan%C3%A7as-clim%C3%A1	0,918540753229909
10/12/2015	https://amazonia.org.br/2015/12/focos-de-queimadas-em-rondonia-aume	0,907212322782824
11/12/2015	https://revistapesquisa.fapesp.br/seca-ameaca-a-amazonia/	0,91443390792154
24/12/2015	https://portalamazonia.com/noticias/mudancas-climaticas-previsoes-alar	0,811030973936838

5.1.2.2 Ocorrência de seca, queimadas e desmatamentos para o ano de 2016

O portfólio de ameaça de ocorrência de seca, queimadas e desmatamentos para o de 2016 possui 43 notícias, com as datas no período de janeiro de 2016 e dezembro de 2016. A Figura 5.5 apresenta o portfólio com informações de Data, URL e Grau de Similaridade com Ameaça.

Figura 5.5 - Visão geral do portfólio de ameaças para a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para o ano de 2016.

Data	URL	Grau de Similaridade com Ameaça (GSA)
04/01/2016	https://amazonia.org.br/2016/01/mudancas-climaticas-previsoes-alarmantes-para-a-amazo	0,898120828087048
30/01/2016	https://amazoniaextrema.webflow.io/	0,910446112698413
15/01/2016	https://domtotal.com/noticias/detalhes.php?notId=983919	0,918958077878661
02/02/2016	https://www.greenme.com.br/informarse/biodiversidade/2837-queimadas-na-amazonia-dest	0,869043535836193
05/02/2016	https://www.socioambiental.org/pt-br/noticias-socioambientais/roraima-tem-pior-seca-em-1	0,904379242675214
29/02/2016	http://amazonia.org.br/2016/02/amazonia-extrema-como-secas-queimadas-e-desmatament	0,91338315555807
01/03/2016	https://amazoniareal.com.br/12872el-nino-godzilla-afeta-norte-da-amazonia-com-seca-prolo	0,873527111271606
01/03/2016	http://g1.globo.com/rr/roraima/noticia/2016/03/rr-tem-segunda-maior-taxa-de-desmatamento	0,85844597912396
02/03/2016	http://www.ihu.unisinos.br/185-noticias/noticias-2016/552127-amazonia-extrema-como-sec	0,905720044245403
26/03/2016	https://museudoamanha.org.br/livro/04-mudancas-climaticas-a-complexa-engrenagem-que	0,913076536387444
22/04/2016	https://amazoniareal.com.br/passaros-raros-fugiram-de-queimadas-no-parque-do-virua-em-	0,887837879062995
03/05/2016	https://amazonia.org.br/2016/05/desmatamento-nos-meses-de-fevereiro-e-marco-de-2016-a	0,849660742700531
19/05/2016	https://agenciabrasil.ebc.com.br/geral/noticia/2016-05/amazonia-ameacada-guardioes-lutar	0,912262384408018
09/06/2016	https://www.socioambiental.org/pt-br/blog/blog-do-xingu/de-olho-no-xingu-parte-ii-evolucao-	0,913300752996534
23/06/2016	https://www.socioambiental.org/pt-br/blog/blog-do-xingu/de-olho-no-xingu-parte-ii-evolucao-	0,913300752996534
29/06/2016	https://brasil.elpais.com/brasil/2016/06/29/ciencia/1467211200_184288.html	0,899049039698475
01/07/2016	https://amazonia.org.br/2016/07/amazonia-deve-ter-recorde-de-queimadas/	0,911581331441661
02/07/2016	https://exame.com/brasil/amazonia-pode-ter-queimada-recorde-em-2016/	0,922914328547721
02/07/2016	https://www.portaldoholanda.com.br/amazonia/amazonia-pode-ter-pior-temporada-de-queim	0,923953975680688
03/07/2016	https://amazonianarede.com.br/acre-decreta-situacao-de-alerta-devido-a-seca-no-rio-acre/	0,850947851221017
06/07/2016	https://epoca.globo.com/colunas-e-blogs/blog-do-planeta/noticia/2016/07/nasa-preve-pior-	0,902435876883004
24/07/2016	https://www.diariodaamazonia.com.br/nasa-divulga-boletim-que-preve-temporada-de-seca-i	0,902031275186594
08/08/2016	https://amazoniareal.com.br/estiagem-severa-e-queimadas-atingem-centro-sul-da-amazoni	0,900845023756936
08/08/2016	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2016/08/08/brasil-ja-t	0,874022226796111
10/08/2016	https://diarionline.com.br/?s=noticia&id=87282	0,893025272272964
18/08/2016	https://www.survivalbrasil.org/ultimas-noticias/11392	0,888663356619299
30/08/2016	http://www.ihu.unisinos.br/185-noticias/noticias-2016/559451-amazonia-e-cerrado-sao-os-b	0,869400729264044
05/09/2016	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2016/09/05/a-amazoni	0,889244933295858
08/09/2016	https://amazonia.org.br/2016/09/no-mato-grosso-cientistas-projetam-o-fim-da-amazonia-pe	0,909919783484553
15/09/2016	https://projetocolabora.com.br/florestas/floresta-em-chamas/	0,86554472421657
26/09/2016	https://dicasdevestibular.blogosfera.uol.com.br/2016/09/26/entenda-os-fatores-naturais-do	0,332937037950278
26/09/2016	https://www.ecodebate.com.br/2016/09/26/numero-de-focos-de-incendios-florestais-subiu-4	0,832005229876259
05/10/2016	https://www.terra.com.br/noticias/ciencia/sustentabilidade/meio-ambiente/desmatamen	0,898500328350407
06/10/2016	http://www.ihu.unisinos.br/78-noticias/560880-aumento-de-desmatamento-na-amazonia-e-s	0,85482777606943
06/10/2016	https://www.socioambiental.org/pt-br/noticias-socioambientais/desmatamento-em-terras-in	0,899864137693955
01/11/2016	https://www.socioambiental.org/pt-br/noticias-socioambientais/incendios-florestais-da-ul	0,921428509779674
11/11/2016	https://www.xapuri.info/amazonia-agenda/amazonia-floresta-que-arde-em-chamas/	0,862625621612414
25/11/2016	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2016/11/25/fogo-na-a	0,89505931012658
25/11/2016	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2016/11/25/incendios	0,901248320710268
30/11/2016	https://epoca.globo.com/ciencia-e-meio-ambiente/blog-do-planeta/noticia/2016/11/como-ex	0,880621855750971
30/11/2016	https://landportal.org/pt/news/2016/12/desmatamento-dispara-na-amaz%C3%B4nia	0,886221010464545
06/12/2016	http://www.ihu.unisinos.br/159-noticias/entrevistas/563108-incendios-na-floresta-amazonic	0,894913248093774
21/12/2016	https://amazonia.org.br/2016/12/queimadas-destroem-78-da-biodiversidade-da-amazonia/	0,902625655383079

5.1.2.3 Ocorrência de seca, queimadas e desmatamentos para o ano de 2017

O portfólio de ameaça de ocorrência de seca, queimadas e desmatamentos para o ano de 2017 possui 41 notícias, com as datas no período de janeiro de 2017 e dezembro de 2017. A Figura 5.6 apresenta o portfólio com informações de Data, URL e Grau de Similaridade com Ameaça.

Figura 5.6 - Visão geral do portfólio de ameaças para a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para o ano de 2017.

Data	URL	Grau de Similaridade com Ameaça (GSA)
05/01/2017	http://www.ihu.unisinos.br/186-noticias/noticias-2017/563643-queimadas-destroem-78-da-biodivers	0,887017797255966
10/01/2017	https://www.xapuri.info/amazonia-agenda/mudancas-climaticas-em-terras-indigenas-ver-de-perto-p	0,846181153032503
12/01/2017	https://www.gazetadopovo.com.br/agronegocio/clima/mesmo-com-grandes-secas-amazonia-se-ma	0,922496147354956
06/02/2017	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2017/02/06/queimadas-no-xin	0,894356277336495
07/02/2017	https://domtotal.com/noticias/detalhes.php?notId=1123739	0,898578049231026
25/02/2017	https://www.ecoamazonia.org.br/2017/02/nivel-desmatamento-seguro-clima-amazonia/	0,879261640394437
02/03/2017	https://jornal.usp.br/ciencias/ciencias-ambientais/nature-discute-efeito-do-desmatamento-em-chuv	0,91219595964157
13/03/2017	https://www.dw.com/pt-br/amaz%C3%B4nia-pode-entrar-em-ciclo-de-desmatamento-e-seca-diz-es	0,854374323798321
18/03/2017	http://crisalida.eco.br/efeitos-da-seca-acentuam-de-maneira-permanente-perda-florestal-na-amazor	0,889010467621521
20/03/2017	https://www.pirelli.com/global/pt-br/life/mudanca-climatica-como-combater-o-aquecimento-global-e	0,886959687392639
29/03/2017	https://www.ecodebate.com.br/2017/03/29/vegetacao-exuberante-da-amazonia-corre-risco-de-ser-s	0,905523987762795
11/04/2017	https://www.funverde.org.br/blog/vegetacao-exuberante-da-amazonia-corre-risco-de-ser-substituida	0,917668689543203
15/04/2017	https://www.goiasmais20.com.br/5-grandes-problemas-que-as-unidades-de-conservacao-enfrentan	0,863267023108934
15/05/2017	https://medium.com/biomas-brasileiros/amaz%C3%B4nia-f16ab3fe4723	0,938652527094786
18/06/2017	https://g1.globo.com/ac/acre/noticia/orgaos-ambientais-fazem-forca-tarefa-contra-desmatamento-e	0,815806283643985
20/06/2017	https://amazonia.org.br/2017/06/orgaos-ambientais-fazem-forca-tarefa-contra-desmatamento-e-que	0,887890958012657
22/06/2017	http://florestasilenciosa.ambiental.media/fogo-na-amazonia/	0,913122533988239
23/06/2017	http://www.barradocorda.com/destaques/transamazonica-crime-e-abandono-estrada-aberta-na-dita	0,903980587177842
29/06/2017	https://exame.com/revista-exame/brasil-destroi-128-campos-de-futebol-de-floresta-por-hora/	0,895489865960617
19/07/2017	https://outraspalavras.net/blog/contra-as-florestas-no-mercado-de-carbono/	0,751555734836827
26/07/2017	https://amazoniareal.com.br/danos-sociais-de-enchente-do-rio-branco-estao-associados-a-ocupaca	0,901269817265861
04/08/2017	https://jp.reuters.com/article/politica-inpe-queimadas-idLTAKCN1VA1H9	0,830962912175405
07/08/2017	http://amazonia.org.br/2017/08/desmatamento-na-floresta-amazonica-maranhense-e-de-75-denunc	0,904043061259438
08/08/2017	http://www.ihu.unisinos.br/78-noticias/570381-acao-humana-contribuiu-para-seca-sem-precedentes	0,894721476810699
25/08/2017	https://epoca.globo.com/ciencia-e-meio-ambiente/blog-do-planeta/noticia/2017/08/praticas-conserv	0,890899752218731
05/09/2017	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2017/09/05/pecuaria-e-respo	0,903892277858036
16/09/2017	https://amazonia.org.br/2017/09/emissoes-de-carbono-podem-aumentar-ate-90-na-amazonia-em-50	0,90610469924533
23/09/2017	https://g1.globo.com/am/amazonas/noticia/am-registra-aumento-nos-focos-de-queimadas-com-ma	0,860732635934409
26/09/2017	https://www.ecodebate.com.br/2017/09/26/com-aumento-da-temperatura-e-secas-extremas-amazo	0,851926139713386
30/09/2017	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2017/09/30/desmatamento-pa	0,907376373344225
11/10/2017	http://www.ihu.unisinos.br/78-noticias/572532-rio-araguaia-pode-secar-em-40-anos-por-causa-do-d	0,854789983641761
18/10/2017	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/ag-estado/2017/10/18/estudo-inedito	0,927359687607147
24/10/2017	https://www.correiodopovo.com.br/not%C3%ADcias/mundo/amaz%C3%B4nia-perdeu-3-7-milh%C	0,800409836906625
01/11/2017	https://www.socioambiental.org/pt-br/noticias-socioambientais/desmatamento-cresce-32-nas-terras	0,896877215018311
06/11/2017	http://www.ihu.unisinos.br/78-noticias/573342-desmatamento-cresce-32-nas-terras-indigenas-da-ar	0,888970882498488
14/11/2017	https://www.uol/noticias/especiais/amazonia-que-arde.htm#amazonia-que-arde	0,891733495100856
27/11/2017	http://www.solam.com.br/blog/?p=5141	0,912208538898117
18/12/2017	https://imazon.org.br/imprensa/brasil-termina-2017-com-numero-recorde-de-queimadas-desde-1999	0,893301891929629
18/12/2017	http://opiniaoenoticia.com.br/brasil/brasil-registra-recorde-de-queimadas/	0,911771204273156
19/12/2017	https://br.sputniknews.com/brasil/2017121910107781-conflitos-fundiarios-principais-responsaveis-	0,903998722072881
19/12/2017	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/ag-estado/2017/12/19/brasil-perdeu-95-m	0,892966126014491

5.1.2.4 Ocorrência de seca, queimadas e desmatamentos para o ano de 2018

O portfólio de ameaça de ocorrência de seca, queimadas e desmatamentos para o ano de 2018 possui 44 notícias, com as datas no período de janeiro de 2018 e dezembro de 2018. A Figura 5.7 apresenta o portfólio com informações de Data, URL e Grau de Similaridade com Ameaça.

Figura 5.7 - Visão geral do portfólio de ameaças para a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para o ano de 2018.

Data	URL	Grau de Similaridade com Ameaça (GSA)
02/01/2018	https://www.agazeta.com.br/brasil/atividade-humana-poe-ecossistemas-brasileiros-em-risco-0118	0,752229159478508
05/01/2018	https://amazonia.org.br/2018/01/ainda-e-pouco-maior-projeto-de-reflorestamento-da-historia-recupera-m	0,889904516556909
18/01/2018	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2018/01/18/desmontado-esquema-ile	0,920171644529957
01/02/2018	http://www.folhadomeio.com.br/fma_nova/noticia.php?id=4516	0,727216945963315
15/02/2018	https://blogdopedlowski.com/2018/02/15/eventos-de-secas-extremas-deverao-aumentar-emissao-de-co2	0,681700866295042
22/02/2018	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2018/02/22/desmatamento-na-amazo	0,933679657940855
02/03/2018	https://istoe.com.br/amazonia-nao-esta-longe-de-virar-savana/	0,910734252161897
02/03/2018	https://conexaoplaneta.com.br/blog/desmatamento-na-amazonia-esta-prestes-atingir-limite-irreversivel/	0,917872679060332
06/03/2018	https://amazonasatual.com.br/mudancas-climaticas-influenciam-os-riscos-de-extincao-de-calangos/	0,893917706286038
13/03/2018	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2018/03/13/terra-de-ninguem-floresta	0,877486055709206
26/04/2018	https://conexaoplaneta.com.br/blog/a-exuberancia-e-o-desmatamento-da-amazonia-sob-o-olhar-de-rogeri	0,75212243514357
26/04/2018	https://conexaoplaneta.com.br/blog/das-araucarias-a-amazonia-os-insaciaveis-desmatadores-sulistas/	0,884319959941168
03/05/2018	https://amazonia.org.br/2018/05/amazonia-maior-afetada-por-queimadas-em-2018/	0,879361995789251
03/05/2018	https://www.climatempo.com.br/noticia/2018/05/03/amazonia-maior-afetada-por-queimadas-em-2018-989	0,752980092951705
08/05/2018	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2018/05/08/mais-de-120-mil-hectare	0,906124862734303
13/05/2018	http://g1.globo.com/economia/agronegocios/globo-rural/noticia/2018/05/operacao-policial-combate-extrac	0,821339054972845
29/05/2018	https://noticias.uol.com.br/saude/ultimas-noticias/redacao/2018/05/29/quilometro-quadrado-desmatado-	0,891581763831906
02/06/2018	https://www.pensamentoverde.com.br/meio-ambiente/desmatamento-na-amazonia-esta-prestes-a-atingir	0,889284698995402
03/06/2018	https://espacoecologiconoar.com.br/desmatamento-na-amazonia-esta-prestes-a-atingir-limite-irreversivel/	0,899493471401729
20/06/2019	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2018/06/20/aquecimento-global-deixa	0,936979710081534
27/06/2018	https://blog.globalforestwatch.org/pt/data-and-research/2017-foi-o-segundo-pior-ano-ja-registrado-de-perd	0,929056086389939
29/06/2018	https://wribrasil.org.br/pt/blog/2018/06/2017-foi-o-segundo-pior-ano-ja-registrado-de-perda-de-cobertura-flo	0,920043361121528
04/07/2018	https://www.nationalgeographicbrasil.com/meio-ambiente/2018/07/diminuicao-florestas-tropicais-desmat	0,93249162912441
25/07/2018	https://pn7.com.br/cerrado-e-um-dos-biomas-mais-vulneraveis-durante-a-estacao-seca/	0,883939119285355
10/08/2018	https://www.greenpeace.org/brasil/blog/fogo-na-floresta-comecou-a-temporada-de-incendios/	0,914165576784173
11/08/2018	https://amazonia.org.br/2018/08/fogo-na-floresta-comecou-a-temporada-de-incendios/	0,915030604630961
22/08/2018	https://www.ecoamazonia.org.br/2018/08/conservar-amazonia-questao-ambiental-social-economica/	0,863632022082438
25/08/2018	http://www.ihu.unisinos.br/78-noticias/582144-imagens-de-satelite-registram-um-mundo-em-chamas	0,882511849809227
26/08/2018	https://g1.globo.com/am/amazonas/noticia/2018/08/26/alteracoes-extremas-no-regime-de-seca-e-cheia-p	0,814705114721011
05/09/2018	https://projetocolabora.com.br/ods15/os-incendios-que-queimam-a-amazonia/	0,866292415941659
07/09/2018	https://projetocolabora.com.br/ods14/floresta-amazonica-arde-e-a-culpa-nao-e-da-vitima/	0,861501412072504
21/09/2018	https://amazonia.org.br/2018/09/dia-da-arvore-desmatamento-vem-caindo-mas-ainda-ha-desafios/	0,889207162963637
15/10/2018	http://www.diretodaciencia.com/2018/10/15/amazonia-tem-mais-chuvas-extremas-no-norte-e-mais-secas	0,908162997594104
23/10/2018	https://www.greenpeace.org/brasil/blog/em-2018-focos-de-calor-se-concentraram-no-oeste-da-amazonia/	0,938837184517717
24/10/2018	https://www.ecodebate.com.br/2018/10/24/eliminar-as-queimadas-e-um-desafio-que-tera-de-ser-enfrentad	0,859831925404942
30/10/2018	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/bbc/2018/10/30/amazonia-perdeu-20-desde-19	0,896699931697627
21/11/2018	https://www.nationalgeographicbrasil.com/meio-ambiente/2018/11/diminuicao-amazonia-floresta-amazon	0,930558351368447
26/11/2018	https://www.nationalgeographicbrasil.com/meio-ambiente/2018/11/desmatamento-amazonia-recorde-falta	0,885602055428058
23/11/2018	https://g1.globo.com/natureza/noticia/2018/11/23/desmatamento-na-amazonia-cresce-137-entre-2017-e-2	0,837097106525933
29/11/2018	https://www.socioambiental.org/pt-br/blog/blog-do-isa/o-descontrole-deliberado-do-desmatamento-na-ama	0,898667872502972
03/12/2018	https://portal.fiocruz.br/noticia/estudo-alerta-para-impactos-das-mudancas-climaticas-na-saude	0,87626084947162
12/12/2018	https://www.socioambiental.org/pt-br/noticias-socioambientais/desmatamento-em-terras-indigenas-cresc	0,908391032495604
14/12/2018	https://www.ecodebate.com.br/2018/12/14/o-desmatamento-na-amazonia-e-sua-influencia-no-regime-de-c	0,863015353632618
17/12/2018	https://amazoniareal.com.br/amazonia-e-o-aquecimento-global-5-secas/	0,872910655922484

5.1.2.5 Ocorrência de seca, queimadas e desmatamentos para o ano de 2019

O portfólio de ameaça de ocorrência de seca, queimadas e desmatamentos para o ano de 2019 possui 50 notícias, com as datas no período de janeiro de 2019 e dezembro de 2019. A Figura 5.8 apresenta o portfólio com informações de Data, URL e Grau de Similaridade com Ameaça.

Figura 5.8 - Visão geral do portfólio de ameaças para a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para o ano de 2019.

Data	URL	Grau de Similaridade com Ameaça (GSA)
08/01/2019	https://www.nationalgeographicbrasil.com/entrevista/2019/01/entrevista-carlos-nobre-clima-amazonia-bolsonar	0,893249044381472
12/01/2019	https://www.redebrasilatual.com.br/ambiente/2019/01/amazonia-esta-se-transformando-em-cerrado-apontam-e	0,873174540998174
19/01/2019	https://www.redebrasilatual.com.br/ambiente/2019/01/leonardo-dicaprio-denuncia-avanco-do-desmatamento-d	0,896152086236944
21/01/2019	https://agazetadoacre.com/2019/01/artigos/o-que-a-california-e-esta-parte-da-amazonia-tem-em-comum-incen	0,902966629179837
11/02/2019	https://www.ecoamazonia.org.br/2019/02/tempestades-vento-intensificam-degradacao-florestas/	0,889488054933935
13/02/2019	http://www.neomondo.org.br/2019/02/13/como-a-floresta-amazonica-enfrenta-a-seca/	0,774346881377425
13/02/2019	https://climainfo.org.br/2019/02/13/secas-na-amazonia-terao-prevalencia-aumentada-e-serao-mais-severas/	0,929517812383247
14/02/2019	https://www.nationalgeographicbrasil.com/meio-ambiente/2019/02/os-ultimos-cinco-anos-foram-os-mais-quent	0,888368183641256
21/02/2019	https://www.amda.org.br/index.php/comunicacao/noticias/5487-fogo-e-desmatamento-deixam-amazonia-vulne	0,920752037020686
02/03/2019	http://www.ihu.unisinos.br/78-noticias/587143-imazon-desmatamento-da-amazonia-aumenta-54-em-janeiro-de	0,839644264830376
07/03/2019	https://amazoniareal.com.br/amazonia-e-o-aquecimento-global-14-conclusao/	0,858432842536081
14/03/2019	https://mst.org.br/2019/03/14/exploracao-da-amazonia-ira-aumentar-conflitos-por-terra-e-dependencia-do-agro	0,880469478780884
25/03/2019	http://www.observatorioeco.com.br/index.php/2019/03/amazonia-sofre-mudancas-intensas-e-nocivas-alerta-an	0,895479896625103
08/04/2019	http://www.cemaden.gov.br/pesquisadores-do-cemaden-desenvolvem-um-sistema-para-monitorar-risco-de-inc	0,9064256505507
28/04/2019	https://g1.globo.com/natureza/noticia/2019/04/28/amazonia-perdeu-18percent-da-area-de-floresta-em-tres-dec	0,826779705631454
30/04/2019	https://www.institutoclaro.org.br/cidadania/nossas-novidades/noticias/desmatamento-na-amazonia-legal-cres	0,900527624172467
07/05/2019	https://www.nationalgeographicbrasil.com/meio-ambiente/2019/05/devastacao-floresta-amazonica-bolsonaro-a	0,921591263226751
14/05/2019	http://www.cemaden.gov.br/estudos-sobre-impactos-dos-incendios-florestais-estimam-perda-acumulada-no-ac	0,908364652814251
22/05/2019	https://noticias.uol.com.br/ultimas-noticias/agencia-estado/2019/05/22/desmatamento-avanca-na-amazonia-qu	0,911661316519356
29/05/2019	http://www.ihu.unisinos.br/78-noticias/589563-imazon-desmatamento-na-amazonia-aumentou-20-entre-agosto	0,856325758621674
29/05/2019	https://www.agencianossa.com/2019/05/29/enfraquecimento-do-meio-ambiente-pode-afetar-abastecimento-de	0,904356661150067
05/06/2019	https://hypescience.com/desmatamento-da-amazonia-brasileira-atinge-recorde/	0,900019711516136
05/06/2019	https://amazonia.org.br/2019/06/ao-menos-13-milhoes-de-arvores-foram-derrubadas-ilegalmente-no-xingu-em	0,893920595565117
10/06/2019	https://projetocolabora.com.br/ods15/desmatamento-e-grilagem-deixam-amazonia-em-chamas/	0,868922954627551
14/06/2019	https://www.nationalgeographicbrasil.com/meio-ambiente/2019/06/governo-bolsonaro-desmatamento-amazon	0,894202892272221
02/07/2019	https://g1.globo.com/natureza/noticia/2019/07/02/desmatamento-na-amazonia-em-junho-e-o-pior-desde-2016.	0,849435360725539
12/07/2019	http://www.obt.inpe.br/OBT/noticias-obt-inpe/areas-de-floresta-amazonica-queimadas-equivalem-a-68-do-tot	0,902841948176177
15/07/2019	https://amda.org.br/index.php/comunicacao/noticias/5629-indice-de-desmatamento-na-amazonia-em-junho-e-o	0,893430448383841
24/07/2019	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2019/07/24/criticado-por-bolsonaro-inpe-e	0,907043106405546
28/07/2019	https://blogdacidadania.com.br/2019/07/nytimes-bolsonaro-ja-levou-1-330-milhoes-de-m%C2%B2-da-amazon	0,929663307158067
19/08/2019	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2019/08/19/numero-de-queimadas-cresce-	0,915405924273928
21/08/2019	https://veja.abril.com.br/blog/impacto/queimadas-na-amazonia-e-o-aumento-do-desmatamento/	0,844210310147013
21/08/2019	https://www.redebrasilatual.com.br/ambiente/2019/08/queimadas-na-amazonia-seguem-o-rastro-do-desmatam	0,901936824629866
21/08/2019	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2019/08/21/queimadas-na-amazonia-percer	0,913191068946151
03/09/2019	https://noticias.r7.com/brasil/como-a-amazonia-chegou-a-30901-focos-de-incendio-03092019	0,927014272821591
04/09/2019	https://www.ovale.com.br/_conteudo/pagina3/2019/09/86784-queimadas-na-amazonia-atingem-maior-indice-d	0,903404098402494
05/09/2019	https://g1.globo.com/natureza/noticia/2019/09/05/amazonia-tem-os-indices-de-chuvas-e-de-queimadas-mais-a	0,842267078884826
05/09/2019	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2019/09/05/desmatada-e-aquecida-amazon	0,900752178812422
01/10/2019	https://g1.globo.com/natureza/noticia/2019/10/01/focos-de-queimadas-na-amazonia-caem-em-setembro-e-sob	0,78637692388546
11/10/2019	https://www.dw.com/pt-br/desmatamento-na-amaz%C3%B4nia-cresce-quase-93-neste-ano-indica-inpe/a-5080	0,889720099086444
12/10/2019	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2019/10/12/menores-focos-de-incendio-na-	0,90227094277679
25/10/2019	https://www.brasil247.com/brasil/por-conta-de-queimadas-e-desmatamento-amazonia-deixara-de-produzir-chu	0,941083678661266
04/11/2019	http://www.somosamazonia.wwf.org.br/2019/11/12/fotos-contam-como-as-queimadas-sao-apenas-uma-parte-d	0,86525327864106
12/11/2019	http://www.somosamazonia.wwf.org.br/2019/11/12/fotos-contam-como-as-queimadas-sao-apenas-uma-parte-d	0,86525327864106
16/11/2019	https://www.redebrasilatual.com.br/ambiente/2019/11/cientistas-desmentem-bolsonaro-desmatamento-record	0,917708594575996
18/11/2019	https://projetocolabora.com.br/ods15/desmatamento-na-amazonia-vai-ultrapassar-os-10-mil-km%C2%B2/	0,867238815840356
02/12/2019	https://www.redebrasilatual.com.br/ambiente/2019/12/exploracao-madeira-queimadas-aumentar-2020/	0,880319447849728
04/12/2019	https://www.ecodebate.com.br/2019/12/04/desmatamento-na-amazonia-aumenta-212-em-outubro-deste-ano-a	0,839204999306272
14/12/2019	https://www.dw.com/pt-br/desmatamento-na-amaz%C3%B4nia-cresce-104-em-novembro/a-51667077	0,891850819475243
18/12/2019	https://reporterbrasil.org.br/2019/12/%EF%BB%BFqueimadas-na-amazonia-sao-3-vezes-mais-comuns-em-are	0,934762112603562

5.1.2.6 Ocorrência de seca, queimadas e desmatamentos para o ano de 2020

O portfólio de ameaça de ocorrência de seca, queimadas e desmatamentos par o ano de 2020 possui 43 notícias, com as datas no período de janeiro de 2020 e dezembro de 2020. A Figura 5.8 apresenta o portfólio com informações de Data, URL e Grau com Similaridade com Ameaça.

Figura 5.9 - Visão geral do portfólio de ameaças para a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para o ano de 2020.

Data	URL	Grau de Similaridade com Ameaça (GSA)
10/01/2020	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/redacao/2020/01/10/amazonia-esta-ficando-mais-seca-e-mais-propensa-as-queim	0,893964702563902
11/01/2020	https://brasil.elpais.com/ciencia/2020-01-11/mudanca-climatica-aumentara-os-incendios-na-floresta-amazonica.html	0,893577612046716
13/01/2020	https://www.jornaldocomercio.com/_conteudo/geral/2020/01/720377-amazonia-tem-risco-de-grandes-queimadas-em-2020.html	0,901696075411638
13/01/2020	http://amazonia.org.br/2020/01/clima-tornara-amazonia-duas-vezes-mais-inflamavel-neste-seculo/	0,909248672671188
30/01/2020	https://conexaoplaneta.com.br/blog/o-ar-antes-sempre-umido-esta-progressivamente-mais-seco-tornando-a-floresta-inflamavel-alerta-antonio	0,901099566811958
09/02/2020	https://climainfo.org.br/2020/02/09/satelite-revela-degradacao-generalizada-da-floresta-amazonica/	0,903992436887861
15/02/2020	https://www.dw.com/pt-br/sob-a-sombra-da-viol%C3%AAncia-na-amaz%C3%B4nia/a-52372640	0,917194963153136
18/02/2020	https://noticias.uol.com.br/meio-ambiente/ultimas-noticias/bbc/2020/02/18/o-que-ameaca-a-amazonia-em-seus-9-paises.htm	0,701143245636252
21/02/2020	https://www.modefica.com.br/desmatamento-secas-inundacoes-como-5-paises-tem-enfrentado-o-colapso-climatico/#.X4rSjCVv-V4	0,891219483900822
08/03/2020	https://www.cnnbrasil.com.br/nacional/2020/03/08/quais-os-riscos-do-avanco-do-desmatamento-na-amazonia	0,909880856239025
11/03/2020	https://www.dw.com/pt-br/amaz%C3%B4nia-pode-entrar-em-colapso-em-50-anos-diz-estudo/a-52723760	0,915482883113796
12/03/2020	https://diariodovale.com.br/colunas/nasa-preve-o-fim-da-amazonia/	0,919621870769726
27/03/2020	https://noticias.uol.com.br/ultimas-noticias/reuters/2020/03/27/exclusivo-brasil-reduz-fiscalizacao-ambiental-em-meio-ao-surto-de-coronaviru	0,896859998023104
20/04/2020	https://www.brasildefato.com.br/2020/04/20/covid-19-sera-cortina-de-fumaca-para-desmatamento-alerta-especialista-do-greenpeace	0,901451655410065
24/04/2020	https://jornal.usp.br/atualidades/queimadas-na-amazonia-tendem-a-ser-mais-intensas-este-ano/	0,897999032955972
30/04/2020	https://veja.abril.com.br/brasil/em-meio-a-pandemia-o-desmatamento-dispara-na-amazonia/	0,873539634778291
30/04/2020	https://revistagloborural.globo.com/Noticias/Sustentabilidade/noticia/2020/04/instituto-de-pesquisa-ambiental-da-amazonia-alerta-para-risc	0,893830352243556
01/05/2020	https://ciclovivo.com.br/planeta/meio-ambiente/desmate-aumenta-estacao-de-fogo/	0,89394605654546
04/05/2020	https://amazonia.org.br/2020/05/imagens-de-satelite-mostram-probabilidade-maior-de-seca-e-incendios-na-amazonia-este-ano/	0,904880914084333
26/05/2020	http://www.cemaden.gov.br/cientistas-alertam-para-a-contencao-das-queimadas-na-amazonia-e-o-colapso-do-sistema-de-saude-na-regiao/	0,876004601967683
27/05/2020	https://www.nytimes.com/pt/2020/05/27/opinion/international-world/amazonia-bolsonaro-quemas.html	0,92420511723515
04/06/2020	https://climainfo.org.br/2020/06/03/explosao-do-desmatamento-e-clima-mais-seco-devem-intensificar-queimadas-na-amazonia/	0,906494779841682
08/06/2020	https://www.dw.com/pt-br/desmatamento-e-fogo-podem-agravar-pandemia-na-amaz%C3%B4nia/a-53724198	0,911491545267731
15/06/2020	https://www.gazetadopovo.com.br/republica/queimadas-governo-plano-contingencia-amazonia/	0,893088717636119
21/06/2020	https://www.oeco.org.br/reportagens/destruicao-da-amazonia-pode-transforma-la-em-deserto-e-desencadear-pandemias/	0,882580789900088
25/06/2020	https://portalamazonia.com/noticias/cidades/acre-pode-ter-a-seca-mais-severa-dos-ultimos-anos-apontam-especialistas	0,926169752890766
09/07/2020	https://veja.abril.com.br/blog/impacto/a-projecao-da-nasa-para-a-temporada-de-queimadas-na-amazonia/	0,852539243790334
10/07/2020	https://isnportal.com.br/editorias/meio-ambiente/2020/07/10/nasa-alerta-para-risco-maior-de-incendios-na-amazonia-2020-esta-programado-p	0,903795929567082
19/07/2020	https://www.nationalgeographicbrasil.com/meio-ambiente/2020/07/temporada-movimentada-de-furacoes-no-atlantico-pode-gerar-mais-incend	0,910002760594493
28/07/2020	https://conexaoplaneta.com.br/blog/desmatamento-na-amazonia-seca-o-brasil-e-pode-levar-agronegocio-ao-colapso-alerta-relatorio-de-orgao	0,889712234392185
05/08/2020	https://projetocolabora.com.br/ods13/pouca-chuva-rios-secos-e-queimadas-assustam-pantanal/	0,862371352645768
07/08/2020	https://infoamazonia.org/pt/2020/08/portugues-avanco-do-desmatamento-e-tempo-seco-fazem-do-acre-o-mais-propicio-para-incendios-flores	0,917786163844833
14/08/2020	https://amazoniareal.com.br/amazonia-em-chamas-20-florestas-do-acre-podem-ser-mais-afetadas-por-incendios-diz-nasa-14-08-2020/	0,903044324136516
18/08/2020	https://noticias.uol.com.br/ultimas-noticias/afp/2020/08/18/desmatamento-e-impunidade-aticam-incendios-na-amazonia.htm	0,936240081226666
28/08/2020	https://g1.globo.com/am/amazonas/noticia/2020/08/28/viagens-de-barco-sao-suspensas-em-duas-cidades-do-alto-solimoes-no-am-por-conta	0,769455112241489
02/09/2020	https://acjornal.com/2020/09/28/sertao-acreano-seca-castiga-familias-da-floresta-estadual-do-antimary-que-caminham-horas-em-busca-de-ag	0,857053481009921
09/09/2020	https://pp.nexojornal.com.br/perguntas-que-a-ciencia-ja-respondeu/2020/As-queimadas-na-Amaz%C3%B4nia-explicadas-em-10-pontos	0,920193428588378
19/09/2020	https://www.em.com.br/app/noticia/nacional/2020/09/15/interna_nacional,1185431/seca-e-o-fogo-criminoso-devastam-quase-todos-os-bioma	0,893926405907694
28/09/2020	https://www.tecmundo.com.br/ciencia/204415-amazonia-aproxima-ponto-nao-se-recuperar.htm	0,916403278271325
03/10/2020	https://www.anda.jor.br/2020/10/03/incendios-na-floresta-amazonica-sao-os-piores-em-uma-decada/	0,610168845237549
09/10/2020	https://amazonia.org.br/2020/10/combustivel-perfeito-desmatamento-provoca-incendios-no-territorio-indigena-do-xingu-mt/	0,916597097719976
09/10/2020	https://www.istoedinheiro.com.br/desmatamento-da-amazonia-cai-pelo-terceiro-mes-consecutivo-mas-queimadas-sobem/	0,916040949741205
13/10/2020	https://amazonia.org.br/2020/10/combustivel-perfeito-desmatamento-provoca-incendios-no-territorio-indigena-do-xingu-mt/	0,916597097719976

As informações contidas nos portfólios de ameaças dos conflitos armados e de ocorrência de seca, queimadas e desmatamentos são utilizados como entrada da função de "*DFA*" para o cálculo do valor de "*slope*"(β). Esse valor de "*slope*" é utilizado como uma das entradas do algoritmo "*P-Model*" para a geração das séries temporais endógenas de ameaças. A utilização das informações dos portfólios com o algoritmo "*P-Model*" permitiu tanto realizar a validação do algoritmo como também apresentar uma das aplicações possíveis dos portfólios gerados. Essa validação do algoritmo é apresentada na Seção 5.2.

5.2 Resultado 2: validação "*P-Model*"

O algoritmo "*P-Model*" utilizado na tese foi apresentado por (BONN, 2019) e aplicado na análise de eventos extremos e na geração de séries temporais endógenas e exógenas por (ROSA et al., 2019). Esse algoritmo apesar de apresentar uma boa resposta para a aplicação de eventos extremos, ainda não passou por uma validação por dados fora do domínio de testes desse fim. Assim, os dados gerados pelos portfólios de ameaças têm como primeira aplicação a validação do algoritmo "*P-Model*" como gerador de séries temporais endógenas para eventos extremos. Para essa validação, consideraram-se as

séries temporais geradas pelos portfólios de ameaça para os conflitos armados entre a Síria e USA, Irã e USA e um possível conflito Global. Os testes consideram séries com 20 elementos devido à representatividade que essa quantidade de elementos demonstrou terna visualização dos dados. Para cada série de conflito considerou o cálculo dos valores de DFA, apresentado na Seção 3.6.1.1 e de autocorrelação (HYNDMAN; ATHANASOPOULOS, 2018). Esses cálculos foram realizados utilizando a biblioteca "*statsmodels*" da linguagem Python 3.7. Posteriormente, gerou-se no "*P-Model*" uma série endógena, com $p = 0,6$, de 2048 elementos, quantidade selecionada para se ter o maior número representativo de subséries, para cada um dos valores de β ("*slope*"), valores de espectro de potência, do conjunto valores: $-0,3$, $-0,4$, $-0,5$, $-0,8$, $-1,0$, $-1,34$ e $-2,0$. Esses valores de espectro de potência foram selecionados devido as suas características de não estacionaridade, como apresentado em Davis et al. (1997).

Para a série endógena do "*P-Model*" não se considerou os picos das séries ou valores maiores que $2,0$, devido ao ruído que esses valores geram nas séries. Assim, foi verificado para cada teste realizado que os valores de pico estavam entre $2,0$ e $5,0$ e esses valores foram desconsiderados, resultando em uma série de 1796 pontos. Essa série do foi convolucionada em séries de 20 pontos. Isso se deu da seguinte forma: a primeira série, consiste nos primeiros vintes pontos da série original, a segunda série, é formada considerando 20 pontos a partir do segundo elemento da série original, a terceira série, é formada considerando 20 pontos a partir do terceiro elemento da série original, e assim sucessivamente. Esse processo gerou 1796 novas séries, e para cada série foi gerado o valor de DFA e autocorrelação em comparação aos valores de DFA e autocorrelação das séries dos conflitos da Síria, Irã e do conflito Global. Os valores de DFA e Autocorrelação das séries dos conflitos foram comparados, em seus respectivos gráficos, com cada série gerada pelo processo de convolução da série endógena do "*P-Model*". No gráfico de DFA, no eixo X representa a disposição espacial dos valores DFA apresentados no eixo y. No gráfico de autocorrelação, no eixo x encontra-se o número de elementos das séries e no eixo y o valor da autocorrelação. A Tabela 5.1 apresenta os valores do cálculo do DFA para cada uma das séries de 20 pontos dos conflitos armados considerados.

Tabela 5.1 - Valores de DFA para as séries dos conflitos armados entre a Síria e USA, Irã e USA e Global.

Conflito Armado	**DFA**
Síria e USA	1,341020221418548
Irã e USA	1,2541438671395748
Global	1,2309744047538853

As Tabelas 5.2, 5.3 e 5.4, apresentam os valores de β ou "*slope*", o DFA dos conflitos e os gráficos de DFA e autocorrelação para as séries analisadas.

Tabela 5.2 - Comparação dos valores de DFA e autocorrelação das séries geradas pelo algoritmo "*P-Model*" com as séries dos conflitos armados entre a Síria e USA, Irã e USA e Global para os valores de $\beta = \{-0,3; -0,4\}$.

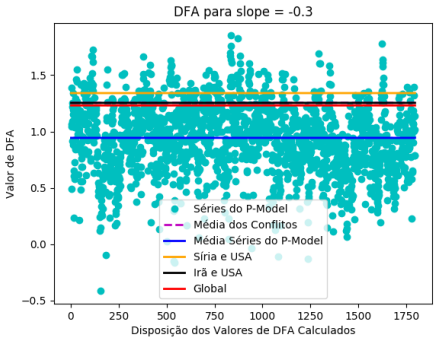
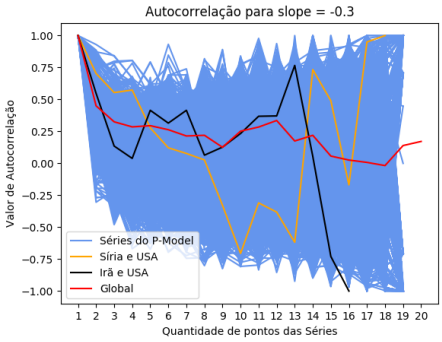
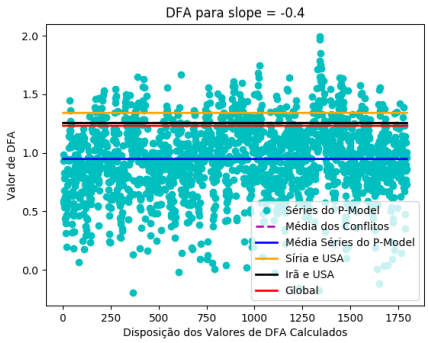
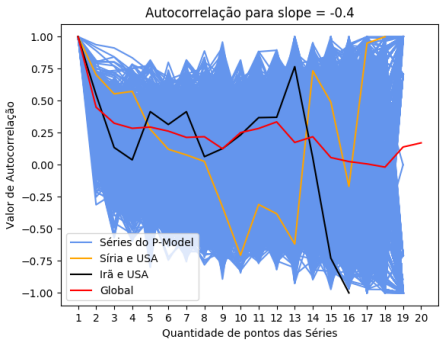
β	Gráfico DFA	Gráfico Autocorrelação
-0,3		
-0,4		

Tabela 5.3 - Comparação dos valores de DFA e autocorrelação das séries geradas pelo algoritmo "*P-Model*" com as séries dos conflitos armados entre a Síria e USA, Irã e USA e Global para os valores de $\beta = \{-0,5; -0,8; -1,0\}$.

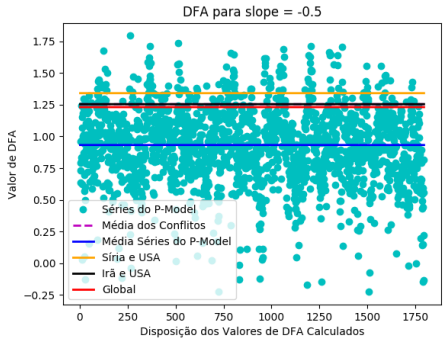
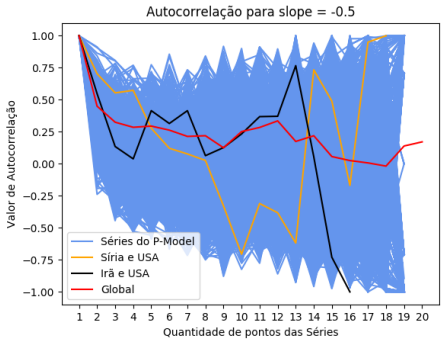
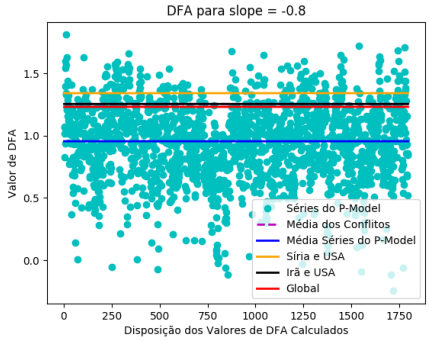
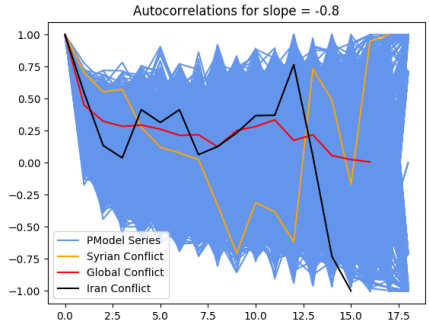
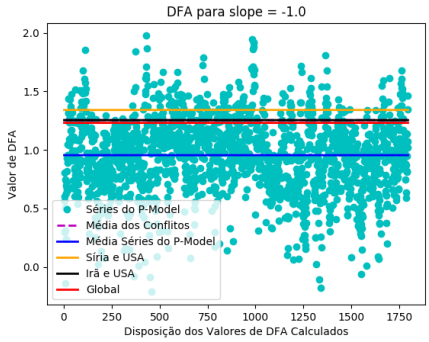
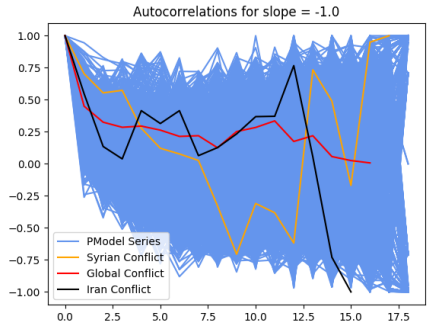
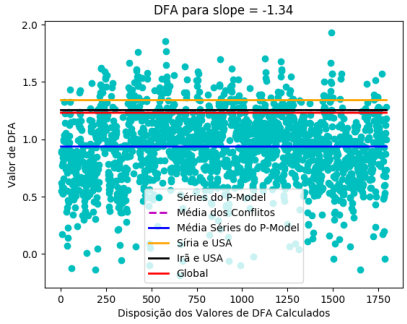
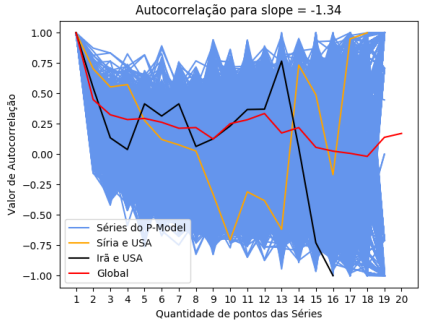
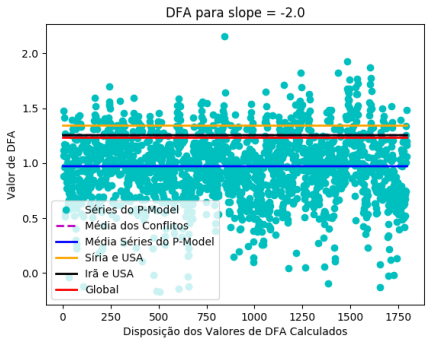
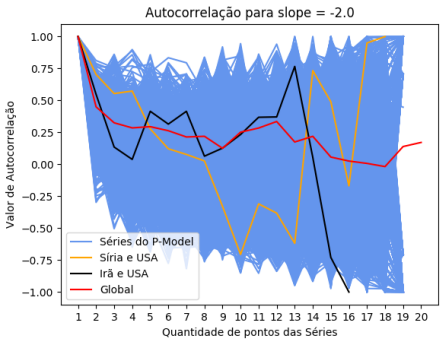
β	Gráfico DFA	Gráfico Autocorrelação
-0,5		
-0,8		
-1,0		

Tabela 5.4 - Comparação dos valores de DFA e autocorrelação das séries geradas pelo algoritmo "*P-Model*" com as séries dos conflitos armados entre a Síria e USA, Irã e USA e Global para $\beta = \{-1,34; -2,0\}$.

β	Gráfico DFA	Gráfico Autocorrelação
-1,34		
-2,0		

A análise de todos os gráficos de DFA e autocorrelação apresenta que os valores de DFA e autocorrelação dos conflitos armados, sempre se mantêm na área gerada pelos valores de DFA e autocorrelação das 1796 séries criadas pela convolução da série gerada pelo "*P-Model*". Dessa forma, como se considerou um range significativo de espectro de potência na validação, pode-se concluir que o "*P-Model*" apresenta-se como um gerador de séries temporais endógenas para qualquer valor de β ou espectro de potência. O portfólio de ameaças com o "*P-Model*" validado são utilizados para gerarem os dados de entrada do modelo de aprendizado profundo que utiliza rede neural LSTM para a predição dos eventos extremos analisados. Esse resultado é apresentado na Seção 5.3.

5.3 Resultado 3: predição dos eventos extremos sociais e naturais

O objetivo principal da tese é realizar a predição dos eventos extremos sociais e naturais, utilizando as informações dos portfólios de ameaças. A predição dos eventos extremos sociais e naturais foi dividida em duas fases: na primeira fase realizou-se a predição de uma fração da série temporal endógena para verificar o funcionamento da metodologia desenvolvida; na segunda fase realizou-se a predição do futuro dos eventos extremos para realizar uma aplicação da metodologia. Dessa forma, nesta seção, apresenta-se uma análise da predição para os dados dos conflitos armados entre a Síria e USA, Irã e USA e conflito Global para os eventos extremos sociais. Para os eventos extremos naturais, apresenta-se uma análise dos dados de ocorrência de seca, queimadas e desmatamentos para os anos de 2015, 2016, 2017, 2018, 2019 e 2020. Apresenta-se, nas Seções 5.3.1 e 5.3.2, a série temporal de ameaça, a série temporal endógena, a predição de uma fração de 20% da série temporal endógena, gerando dados para comparação com os dados restantes da série e a predição do futuro dessa série, que gera novos dados de variação de ameaça. A fração de 20% foi selecionada com base na divisão dos dados de treinamento, validação e testes de 60%, 20%, 20%, sugerida em (HAYKIN, 2008). Essa análise foi realizada para os três casos de estudo de conflito armado e para cada ano de ocorrência de seca, queimadas e desmatamento abordados, respectivamente.

5.3.1 Série temporal, série temporal endógena e predição dos eventos extremos sociais

Na predição dos eventos extremos sociais utilizam-se os dados contidos nos portfólios de ameaça para os conflitos armados. A série temporal de ameaça gerada pelo processamento do portifólio, é aplicada no cálculo do fator "*slope*" ou β, sendo uma das entradas do algoritmo "*P-Model*". Tal algoritmo gera como saída uma série temporal endógena de ameaças com 16384 elementos. O número de elementos da série foi definido considerando a fase de experimentação da tese, na qual se obteve a melhor representação da série temporal endógena para o valor selecionado. Os outros valores testados foram 1024, 2048, 4096 e 8192, que são apresentados na Seção C.1. A predição foi realizada em duas fases e ambas as fases utilizaram-se a arquitetura de aprendizado profundo com rede neural LSTM, como apresentado nas Seções 5.3.1.1 e 5.3.1.2.

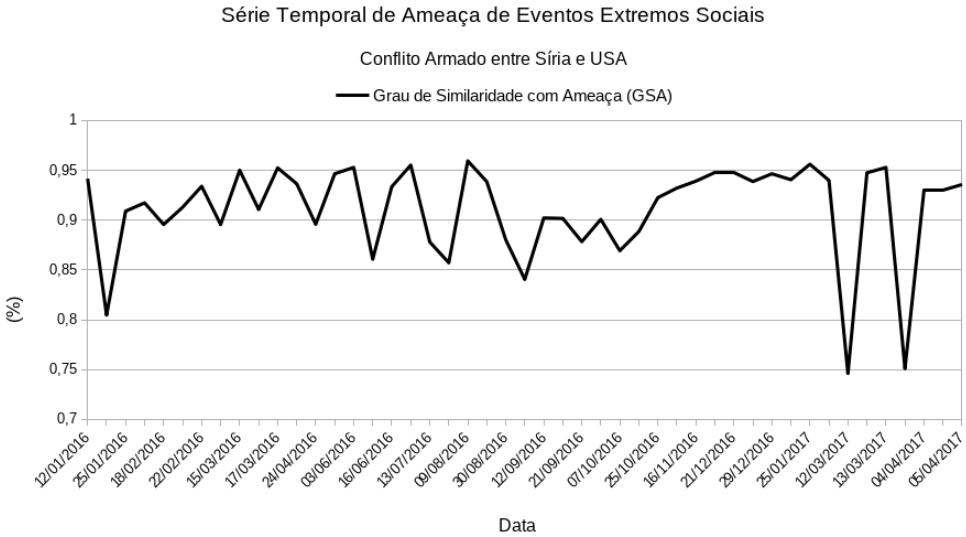
5.3.1.1 Primeira fase de predição

Na primeira fase, realizou-se a predição de uma fração de 3276 pontos da série temporal endógena, gerada pelo "*P-Model*", para verificar o funcionamento da metodologia. A seguir são apresentados, para cada conflito abordado, a série temporal de ameaça, a série temporal endógena, com uma análise, utilizando os parâmetros de Média, Variância, Curtose e Assimetria, para a verificação da representação de variação de ameaça da série endógena e a predição da fração dessa série.

a) **Conflito armado entre a Síria e USA**

A utilização do portfólio de ameaça do conflito armado entre a Síria e USA, apresentado na Seção 5.1.1.1, permitiu que se gerasse uma série temporal de ameaça com 47 elementos, como apresentado na Figura 5.10. Essa série temporal de ameaça apresenta a variação do grau de similaridade com ameaça (GSA), sendo calculado pela aplicação da técnica de análise de sentimento nas notícias coletadas das mídias sociais, como apresentada na Seção 4.3. Assim, são consideradas somente as notícias de ameaça entre os chefes dos países envolvidos, sendo coletadas no período entre janeiro de 2016 e abril de 2017, e que antecedem o evento considerado extremo, devido às suas consequências para a vida humana, referente ao lançamento dos mísseis "*Tomahawk*" ocorrido em 05/04/2017.

Figura 5.10 - Série temporal de ameaça de eventos extremos sociais referentes ao conflito armado entre a Síria e USA.



Os dados da série temporal de ameaça foram utilizados no processo de criação da série temporal endógena de ameaças por meio do algoritmo "*P-Model*", como apresentado no pseudocódigo da Seção 4.5. A Tabela 5.5 apresenta o número de elementos da série, o valor de parâmetro para geração de série endógena e o valor de β, sendo gerado pelo cálculo do valor de DFA.

Tabela 5.5 - Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena do conflito armado entre a Síria e USA.

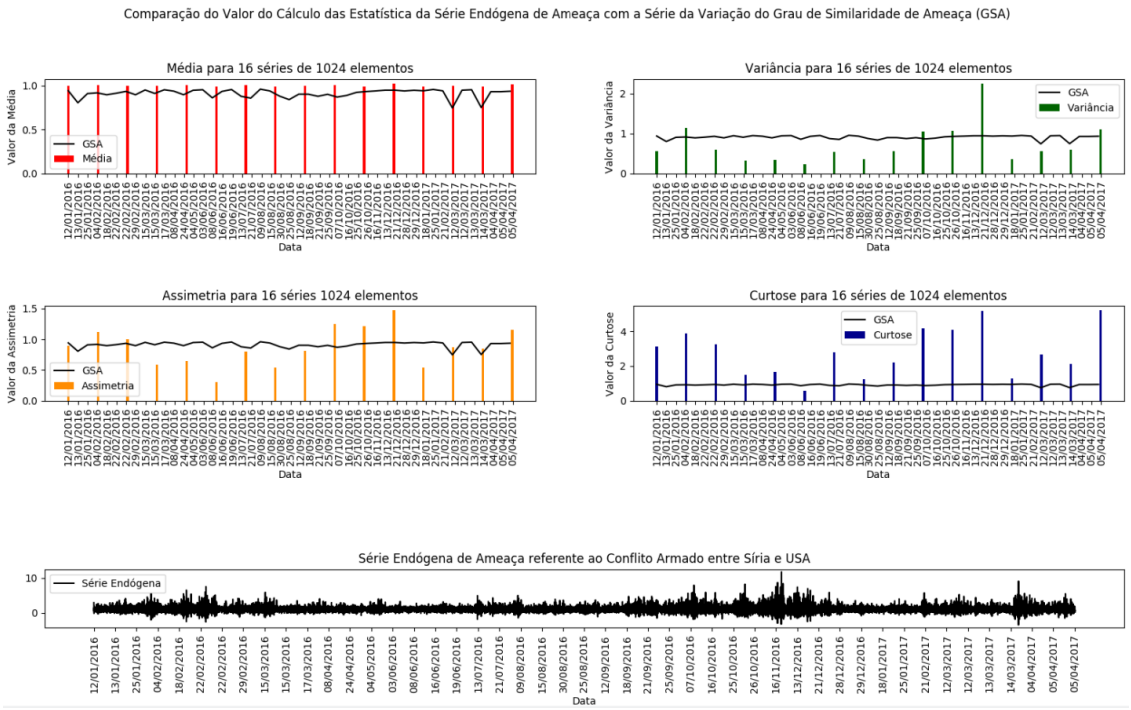
Número de Elementos da Série	p	**"*slope*"**(β)
16384	0,60	0,5606194965951611

Os dados da Tabela 5.5 geraram uma série endógena de ameaças apresentada na Figura 5.11. Com essa série, apresentam-se os gráficos dos parâmetros de média, variância, assimetria e curtose com a série temporal de ameaça, que representa a variação da similaridade de ameaça. Essa análise foi realizada com a intenção de verificar se a variação de ameaça, encontrada na série temporal de ameaça (GSA), é representada pela série endógena de ameaças e se a mesma variação de ameaça, sugere que algum evento maior ou evento extremo possa ocorrer posteriormente ao período de ameaça. Para o cálculo desses parâmetros, dividiu-se a série endógena de ameaças de 16384 elementos em 8, 16 e 32 séries de 2048, 1024 e 512 elementos, respectivamente, e calcularam-se os valores de média, variância, curtose e assimetria para cada uma das séries. Dessa análise inicial, sobre a melhor quantidade de elementos da série, verificou-se que o cálculo com 16 séries de 1024 elementos obteve uma melhor representação da análise dos parâmetros. As análises para as quantidades de 512 e 2048 elementos, são apresentadas na Seção C.1 e C.2. Pela análise dos parâmetros, observam-se os seguintes comportamentos:

- O parâmetro de Média não gerou uma representação significativa da variação de ameaça comparando com a série temporal do Grau de Similaridade com Ameaça (GSA).
- O parâmetro de Variância teve uma boa resposta na comparação da variação da série do GSA, representando bem todos os picos e vales dessa série.

- O parâmetro de Curtose também respondeu de forma adequada a mesma comparação. No entanto, representou de forma menos precisa os picos e vales da série. Porém, como os valores de curtose são maiores que 3, têm-se um excesso > 0 com caldas mais longas e grossas, chamado de curtose leptocúrtica e isso pode representar uma hipótese maior de ocorrência de algum evento maior ou evento extremo após o período de ameaças.
- O parâmetro de Assimetria apresenta todos os vales e picos com boa precisão. Embora os seus valores serem maiores que zero, verifica-se que o gráfico apresenta um calda mais deslocada para esquerda, o que também pode representar uma maior possibilidade de ocorrência de algum evento maior ou evento extremo após o período de ameaças.

Figura 5.11 - Análise do cálculo dos parâmetros média, variância, assimetria e curtose para a série temporal endógena do conflito armado entre a Síria e USA.

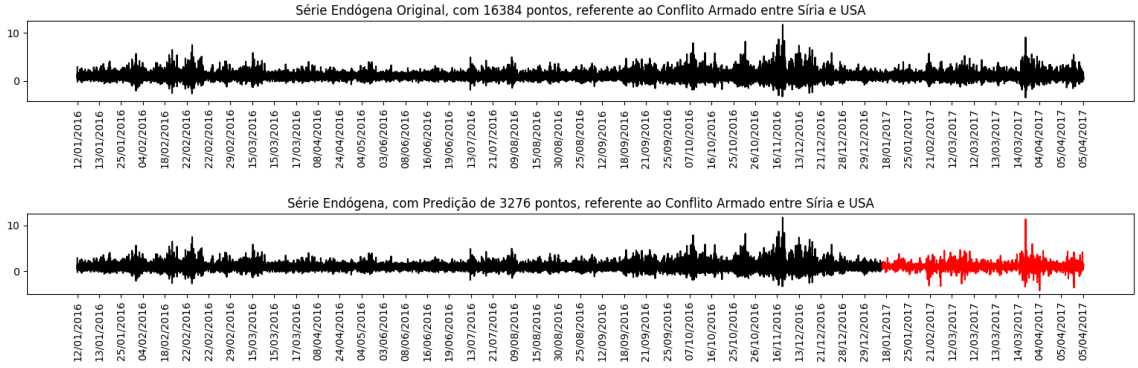


No caso do conflito armado entre a Síria e USA, verifica-se que esses resultados apresentados pelos parâmetros de curtose e assimetria, mostram que a variação de ameaça presente na série endógena, sugere o acontecimento de um evento extremo após o período de ameaça, que como mencionado anteriormente, nesse caso, refere-se ao lançamento dos mísseis "*Tomahawk*"

ocorrido em 05/04/2017. Esses resultados induzem ao entendimento, de que a série endógena de ameaças, representa de forma adequada as variações de ameaça e essas variações sugerem o acontecimento de um possível evento extremo.

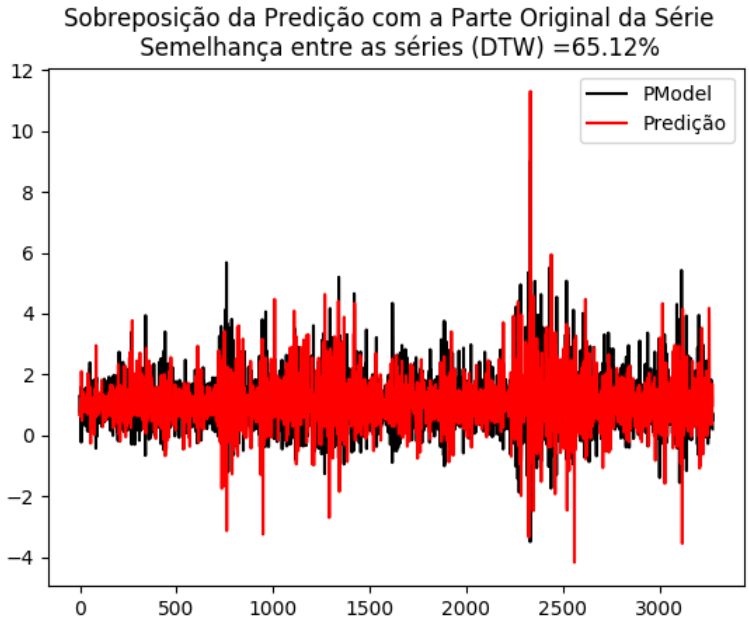
Depois da criação da série temporal endógena de ameaças, realizou-se a aplicação da arquitetura de aprendizado profundo com a rede neural LSTM, apresentada na Seção 4.6, na predição de uma fração dessa série temporal endógena para a aplicação da metodologia desenvolvida. A Figura 5.12 apresenta duas séries temporais endógenas: a primeira é a série original do "*P-Model*" que está toda na cor preta e a segunda é a série que apresenta a predição de 3276 pontos da série original na cor vermelha.

Figura 5.12 - Predição de uma fração da série temporal endógena de ameaças do conflito entre a Síria e USA no período de janeiro de 2016 e abril de 2017.



Para a verificação da precisão do resultado apresentado na Figura 5.12, gerou-se um gráfico de sobreposição dos dados preditos com os dados originais da série em conjunto com o cálculo de DTW, que representa a distância entre duas séries ou a dissimilaridade. No gráfico da Figura 5.13, com a sobreposição das séries, apresenta-se o valor de DTW subtraído de 100, para se obter o valor da similaridade entre os trechos de séries analisados.

Figura 5.13 - Sobreposição da predição da série endógena com o cálculo do valor de DTW.

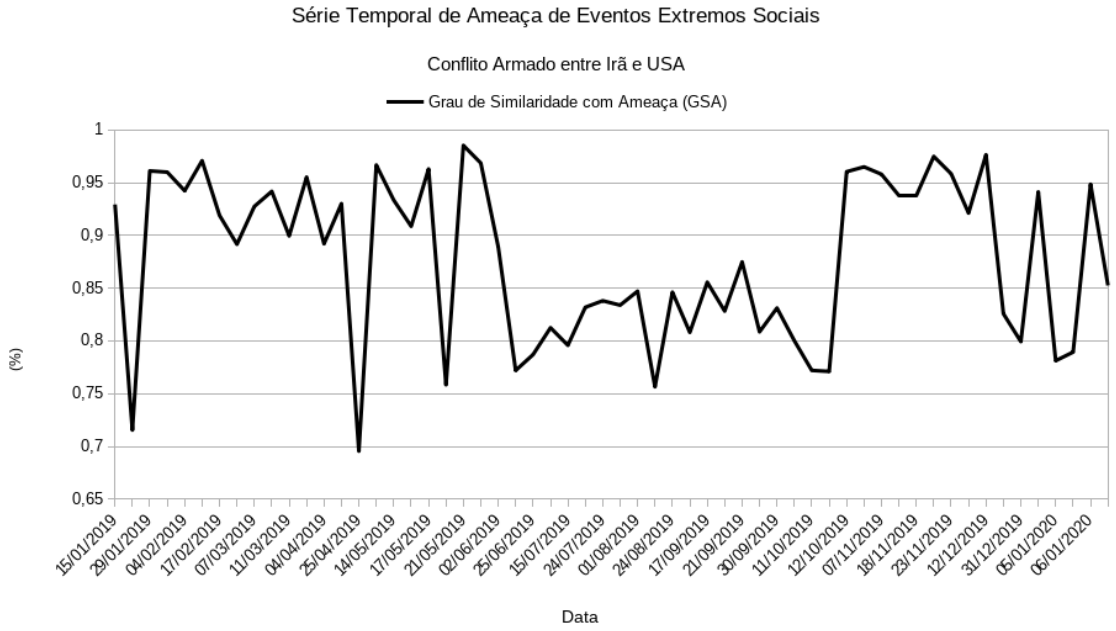


A Figura 5.13 apresenta uma similaridade de cerca de 65,12%, entre as partes das séries analisadas. Esse valor de similaridade também representa a precisão da predição, dessa série endógena, pela arquitetura de aprendizado profundo com rede LSTM desenvolvida na metodologia.

b) **Conflito armado entre Irã e USA**

Outra utilização para o portfólio de ameaça, foi a aplicação dos seus dados, apresentados na Seção 5.1.1.2, na análise do conflito armado entre Irã e USA. Esses dados, geraram uma série temporal de ameaças com 58 elementos coletados considerando o período entre janeiro de 2019 e janeiro de 2020, como apresentado na Figura 5.14.

Figura 5.14 - Série temporal de ameaças de eventos extremos sociais referentes ao conflito armado entre Irã e USA.



Da mesma forma, que na análise do conflito armado entre a Síria e USA, os dados da série temporal de ameaça para o conflito armado entre Irã e USA, foram processados pelo algoritmo "*P-Model*" para a criação da série temporal endógena de ameaças. Esse processamento utilizou os dados da Tabela 5.6, no qual o valor de "*Slope*" ou β também foi calculado utilizando o conceito de DFA.

Tabela 5.6 - Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena do conflito armado entre Irã e USA.

Número de Elementos da Série	p	**"*Slope*"(β)**
16384	0,60	0,9132939832232201

A série endógena de ameaças, resultante do processamento dos dados da Tabela 5.6, é apresentada na Figura 5.15 e da mesma forma, são apresentados os cálculos dos valores de média, variância, curtose e assimetria, para a verificação da variação de ameaça e representação, dessa informação de ameaça, de um possível acontecimento de evento maior ou evento extremo. No caso do conflito armado entre Irã e USA, analisa-se como

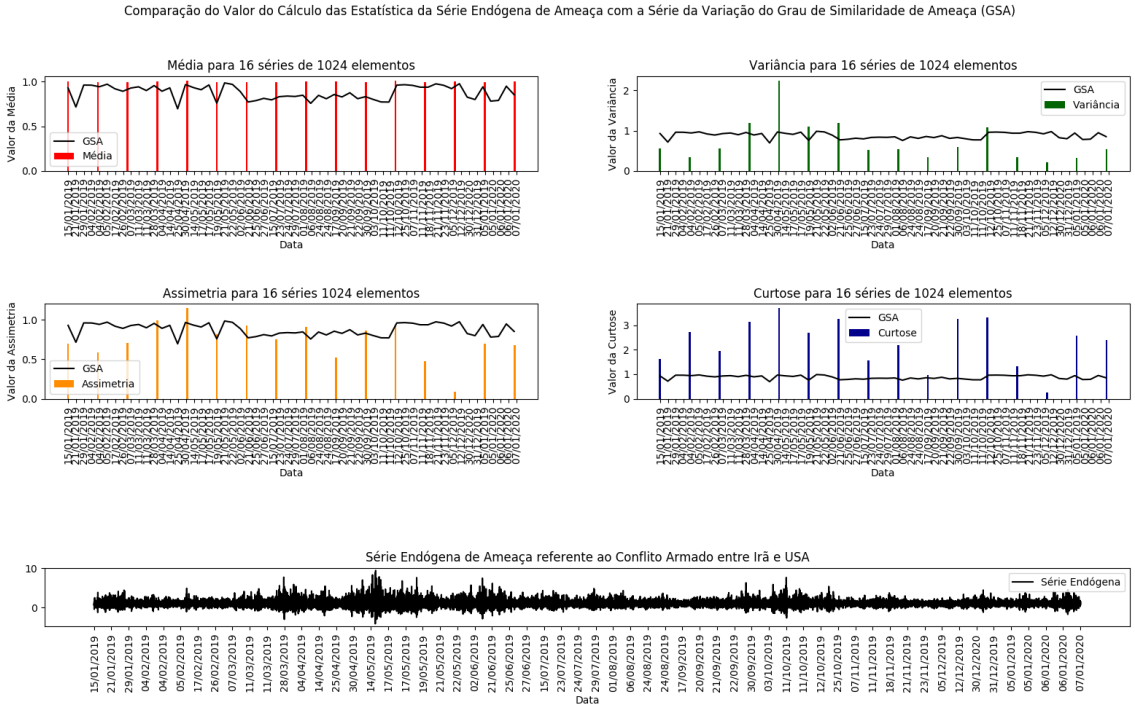
possível evento maior, o ataque realizado, pelo grupo jihadista, à base no Iraque que abrigava forças americanas, ocorrido em 07/01/2020.

A análise desses parâmetros para os dados do conflito armado entre Irã e USA, permitiu observar as seguintes características.

- O parâmetro de Média, também, não teve uma boa representação da variação de ameaça para os dados do conflito armado entre Irã e USA.
- O parâmetro de Variância, no que lhe concerne, apresentou de forma menos acentuada os picos e vales da série de ameaça.
- O parâmetro de Curtose, também apresentou de forma menos acentuada os picos e vales da série. O valor de Curtose no final da análise apresentou-se entre 2 e 3, ou seja, ≈ 3, curtose mesocúrtica. Esse resultado da curtose, apresenta-se coerente com a análise, já que espera que se ocorra após o período de ameaças analisado, um evento maior e não um evento extremo. Considerou-se o ataque jihadista como um evento maior, pois o mesmo não gerou tantos danos a vida humana quanto o evento extremo considerado no conflito armado entre a Síria e USA.
- O parâmetro de Assimetria, apresentou de uma forma um pouco mais acentuada os picos e vales da série temporal de ameaça. Os valores de assimetria apresenta-se sempre maior que zero. No entanto, as suas caldas continuam com uma leve tendência para a esquerda. Assim, da mesma forma sugere-se que algo maior que uma ameaça possa acontecer.

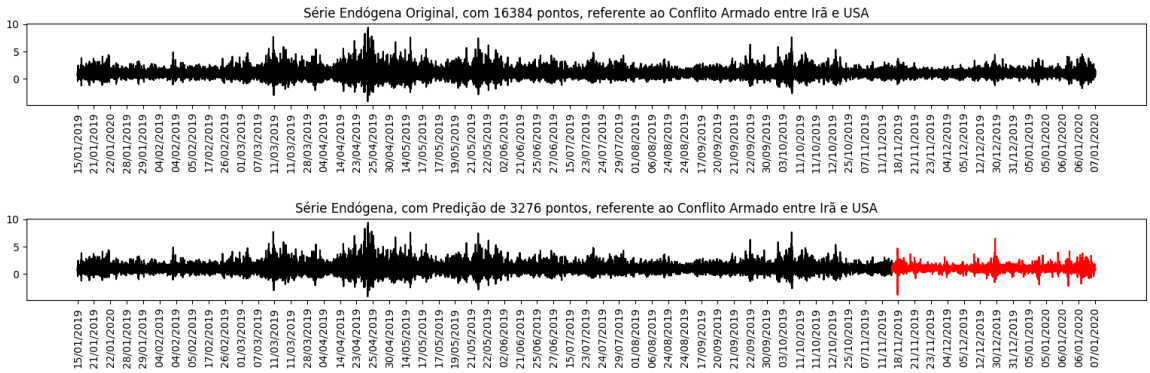
Os resultados da análise dos parâmetros de média, variância, curtose e assimetria, demonstraram que tanto a curtose quanto a assimetria sugeriram, novamente, que a série endógena faz uma boa representação da variação de ameaça e de um possível acontecimento de um evento maior, para os dados do conflito entre Irã e USA.

Figura 5.15 - Variação do cálculo dos parâmetros média, variância, curtose e assimetria para a série temporal endógena do conflito armado entre Irã e USA.



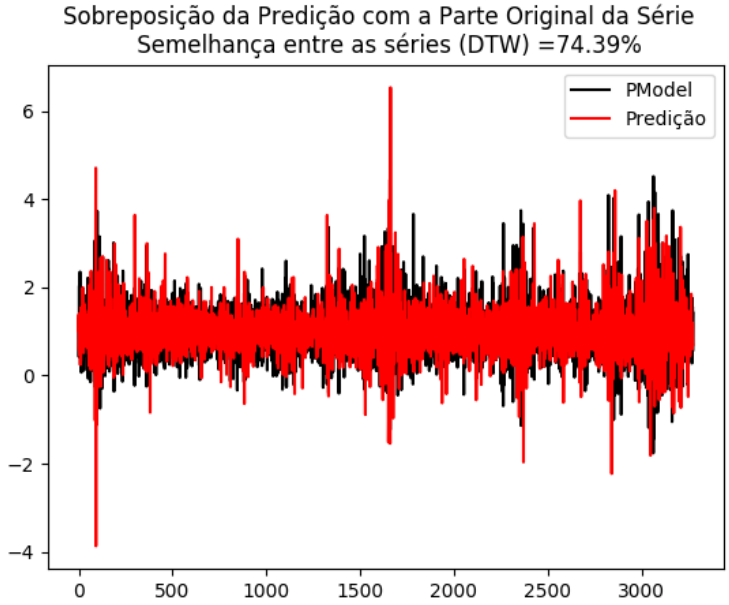
Após a análise da série endógena de ameaças, realizou-se, também por meio da aplicação da arquitetura de aprendizado profundo com rede neural LSTM, a predição de uma fração da série temporal endógena, como mais um teste para verificar o funcionamento da metodologia desenvolvida. E de forma padronizada, a Figura 5.16 apresenta duas séries temporais endógenas: a primeira é a representação do "*P-Model*" para a variação de ameaça, que está toda na cor preta e a segunda é a representação da arquitetura de aprendizado profundo, que apresenta a predição de 3276 pontos da série original na cor vermelha.

Figura 5.16 - Predição de uma fração da série temporal endógena de ameaças do conflito entre Irã e USA no período de janeiro de 2019 a janeiro de 2020.



Para a comprovação do resultado da predição da fração da série, também se gerou um gráfico, de sobreposição dos dados preditos com os dados originais da série em conjunto com o cálculo de DTW, sendo representado pelo valor de similaridade. No gráfico da Figura 5.17, com a sobreposição das séries, apresenta-se esse valor da similaridade entre os trechos de séries analisadas.

Figura 5.17 - Sobreposição da predição da série endógena com o cálculo do valor de DTW para o conflito entre Irã e USA.

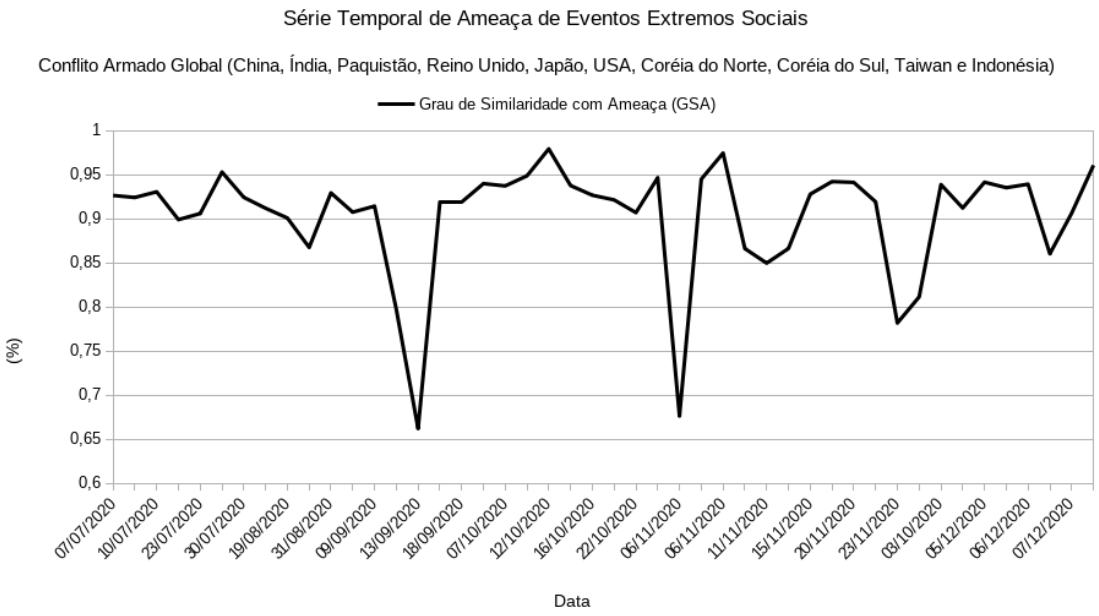


A Figura 5.17 mostra que a similaridade entre o trecho predito e o mesmo trecho original é de 74,39% e esse valor também demonstra o bom resultado da predição da série endógena de ameaças utilizando a arquitetura de aprendizado profundo com a rede LSTM, desenvolvida para a metodologia.

c) **Conflito armado Global**

Os dados para o conflito armado global foram coletados entre o período de julho de 2020 a dezembro de 2020, com a intenção de monitorar a variação de ameaças entre os chefes de estado dos países, China, Índia, Paquistão, Reino Unido, Japão, USA, Coreia do Norte, Coreia do Sul e Indonésia. Esse monitoramento se deve a movimentação que está ocorrendo no mar do Sul da China. Assim, o evento maior ou evento extremo considerado seria o aumento de ameaças e/ou um ato de ataque ou conflito entre esses países. Esses dados geraram uma série temporal de ameaça com 46 elementos, como apresentado na Figura 5.18.

Figura 5.18 - Série temporal de ameaças de eventos extremos sociais referentes ao conflito armado global.



Dando continuidade na verificação do funcionamento da metodologia desenvolvida para a predição das séries temporais endógenas, utilizaram-se os dados dessa série temporal de ameaça para realizar o cálculo do valor de DFA, considerado como valor de β na tese. Esse valor de DFA, com

o número de elementos da série e do parâmetro p, são apresentados na Tabela 5.7.

Tabela 5.7 - Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena do conflito armado global.

Número de Elementos da Série	p	"***Slope***"(β)
16384	0,60	1,2216174784781086

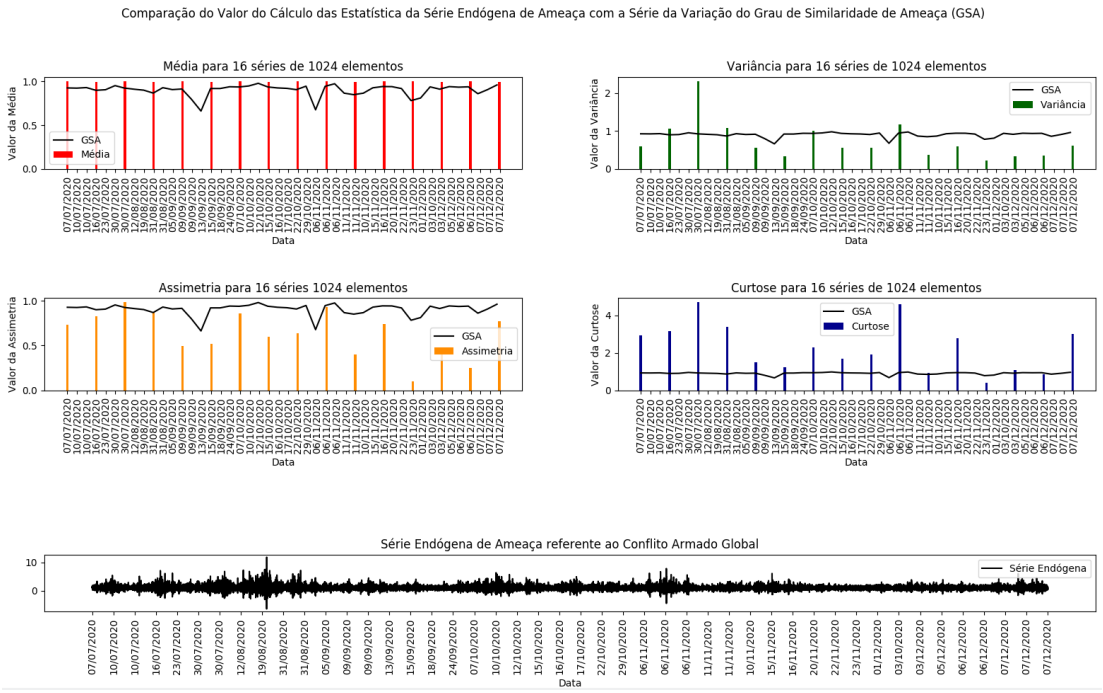
O resultado do processamento dos dados apresentados na Tabela 5.7 é uma série endógena de ameaças, apresentada na Figura 5.19. Essa série endógena de ameaças representa a variação de ameaça contida na série gerada pela aplicação da técnica de análise de sentimento nas notícias coletadas das mídias sociais. Essa variação de ameaça, novamente foi analisada considerando os parâmetros de média, variância, curtose e assimetria. Dessa análise, observou as seguintes características.

- O parâmetro de Média, também nesse caso, não demonstrou uma representação satisfatória dos picos e vales da variação de ameaça.
- O parâmetro de Variância conseguiu representar de maneira satisfatória as varições de picos e vales da série de ameaça.
- O parâmetro de Curtose apresentou uma representação considerável dos picos e vales. Os valores de curtose, no final da série, apresentaram-se maiores que 3 ou excesso > 0, indicando também uma curtose leptocúrtica. Assim, as caldas se mantiveram a esquerda e mais longas, o que pode representar um possível acontecimento de um evento maior ou de um evento extremo.
- O parâmetro de Assimetria, para esse caso, também apresentou uma boa representação das variações dos picos e vales da série de ameaça. No entanto, os seus valores mantiveram-se maiores que zero, mas suas caldas demonstraram ter uma leve inclinação para a esquerda. Esse conjunto de características, também pode representar uma possível ocorrência de algum evento maior que uma ameaça.

A análise dos parâmetros de média, variância, curtose e assimetria, também demonstraram, considerando as características de curtose e assimetria apresentadas, que a série endógena de ameaças representa de forma ade-

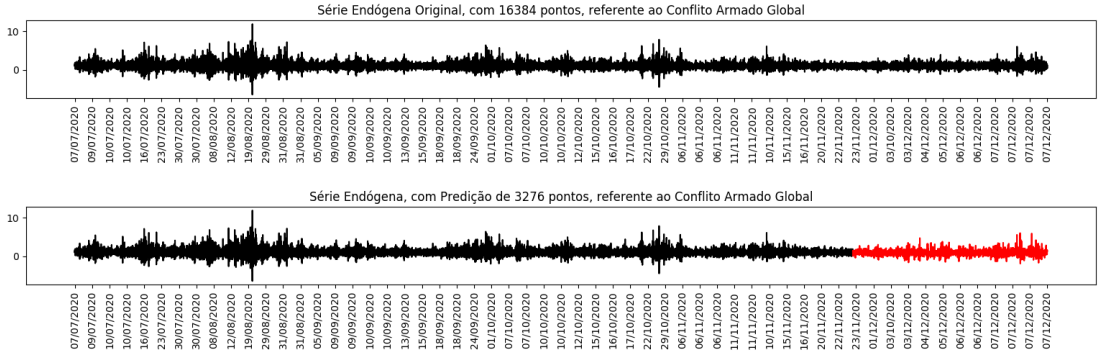
quada a variação de ameaça da série de ameaça e essa variação também pode representar a ocorrência de algum evento maior.

Figura 5.19 - Variação do cálculo dos parâmetros média, variância, curtose e assimetria para a série temporal endógena do conflito armado Global.



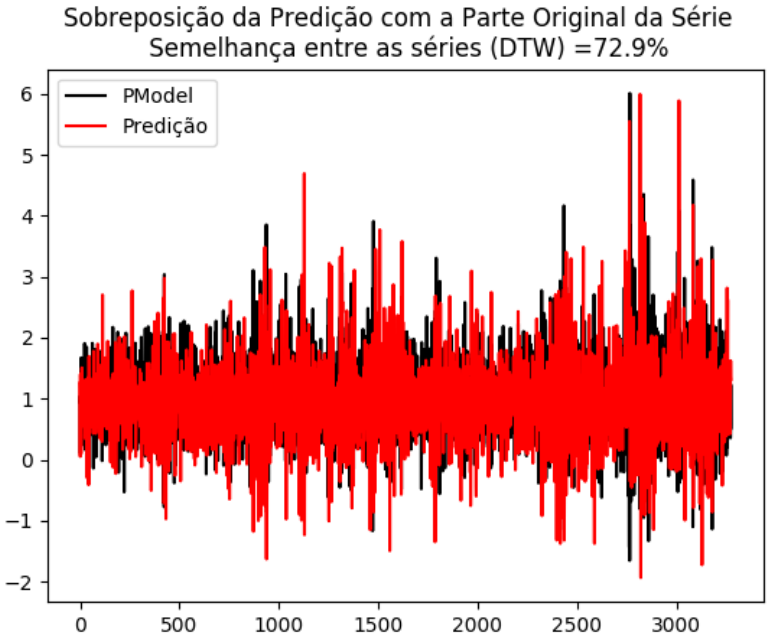
Depois da análise da série temporal endógena de ameaças, realizou-se pela aplicação da mesma arquitetura de aprendizado profundo com rede neural LSTM, utilizada anteriormente, a predição de uma fração da série temporal endógena. A Figura 5.20 apresenta duas séries temporais endógenas: a primeira é a representação do "*P-Model*" da série de ameaça e está toda na cor preta e a segunda é a representação da predição, pela arquitetura de aprendizado profundo, de 3276 pontos da série original na cor vermelha.

Figura 5.20 - Predição de uma fração da série temporal endógena de ameaças do conflito Global no período de julho de 2020 e dezembro de 2020.



A verificação da precisão da predição da fração da série endógena de ameaças, foi realizada gerando um gráfico de sobreposição dos dados preditos com os dados originais da série em conjunto com o valor de similaridade entre essas séries (DTW). No gráfico da Figura 5.21, com a sobreposição das séries, apresenta-se o valor da similaridade entre os trechos de séries analisados.

Figura 5.21 - Sobreposição da predição da série endógena com o cálculo do valor de Similaridade para o conflito armado Global.



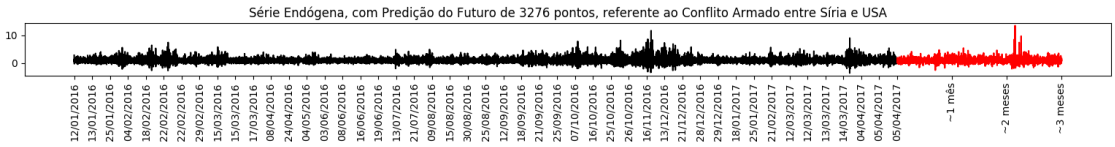
A Figura 5.21 mostra que a sobreposição de uma fração da série apresenta uma similaridade entre às duas representações de 74,39%. Esse resultado, também representa, o bom funcionamento da metodologia e arquitetura de aprendizado profundo com rede LSTM desenvolvida na tese.

5.3.1.2 Segunda fase de predição

Na segunda fase, realizou-se a predição do futuro das séries temporais de ameaças dos conflitos armados entre a Síria e USA, Irã e USA e Global, para realizar uma aplicação da metodologia da tese, como apresentada nas Seções a, b e c. Essa predição utilizou a mesma arquitetura de aprendizado profundo usada na predição das frações das séries endógenas dos conflitos analisados. A diferença, consiste na aplicação de uma estrutura de repetição que em cada novo conjunto de valores preditos, realimentava a rede. Esse processo se repetiu até se obter a predição de 3276 elementos ou 20% do número de dados total da série de entrada. Essa quantidade de pontos representa um período aproximado de tempo de 3 meses.

a) **Predição do futuro para o conflito armado entre a Síria e USA**
A primeira aplicação realizada da metodologia da tese, foi a predição do futuro da série temporal endógena de ameaças do conflito armado entre a Síria e USA. Essa predição utilizou como dado de entrada da rede a série endógena de ameaças referente a esse conflito. A Figura 5.22, mostra o resultado, na cor vermelha, da predição do futuro de cerca de 3 meses.

Figura 5.22 - Predição do futuro da série temporal endógena de ameaças do conflito entre a Síria e USA.

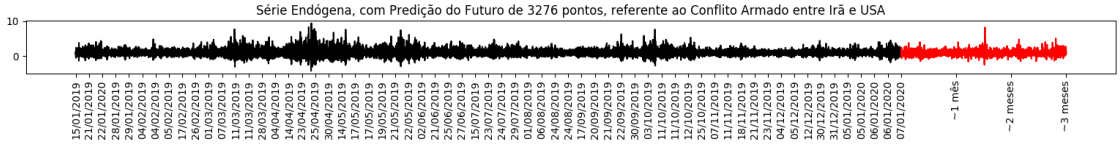


A Figura 5.22 apresenta que as variações de ameaças futuras continuam na mesma intensidade, ao longo do período predito. Considerando que o conflito armado entre a Síria e USA teve uma continuidade nas ameaças (DEJEVSKY, 2017), que perdura até agora, o resultado sugere que a arquitetura de aprendizado profundo fez uma boa representação do futuro desse conflito.

b) **Predição do futuro para o conflito armado entre Irã e USA**

Outra aplicação da metodologia utilizada, verifica-se na predição do futuro da variação de ameaças do conflito armado entre Irã e USA. Da mesma forma, utilizou-se como entrada a série temporal endógena de ameaças desse conflito e foram preditos cerca de 3 meses de futuro para esse conflito, representados na cor vermelha.

Figura 5.23 - Predição do futuro da série temporal endógena de ameaças do conflito entre Irã e USA.

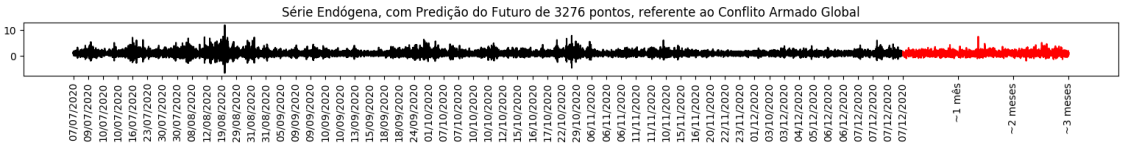


A análise da predição da Figura 5.23, apresenta que as ameaças continuam entre os dois países envolvidos no conflito. Considerando que as ameaças entre os chefes de estado desses países continuou, após o período analisado (MARCUS, 2020), verifica-se uma boa resposta da metodologia e arquitetura de aprendizado profundo, para a predição do futuro desse conflito.

c) **Predição do futuro para o conflito armado global**

Uma terceira aplicação da metodologia na área de eventos extremos sociais, foi a predição do futuro dos dados do conflito armado Global. Da mesma forma, para a realização da predição de cerca de 3 meses, utilizou-se a série endógena de ameaças desse conflito, como entrada da rede neural que utiliza aprendizado profundo. A Figura 5.24, mostra o resultado dessa predição do futuro para esse conflito, representada na cor vermelha.

Figura 5.24 - Predição do futuro da série temporal endógena de ameaças do conflito entre Global e USA.



A predição do futuro para o conflito global, apresentou uma instabilidade na série endógena, como observado no fim do trecho em vermelho da Fi-

gura 5.24. Essa instabilidade pode sugerir, que pelo menos o aumento das ameaças poderão ocorrer durante o período predito (REUTERS, 2021).

5.3.2 Série temporal, série temporal endógena e predição dos eventos extremos naturais

Os dados sobre os eventos extremos naturais são utilizados com a intenção de verificar a funcionalidade e fazer uma aplicação da metodologia, com um conjunto de dados fora do domínio de conflitos armados. Assim, pode-se verificar o desempenho da metodologia desenvolvida em conjuntos de dados variados. Da mesma forma, que na predição dos eventos extremos sociais, essa predição foi realizada em duas fases: na primeira fase apresenta-se a predição de uma fração da representação do "*P-Model*" da série de ameaças e na segunda fase apresenta-se a predição do futuro, como uma representação da arquitetura de aprendizado profundo da série de ameaças. Ambas as fases foram geradas para cada ano de dado de ocorrência de seca, queimadas e desmatamentos, ou seja, foram criados 6 séries temporais para os anos de 2015 a 2020. Esses resultados são apresentados nas Seções 5.3.2.1 e 5.3.2.2, respectivamente.

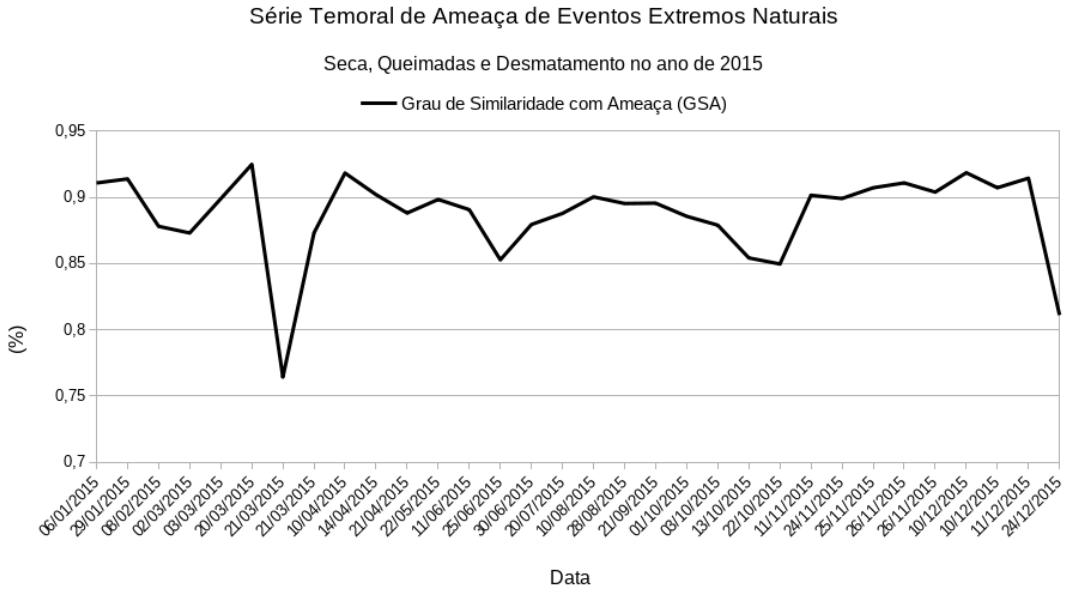
5.3.2.1 Primeira fase de predição

A primeira fase da predição apresenta como resultado a representação da arquitetura de aprendizado profundo com rede neural LSTM, para um trecho da série endógena de ameaças. Esse resultado é comparado com a representação, do algoritmo "*P-Model*", do mesmo trecho da série endógena de ameaças. Esses dois processos foram realizados para os dados de ocorrência de seca, queimadas e desmatamentos na região da Floresta Amazônica para os anos de 2015, 2016, 2017, 2018, 2019 e 2020.

a) **Ocorrência de seca, queimadas e desmatamentos para o ano de 2015**

A análise dos dados de ocorrência de seca, queimadas e desmatamentos foi realizada utilizando as informações contidas no portfólio de ameaças da Seção 5.1.2.1. Para esses dados, gerou-se uma série temporal de ameaças com 32 elementos com dados coletados no período de janeiro de 2015 e dezembro de 2015, como apresentado na Figura 5.25.

Figura 5.25 - Série temporal de ameaças de eventos extremos naturais referentes às notícias sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2015.



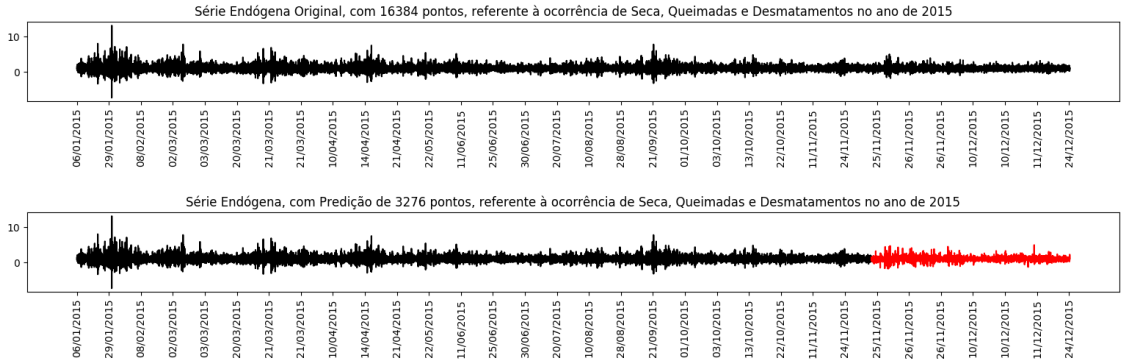
O processamento dos dados da série temporal de ameaças da Figura 5.25 também foi realizado pelo cálculo do valor "*slope* (β)", pela função de DFA. Esse cálculo com os valores do parâmetro p e do número de elementos da série endógena são fornecidos como entrada do algoritmo "*P-Model*". Os valores para cada um desses parâmetros é apresentado na Tabela 5.8.

Tabela 5.8 - Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2015.

Número de Elementos da Série	p	"***Slope***(β)"
16384	0,60	1,3082499264176222

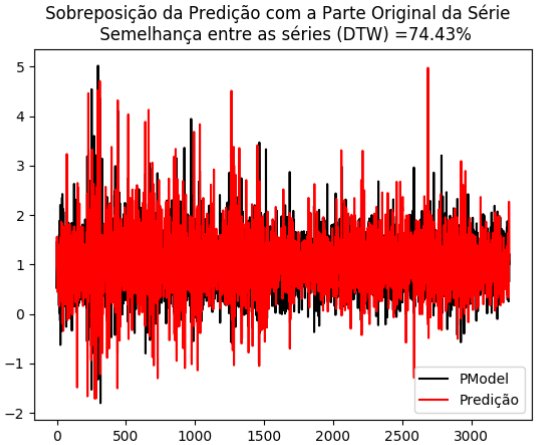
O resultado do processamento do algoritmo "*P-Model*", gerou uma série temporal endógena de ameaças, representada pela série endógena original na Figura 5.26. Nessa figura, também apresenta-se a série endógena com predição de 3276 elementos da série temporal endógena, representada na cor vermelha. Para alcançar esse resultado, utilizou-se a mesma arquitetura de aprendizado profundo com rede neural LSTM, apresentada anteriormente.

Figura 5.26 - Predição de uma fração da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para ano de 2015.



A comprovação do resultado dessa predição, também foi realizada fazendo a sobreposição dos dados preditos, que são a representação da arquitetura de aprendizado profundo da variação de ameaças, com os dados originais da série, que são a representação do *P-Model* da mesma variação de ameaça. Essa sobreposição é apresentada em conjunto com o valor de similaridade (semelhança) entre as representações dos trechos de série, sendo calculada utilizando o conceito de DTW, como realizado anteriormente. Essa comparação é apresentada no gráfico da Figura 5.27.

Figura 5.27 - Comparação da predição do trecho da série endógena com o trecho original pelo valor de similaridade para o ano de 2015.



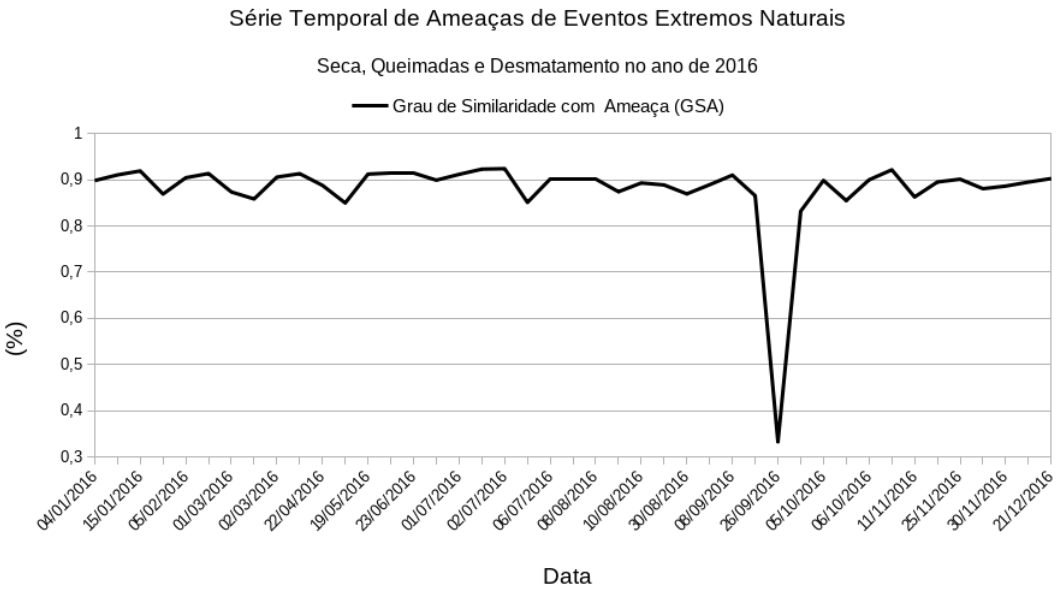
A Figura 5.27 apresenta uma similaridade de cerca de 74,43%. O bom re-

sultado da comparação dos trechos de séries, também pode ser considerado, como uma boa resposta de precisão da predição, utilizando a arquitetura de aprendizado profundo com rede neural LSTM, desenvolvida na metodologia da tese.

b) **Ocorrência de seca, queimadas e desmatamentos para o ano de 2016**

O portfólio de ameaças, também, foi utilizado na análise da ocorrência de seca, queimadas e desmatamentos para o ano de 2016. Esses dados coletados no período de janeiro de 2016 e dezembro de 2016, são apresentados na Seção 5.1.2.2. Essas informações geraram a série temporal de ameaças, apresentada na Figura 5.28.

Figura 5.28 - Série temporal de ameaças de eventos extremos naturais referentes às notícias sobre ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2016.



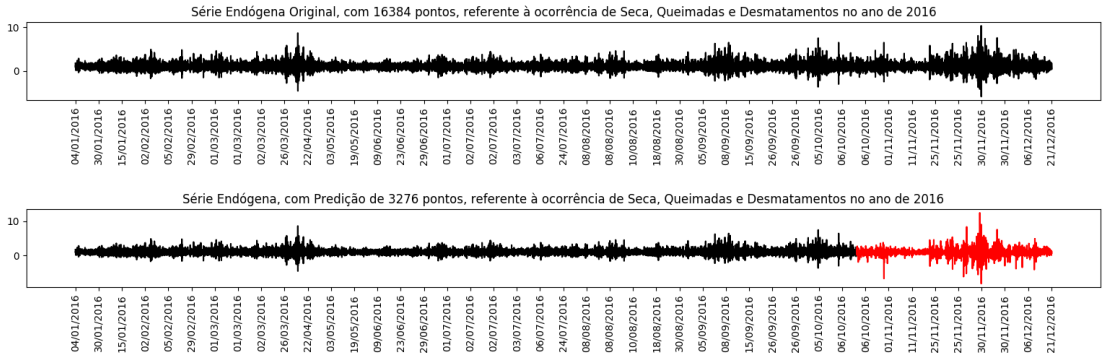
Da mesma forma, esses dados, da série temporal de ameaças, foram processados pela função de DFA, gerando, além do valor de "*slope*", as informações apresentadas na Tabela 5.9.

Tabela 5.9 - Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2016.

Número de Elementos da Série	p	"***Slope***(β)"
16384	0,60	1,3905561526142236

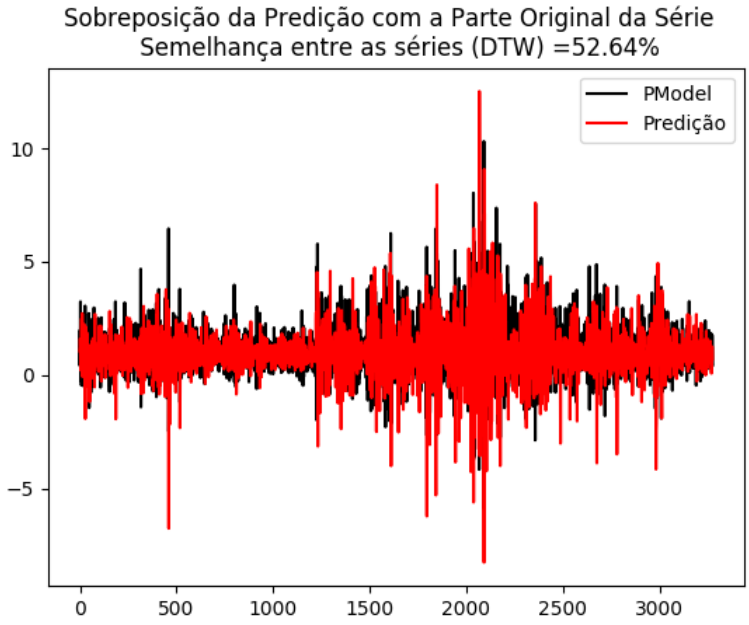
Igualmente, o processamento das informações da Tabela 5.9 pelo "*P-Model*", gerou a série endógena de ameaças apresentada na Figura 5.29, que também mostra a predição de uma fração dessa série temporal endógena, com 3276 pontos, na cor vermelha.

Figura 5.29 - Predição de uma fração da série temporal endógena sobre ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2016.



De modo igual, a comprovação do resultado da predição é realizada pela comparação dos dois trechos de série, visualmente e pelo cálculo da similaridade, considerando o cálculo de DTW. No gráfico da Figura 5.30, apresenta-se a comparação desses trechos.

Figura 5.30 - Comparação da predição do trecho da série endógena com o trecho original pelo valor de similaridade para o ano de 2016.

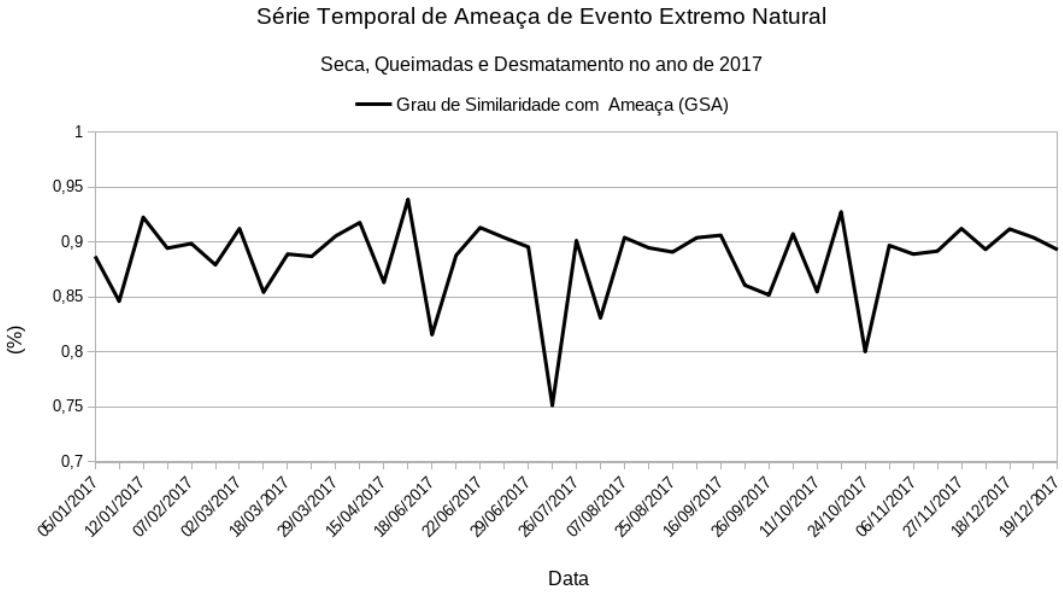


Novamente, chama-se a atenção para o bom resultado do valor da similaridade de 52,64%, entre os trechos de séries, que representa a precisão da arquitetura de aprendizado profundo utilizada na metodologia.

c) **Ocorrência de seca, queimadas e desmatamentos para o ano de 2017**

Os dados do portfólio de ameaças de ocorrência de seca, queimadas e desmatamentos para o ano de 2017, apresentados na Seção 5.1.2.3, foram processados de forma igual aos anos de 2015 e 2016, apresentados anteriormente. Assim, esses dados, coletados no período de janeiro de 2017 e dezembro de 2017, geraram a série temporal de ameaças com 41 elementos, apresentada na Figura 5.31.

Figura 5.31 - Série temporal de ameaça de eventos extremos naturais referentes às notícias sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2017.



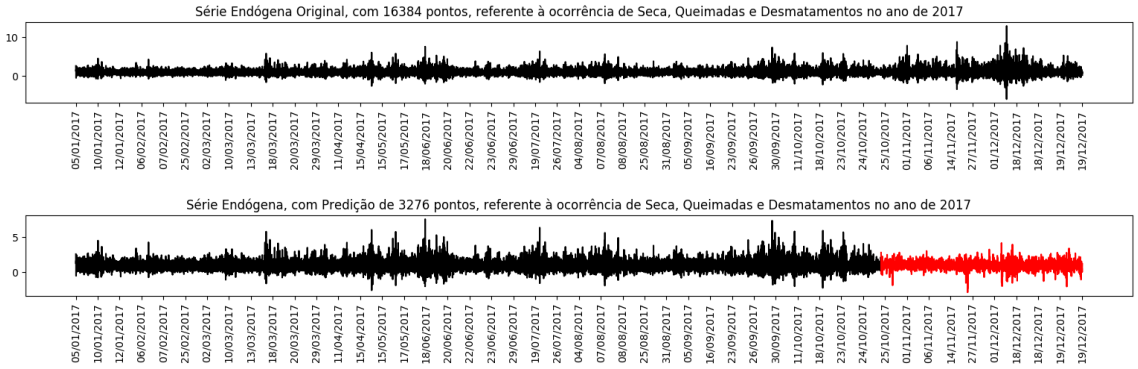
O tratamento dessas informações, por sua vez, geraram as informações da Tabela 5.10.

Tabela 5.10 - Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena sobre ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2017.

Número de Elementos da Série	p	"***Slope***(β)"
16384	0,60	0,6792655564992033

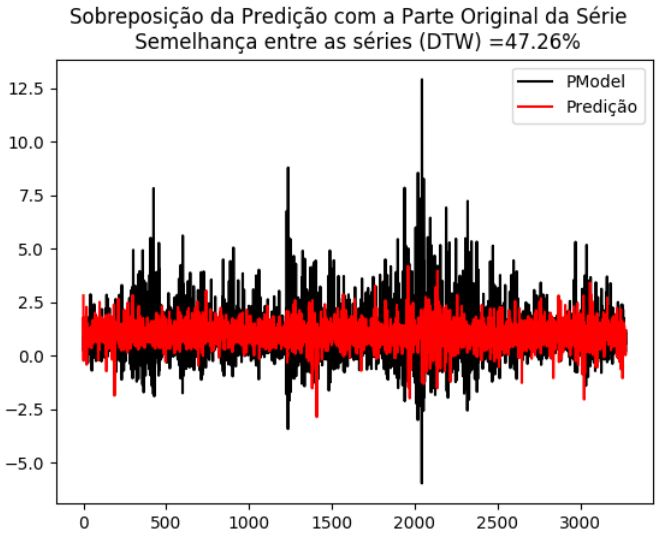
Já o processamento dessas informações, pelo "*P-Model*", geraram a série temporal endógena de ameaças apresentada na Figura 5.32. Na mesma figura, apresenta-se uma segunda série com a predição de 3276 pontos da série original, na cor vermelha.

Figura 5.32 - Predição de uma fração da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2017.



A comprovação do resultado apresentado pela predição, também, foi realizado pela comparação dos dois trechos de séries: o predito, que é a representação da arquitetura de aprendizado profundo, e o original, que é a representação do "*P-Model*" da variação de ameaças. Essa comparação resultou no gráfico de sobreposição dos dados preditos com os dados originais em conjunto com o cálculo de DTW, representado pelo valor de similaridade, definido como semelhança, entre os trechos de séries. A Figura 5.33, apresenta os resultados descritos.

Figura 5.33 - Comparação da predição do trecho da série endógena com o trecho original pelo valor de similaridade para o ano de 2017.

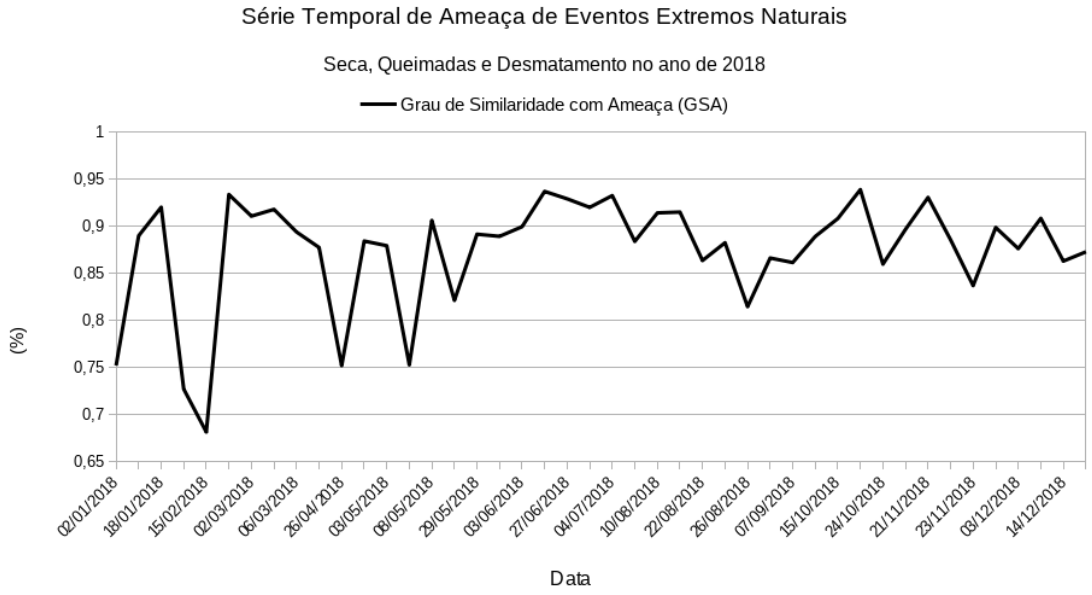


A similaridade alcançada pela análise dos dados do ano de 2017, também geraram resultados satisfatórios de 47,26%, o que apresenta novamente um desempenho satisfatório da metodologia e arquitetura de aprendizado profundo desenvolvida na tese.

d) **Ocorrência de seca, queimadas e desmatamentos para o ano de 2018**

Da mesma forma, os dados dos portfólio de ameaças, apresentados na Seção 5.1.2.4, referentes à ocorrência de seca, queimadas e desmatamentos para o ano de 2018, foram utilizados para gerar uma série temporal de ameaça com 44 elementos, como apresentada na Figura 5.34.

Figura 5.34 - Série temporal de ameaça de eventos extremos naturais referentes às notícias sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2018.



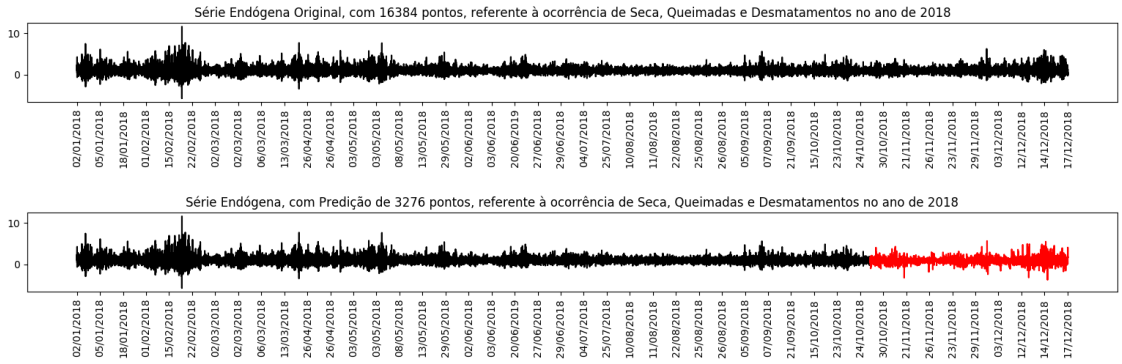
Os dados da série também foram utilizados no cálculo do valor de DFA, representado pelo valor de "*slope*", sendo apresentado na Tabela 5.11 , com as informações de número de elementos da série e valor do parâmetro p.

Tabela 5.11 - Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2018.

Número de Elementos da Série	p	"***Slope***(β)"
16384	0,60	0,8709826305556972

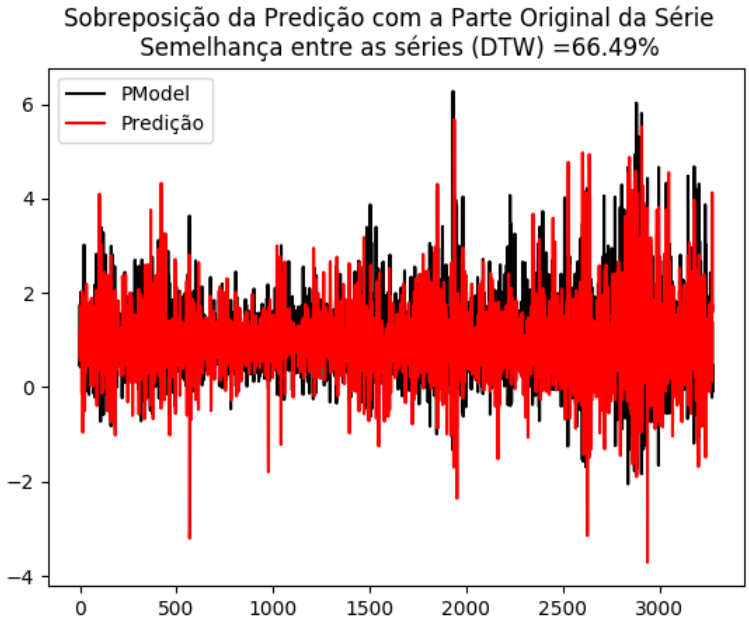
O processamento dessas informações, pelo algoritmo "*P-Model*", também, resultou em uma série temporal endógena de ameaças, apresentada na Figura 5.35. Da mesma maneira, com essa série, apresenta-se uma segunda série que mostra a predição dos 3276 pontos da série original, na cor vermelha.

Figura 5.35 - Predição de uma fração da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2018.



A comprovação do resultado dessa predição, foi realizado pelo cálculo da similaridade utilizando os conceitos de DTW. Esse resultado é apresentado no gráfico da Figura 5.36, com a sobreposição dos trechos de séries analisados.

Figura 5.36 - Comparação da predição do trecho da série endógena com o trecho original pelo valor de similaridade para o ano de 2018.

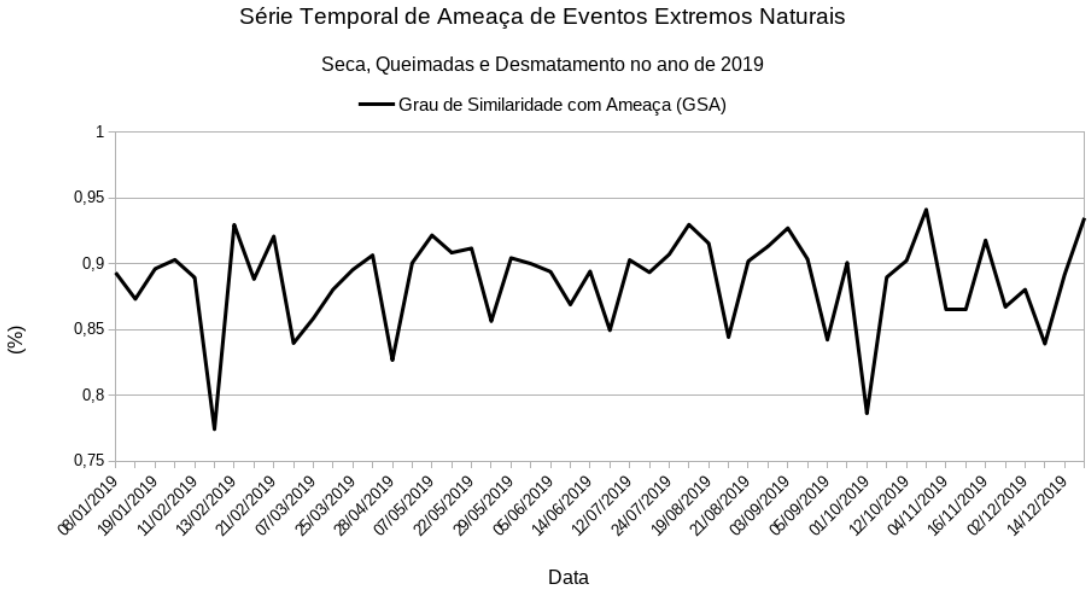


O resultado da comparação dos trechos de séries, apresentou uma similaridade bastante considerável de 66,49%, demostrando o bom resultado da metodologia de predição desenvolvida na tese.

e) **Ocorrência de seca, queimadas e desmatamentos para o ano de 2019**

A análise dos dados de ocorrência de seca, queimadas e desmatamentos para o ano de 2019, apresentados no portfólio de ameaças na Seção 5.1.2.5, foi realizada seguindo as mesmas etapas da metodologia desenvolvida e aplicada para alcançar os resultados apresentados na análise do ano de 2015, 2016, 2017 e 2018. Assim, a Figura 5.37, mostra a série temporal de ameaça com 50 elementos, gerada com esses dados.

Figura 5.37 - Série temporal de ameaça de eventos extremos naturais referentes às notícias sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2019.



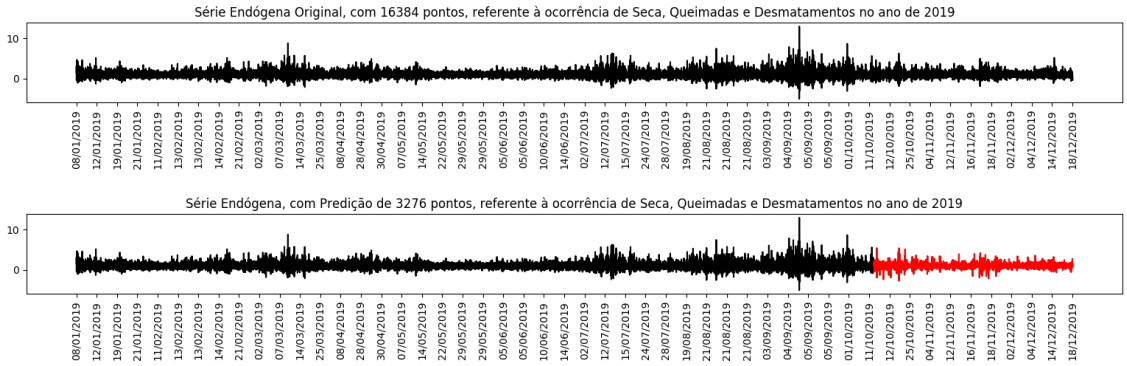
O tratamento desses dados da série, também gerou os valores de "*slope*", parâmetro p e número de elementos para a geração da série endógena de ameaças pelo algoritmo "*P-Model*", sendo apresentados na Tabela da Tabela 5.12.

Tabela 5.12 - Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2019.

Número de Elementos da Série	p	"***Slope***(β)"
16384	0,60	0,6708664045390897

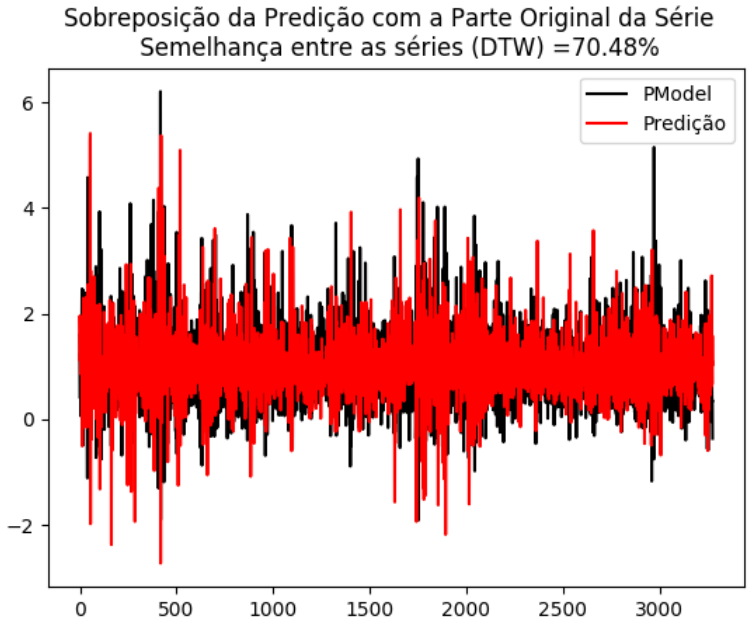
A série temporal endógena gerada por esse processamento do algoritmo "*P-Model*", é apresenta na Figura 5.38, com predição de 3276 pontos da série original, apresentada na cor vermelha.

Figura 5.38 - Predição de uma fração da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2019.



A comprovação do resultado dessa predição, também, foi realizada por meio da comparação dos dois trechos de séries analisados, o trecho da série original do *P-Model* e o trecho predito pela arquitetura de aprendizado profundo com rede neural LSTM. Essa comparação é apresentada no gráfico da Figura 5.39, de forma visual e também pelo valor de similaridade, calculado utilizando o conceito de DTW.

Figura 5.39 - Comparação da predição do trecho da série endógena com o trecho original pelo valor de similaridade para o ano de 2019.

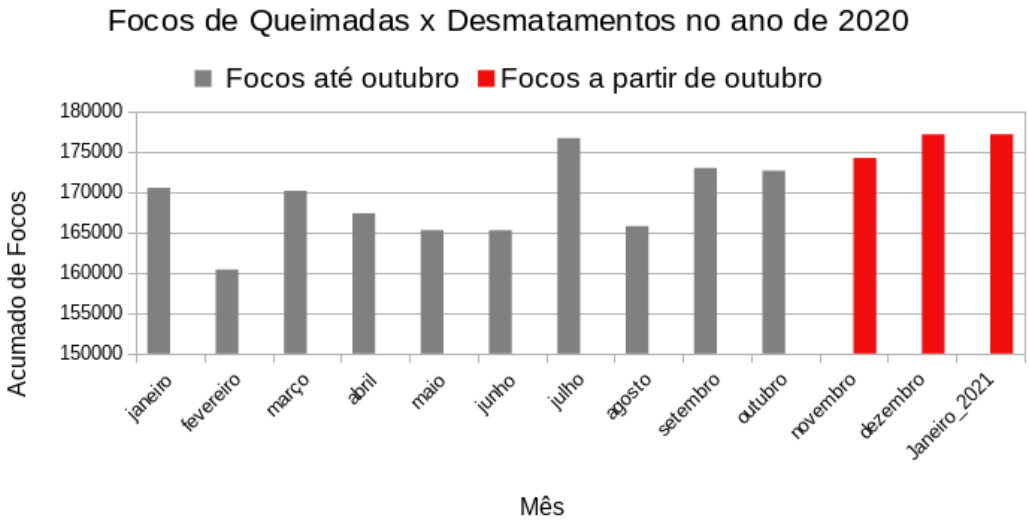


Para o ano de 2019, obteve-se o valor de similaridade de 70,48%, que continua sugerindo a boa precisão da arquitetura de rede utilizada e o bom funcionamento da metodologia da tese.

f) **Ocorrência de seca, queimadas e desmatamentos para o ano de 2020**

A ocorrência de seca, queimadas e desmatamentos para o ano de 2020, foi o último ano de dado analisado para os eventos extremos naturais. A análise desse modelo de dados, considerou como evento maior ou evento extremo, a continuidade da variação de ameaças. Para os anos de 2015, 2016, 2017, 2018 e 2019, essa variação é verifica pelos dados do ano sucessor. No entanto, para o ano de 2020, conseguiu-se dados de fontes confiáveis, como o portal do programa TerraBrasilis, do INPE (INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE, 2019), com informações de foco de queimadas e desmatamentos. Essas informações de foco são apresentadas na Figura 5.40.

Figura 5.40 - Focus de queimadas x desmatamentos para o ano de 2020.

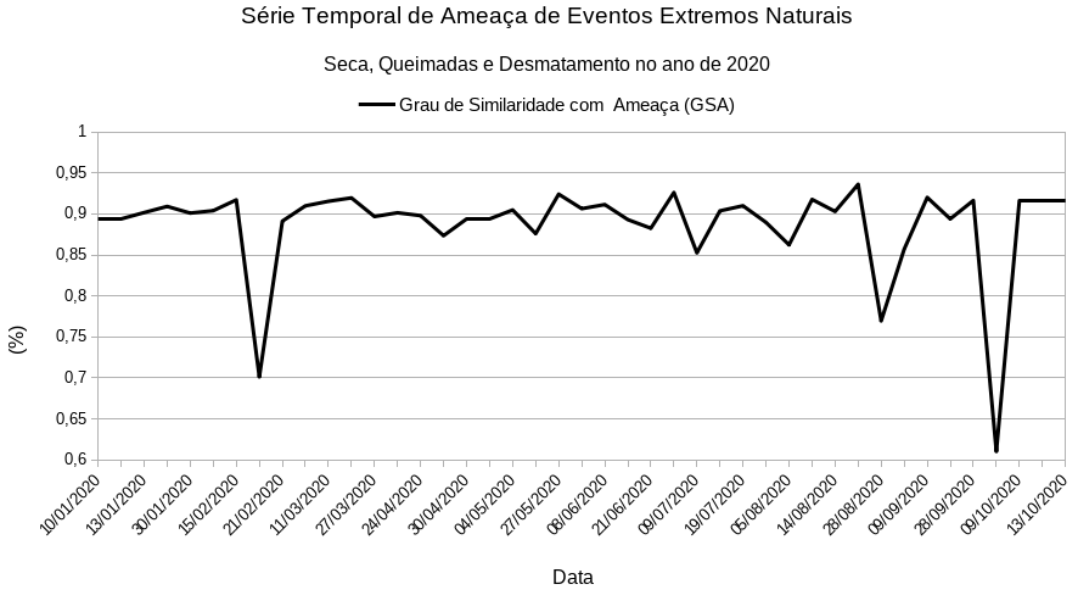


Fonte: Adaptado de INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE (2019)

A análise dos dados de 2020, seguiram as mesmas etapas consideradas na análise dos dados de eventos extremos sociais para as informações dos

conflitos armados. Assim, as informações contidas no portfólio de ameaças da Seção 5.1.2.6, geraram uma série temporal de ameaça com 43 elementos, como apresentada na Figura 5.41.

Figura 5.41 - Série temporal de ameaça de eventos extremos naturais referentes às notícias sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2020.



Da mesma forma, os dados da série temporal de ameaça foram utilizados para a geração do fator de "*slope* (β)", pelo cálculo do valor de DFA. A Tabela 5.13 apresenta o valor de DFA, como os valores de quantidade de elementos da série endógena a ser gerada e o parâmetro p, que define o modelo da série a ser gerada pelo algoritmo "*P-Model*".

Tabela 5.13 - Valores de entrada utilizados no algoritmo "*P-Model*" para a criação da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2020.

Número de Elementos da Série	p	"***Slope***(β)"
16384	0,60	0,5885749951561488

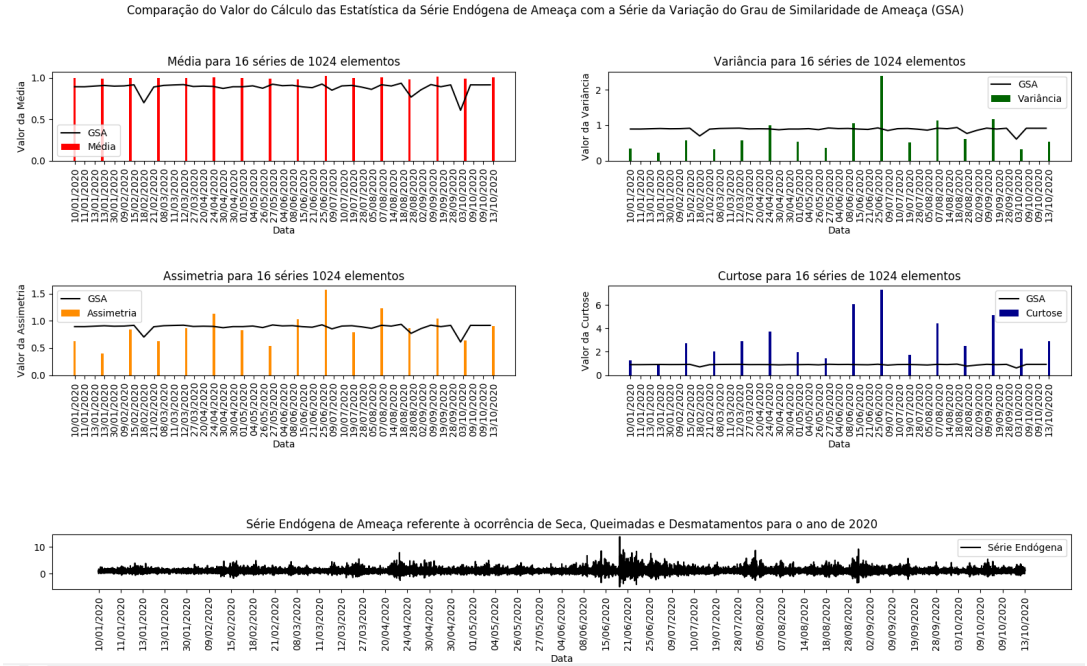
O resultado do processamento desses dados da Tabela 5.13 pelo "*P-Model*",

gerou uma série endógena de ameaças apresentada na Figura 5.42. Com essa série apresenta-se a análise estatística com os parâmetros de média, variância, curtose e assimetria, para verificar se a série endógena realiza uma boa representação da variação de ameaça presente na série temporal de ameaça. A análise desses parâmetros permitiu identificar as seguintes características.

- O parâmetro de Média, para esses dados de 2020, também, não representou de forma adequada os picos e vales da série de ameaças.
- O parâmetro de Variância demonstrou de maneira bem satisfatória a variação de ameaças das séries de ameaças, apresentando de forma definida todos os picos e vales da série.
- O parâmetro de Curtose, também apresentou de maneira satisfatória a variação de ameaças, apresentando todos os seus picos e vales. O parâmetro de curtose, no final do período analisado, apresentou valores maiores que 3 ou excesso > 0, definindo, também, uma curtose leptocúrtica. Dessa forma, observa-se que as caldas da curtose são mais longas e gordas a esquerda. Essas características sugerem a possibilidade de acontecimento de algum evento maior ou evento extremo, posteriormente ao período analisado.
- O parâmetro de Assimetria, apresentou as variações de picos e vales da série de ameaças. Apesar de seus valores, serem sempre maior que zero, as suas caldas apresentam uma leve inclinação negativa para a esquerda. Essas características, também sugerem a possibilidade de ocorrer algum evento maior.

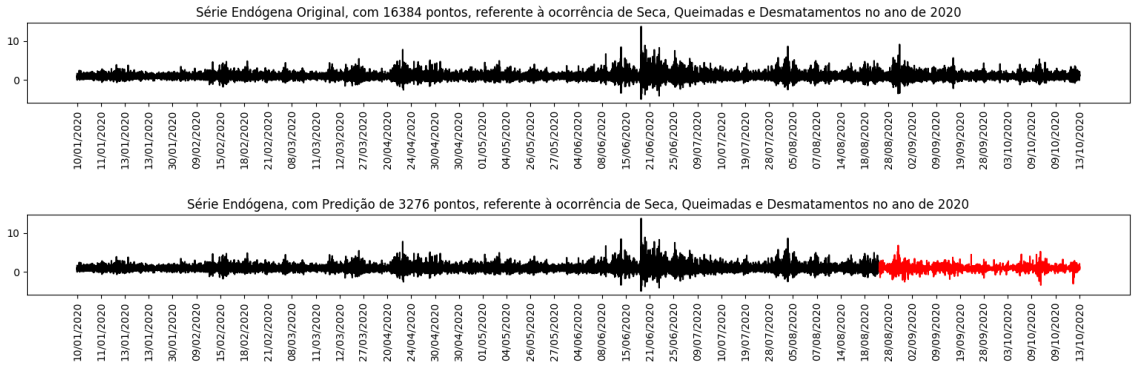
A análise dos parâmetros de média, variância, curtose e assimetria, apresentou que os parâmetros de curtose e assimetria, tiveram uma melhor resposta para sugerir que a série endógena estaria representando as variações de ameaças e que isso poderia sugerir algum acontecimento maior do que as ameaças analisadas. Considerando as análises desses parâmetros, para todos os modelos de dados dos eventos extremos sociais e dos eventos extremos naturais, apesar da variância apresentar em todas as análises uma boa resposta para a representação dos picos e vales das séries, não se encontrou uma relação dos valores de variância entre os diferentes populações ou modelos de dados analisados.

Figura 5.42 - Variação do cálculo dos parâmetros média, variância, curtose e assimetria para a série temporal endógena de ocorrência de seca, queimadas e desmatamentos do ano de 2020.



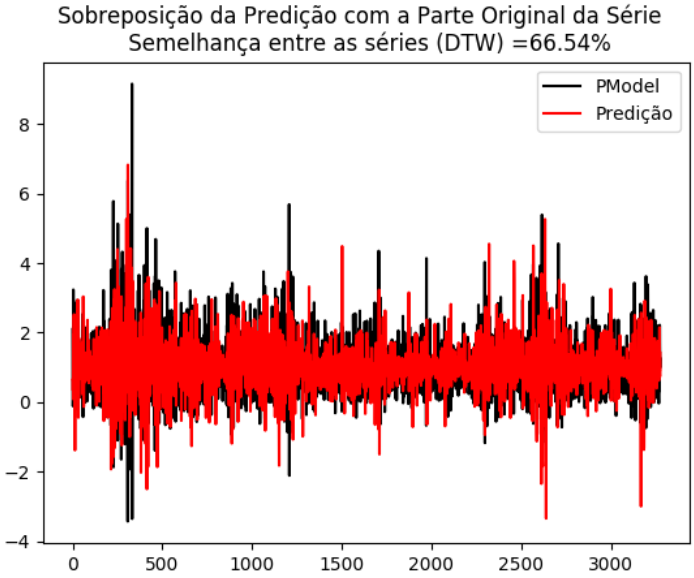
Igualmente, após a análise da série endógena, foi realizada a predição de uma fração de 3276 elementos dessa série. A Figura 5.43 apresenta duas séries temporais endógenas: a primeira série, é a representação original do "*P-Model*" da variação de ameaças e está toda na cor preta e a segunda série é a representação da arquitetura de aprendizado profundo, apresentada na cor vermelha.

Figura 5.43 - Predição de uma fração da série temporal endógena sobre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica no ano de 2020.



Para a comprovação do resultado da predição para a ocorrência de seca, queimadas e desmatamentos do ano de 2020, gerou-se um gráfico de sobreposição dos dados preditos com os dados originais da série em conjunto com o cálculo de DTW, que representa a similaridade entre os dois trechos de série. No gráfico da Figura 5.44, com a sobreposição das séries, apresenta-se o valor de similaridade para os trechos de séries analisados.

Figura 5.44 - Comparação da predição do trecho da série endógena com o trecho original pelo valor de similaridade para o ano de 2020.



A Figura 5.44, mostra uma similaridade entre os trechos de séries analisados de 66,52%. Esse valor continua demonstrando a boa resposta da metodologia e arquitetura de aprendizado profundo para a predição dos eventos extremos analisados.

5.3.2.2 Segunda fase de predição

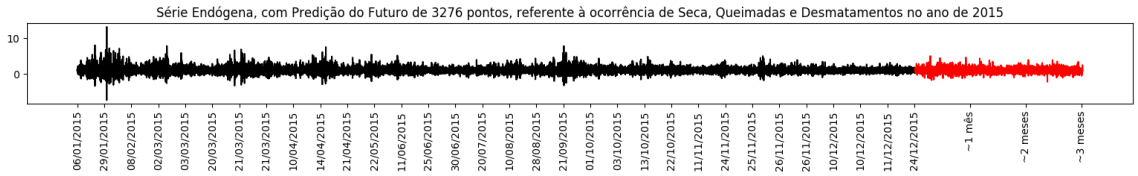
A segunda fase de predição apresenta como resultado a representação de futuro, das informações coletadas para os dados de eventos extremos naturais, gerada pela arquitetura de aprendizado profundo. Essa representação da arquitetura de rede é comparada com a representação do "*P-Model*", das mesmas informações, considerando uma quantidade de elementos preditos de 3276 pontos, que correspondem cerca de a 3 meses de dados, dos meses sucessores. O período das coletas, dos dados

de eventos extremos naturais, ocorreu entre janeiro e dezembro para os anos de 2015 a 2019 e de 2020, coletou-se dados de janeiro a outubro. Essa predição de futuro, dos dados de eventos extremos naturais, também, foi realizada utilizando a mesma arquitetura de aprendizado profundo com rede neural LSTM e com a realimentação da rede para cada grupo de valores de saída, até se chegar ao conjunto de valores preditos. Esses resultados são apresentados nas Seções a, b, c, d, e e f.

a) **Predição do futuro para a ocorrência de seca, queimadas e desmatamentos para o ano de 2015**

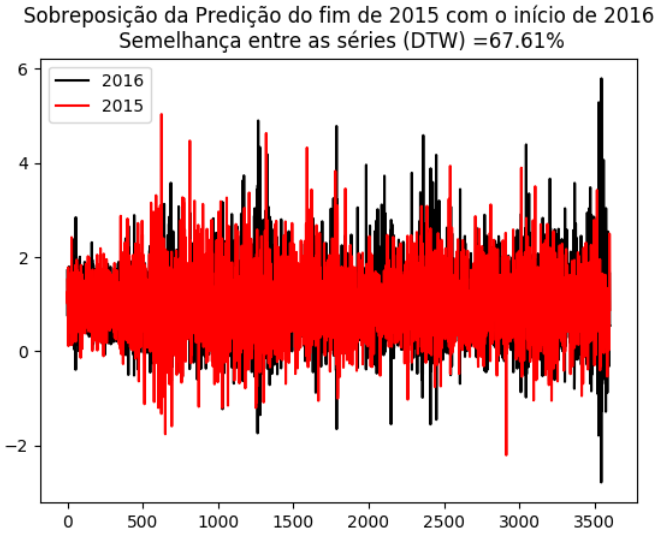
A predição do futuro da série endógena de ameaças do ano de 2015, gerada pelo algoritmo "*P-Model*" na Seção a, é apresentada na Figura 5.45, na coloração vermelha e uma separação aproximada do intervalo de tempo de 3 meses dos dados preditos.

Figura 5.45 - Predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos para o ano de 2015.



A comprovação desses resultados foi realizada utilizando a mesma quantidade de elementos de 3276 do início da série endógena de ameaças do ano de 2016, apresentada na Seção b. Da mesma forma, foi utilizado na comparação desses trechos de séries, o cálculo do valor de similaridade, cálculo de DTW, e a sobreposição dos elementos preditos com os elementos iniciais da série do ano de 2016. Esses resultados são apresentados na Figura 5.46.

Figura 5.46 - Sobreposição dos 3276 pontos preditos com os 3276 pontos iniciais da série de 2016 e cálculo do valor de similaridade para esses elementos.

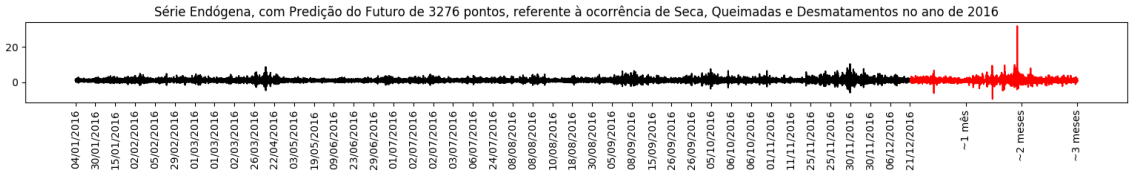


Observa-se, na Figura 5.46, que o resultado apresenta uma similaridade de 67,61% entre os trechos preditos das séries e que isso também demonstra o bom desempenho da metodologia desenvolvida na tese.

b) **Predição do futuro para a ocorrência de seca, queimadas e desmatamentos para o ano de 2016**

A predição do futuro para os dados de ocorrência de seca, queimadas e desmatamentos para o ano de 2016, foi realizada seguindo o mesmo processo utilizado para a predição dos dados do ano de 2015. Utilizaram-se os dados da série endógena, da Seção b, para realizar a predição de 3276 elementos, apresentados na Figura 5.47, na cor vermelha.

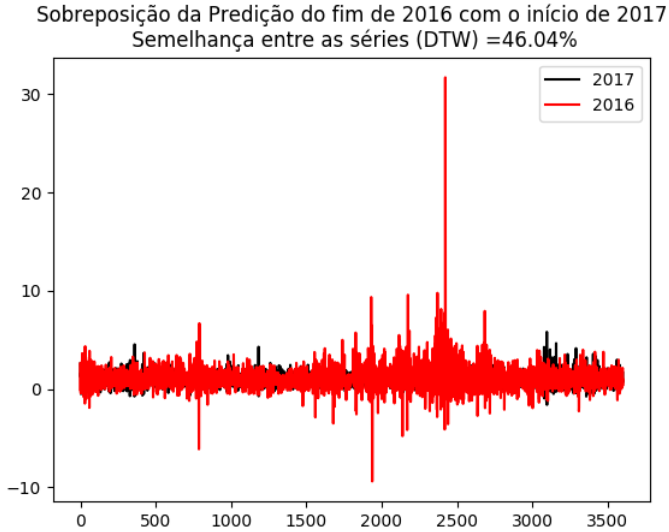
Figura 5.47 - Predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos para o ano de 2016.



A comprovação do resultado da predição, também utilizou a mesma quantidade de elementos do início da série endógena do ano de 2017, apresentada

na Seção c. Identicamente, realizou-se o cálculo do valor de similaridade e a sobreposição dos 3603 dos pontos preditos em 2016 nos pontos iniciais da série do ano de 2017. Esses resultados são apresentados na Figura 5.48.

Figura 5.48 - Sobreposição dos 3276 pontos preditos com os 3276 pontos iniciais da série de 2017 e cálculo do valor de similaridade para esses elementos.

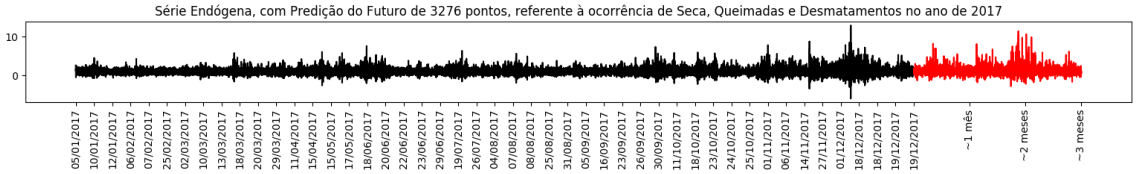


A comparação da representação de futuro, da variação de ameaças, da arquitetura de aprendizado profundo com a representação do "*P-Model*" para a mesma variação de ameaças, apresentada na Figura 5.48, mostra uma similaridade de 46,04%, o que demonstra que a metodologia teve um acerto satisfatório na predição desses pontos futuros.

c) **Predição do futuro para ocorrência de seca, queimadas e desmatamentos no ano de 2017**

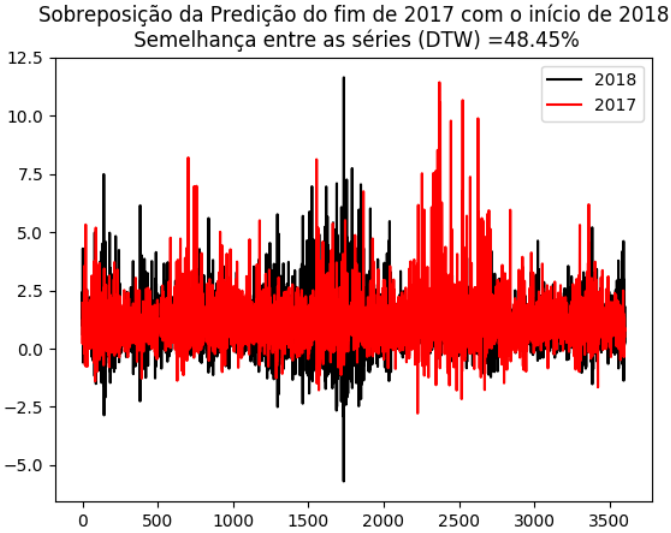
Da mesma forma, para a predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos no ano de 2017, utilizou-se os mesmos dados apresentados na Seção c. A Figura 5.49, mostra o resultado da predição do futuro para 3276 elementos, na cor vermelha.

Figura 5.49 - Predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos no ano de 2017.



O futuro predito de aproximadamente 3 meses, foi comparado com a mesma quantidade de dados do início da série endógena de ameaças do ano de 2018, apresentada na Seção d. Além da forma visual da comparação, também se apresenta na Figura 5.50, o valor de similaridade, calculado por meio do conceito de DTW, entre os trechos de séries comparados.

Figura 5.50 - Sobreposição dos 3276 pontos preditos com os 3276 pontos iniciais da série de 2018 e cálculo do valor de similaridade para esses elementos.

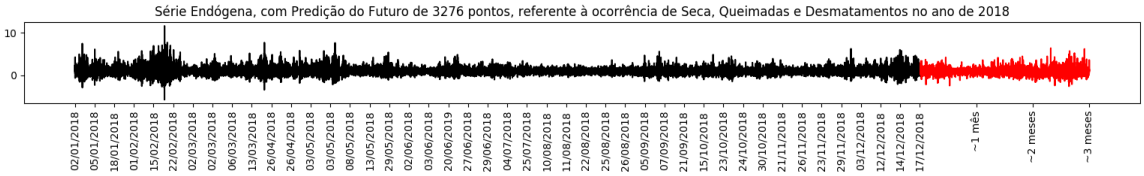


A Figura 5.50 apresenta uma similaridade entre as séries de 48,7%. Esse valor também indica um desempenho satisfatório da metodologia desenvolvida na tese.

d) **Predição do futuro para ocorrência de seca, queimadas e desmatamentos para o ano de 2018**

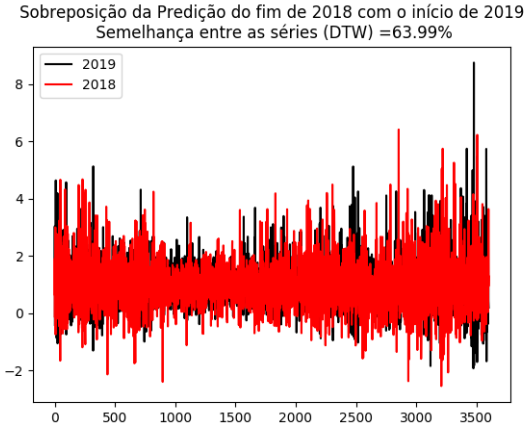
A predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos para o ano de 2018, foi desenvolvida utilizando as mesmas sequências de passos dos anos de 2015, 2016 e 2017. Na análise utilizou-se os mesmos dados apresentados na Seção d, sobre a série endógena de ameaças do ano de 2018. A Figura 5.51 mostra o resultado da predição do futuro para 3276 elementos, que também representam o valor aproximado de 3 meses de intervalo de tempo.

Figura 5.51 - Predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos para o ano de 2018.



O resultado da predição foi comparado com a mesma quantidade de dados do início da série endógena de ameaças do ano de 2019. Essa comparação foi realizada pela sobreposição dos dados preditos do ano de 2018 com os dados do início de 2019 e também pelo cálculo do valor de similaridade, utilizando o conceito de DTW. Esses resultados podem ser observados na Figura 5.52.

Figura 5.52 - Sobreposição dos 3276 pontos preditos com os 3276 pontos iniciais da série de 2019 e cálculo do valor de similaridade para esses pontos.

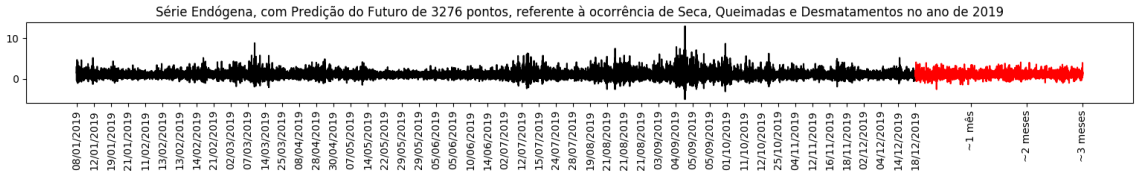


Observa-se, na Figura 5.52, que o resultado de similaridade apresenta um valor de 63,99% entre os trechos de séries analisados. Esse resultado também mostra o bom desempenho da metodologia desenvolvida na tese.

e) **Predição do futuro para a ocorrência de seca, queimadas e desmatamentos para o ano de 2019**

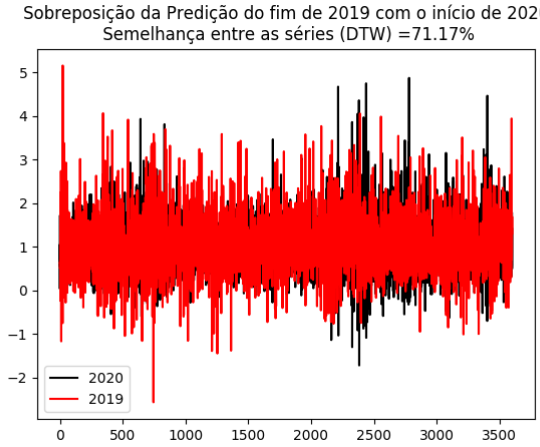
Da mesma forma, a predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos para o ano de 2019, utilizou os mesmos dados da série endógenas de ameaças apresentados na Seção e. A Figura 5.53, mostra o resultado da predição do futuro para 3276 elementos ou 3 meses de intervalo de tempo, na cor vermelha.

Figura 5.53 - Predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos para o ano de 2019.



A verificação da precisão desses resultados, também foi desenvolvida fazendo a comparação entre o futuro de 3 meses preditos com a representação de 3 meses do ano de 2020. Essa comparação é apresentada pelo valor de similaridade entre os trechos de séries analisados. A comprovação desses resultados pode ser visualizada na Figura 5.54, na qual se realizou a sobreposição dos 3276 elementos preditos nos elementos iniciais da série do ano de 2020 e apresenta-se a similaridade entre esses elementos das séries.

Figura 5.54 - Sobreposição dos 3276 elementos preditos com os 3276 elementos iniciais da série de 2020 e cálculo do valor de similaridade para esses elementos.

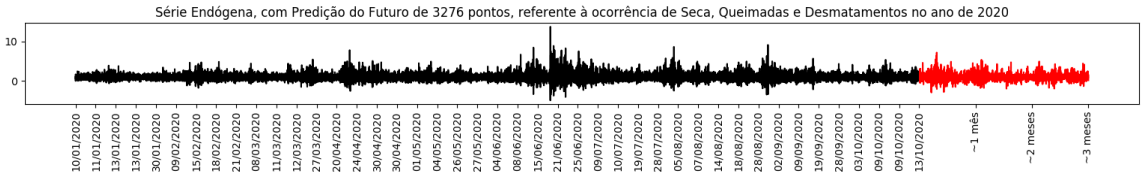


A similaridade alcançada com a comparação dos trechos de séries foi de 71,17%. Esse valor também demonstra o bom desempenho da metodologia com a arquitetura de aprendizado profundo desenvolvida na tese.

f) **Predição do futuro para a ocorrência de seca, queimadas e desmatamentos para o ano de 2020**

Para a finalização dos testes de aplicação dos dados dos eventos extremos naturais na metodologia desenvolvida na tese, realizou-se a predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamentos para o ano de 2020. Nessa aplicação, como nas anteriores, utilizaram-se os dados, da série temporal endógena, apresentados na Seção f. De forma padrão, a Figura 5.55, mostra o resultado dessa predição do futuro para 3276 elementos, na cor vermelha.

Figura 5.55 - Predição do futuro da série temporal endógena de ameaças de ocorrência de seca, queimadas e desmatamento para o ano de 2020.



Considerando as análises dos resultados alcançados para as predições dos anos de 2015 (Figura 5.46), 2016 (Figura 5.48), 2017 Figura (5.50), 2018 (Figura 5.52) e 2019 (Figura 5.54), observa-se que a variação de ameaças continua nos 3 meses posteriores ao período de análise, como observado na predição do futuro da série da Figura 5.55. Assim, sugere-se que para os meses de novembro, dezembro de 2020 e janeiro de 2021, as ameaças de seca, queimadas e desmatamentos devem ter continuidade. Isso se comprova pelo gráfico da Figura 5.40, que apresenta a quantidade de focos de queimadas e desmatamentos para esses meses.

5.3.2.3 Resumo dos resultados das predições realizadas para os eventos extremos sociais e eventos extremos naturais

Nesta seção, apresenta-se, na Tabela 5.14, os resultados das similaridades ou precisões alcançadas com as comparações das predições realizadas para os eventos extremos sociais, utilizando os dados sobre os conflitos armados entre a Síria e USA, Irã e USA e Global. Já na Tabela 5.15, apresentam-se os resultados das similaridades ou precisões alcançadas com as comparações das predições realizadas para os eventos extremos naturais, considerando os dados de ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica, para os anos de 2015, 2016, 2017, 2018, 2019 e 2020. Ambas as predições foram realizadas para a quantidade de 3276 elementos ou para um período aproximado de 3 meses.

Tabela 5.14 - Resultados das similaridades ou precisões alcançadas com as predições dos dados sobre os conflitos armados entre a Síria e USA, Irã e USA e Global.

Conflito Armado	**Similaridade entre trechos das Séries**
Síria e USA	65,12%
Irã e USA	74,39%
Global	72,9%

Tabela 5.15 - Resultados das similaridades ou precisões alcançadas com as predições dos dados sobre a ocorrência de seca, queimadas e desmatamentos para os anos de 2015, 2016, 2017, 2018, 2019 e 2020.

Ano	**Similaridade entre os trechos das séries**	**Similaridade do futuro entre os trechos das séries**
2015	74,43%	67,61%
2016	52,64%	46,04%
2017	47,26%	48,45%
2018	66,49%	63,99%
2019	70,48%	71,17%
2020	66,54%	–

Os resultados apresentados nas Tabelas 5.14 e 5.15, demonstram que a metodologia, que utiliza a técnica de aprendizado profundo com a rede neural LSTM, desenvolvida na tese, apresentou uma variação de precisão constante tanto para os dados de eventos extremos sociais quanto para os dados de eventos extremos naturais. Assim, esses resultados sugerem que essa metodologia pode ser aplicada para qualquer modelo de evento extremo. Também verificou-se, que as similaridades apresentadas para as predições de futuro dos eventos extremos sociais, demonstraram-se consistentes com as notícias publicadas durante os 3 meses preditos (DEJEVSKY, 2017), (MARCUS, 2020), (REUTERS, 2021). Já as predições de futuro para os eventos extremos naturais, apresentaram uma precisão, na comparação com o respectivo ano sucessor, variando entre 46% e 71%, o que também apresenta o bom desempenho da metodologia desenvolvida na tese. Para os dados de eventos extremos naturais do ano de 2020, a predição do futuro apresentou que as ameaças se manteriam. Assim, demonstrou-se, utilizando as informações do programa TerraBrasilis, Figura 5.40, que durante o período de três meses, posterior ao período de análise dos dados, a quantidade de foco de queimadas e desmatamentos aumentaram. Isso novamente demonstra o bom desempenho da metodologia desenvolvida.

6 CONCLUSÕES

O advento das mídias e redes sociais possibilitaram que circulassem as mais diversos categorias de informações e opiniões entre a sociedade. Um modelo de informação que chama atenção para uma análise mais profunda são as notícias sobre os eventos extremos sociais e naturais. Nesta tese, o evento extremo social é o conflito armado, e o evento extremo natural é o desmatamento gerado pelas queimadas. Assim, devido aos grandes problemas gerados por esses eventos, esta tese visa analisar e gerar um modelo de previsão do futuro de ameaças de eventos extremos sociais e naturais, utilizando dados das mídias sociais, técnicas de análise de sentimentos, séries temporais com o algoritmo "*P-Model*" e aprendizado de máquina, com a intenção de auxiliar na prevenção, alerta e consequente redução dos problemas gerados por esses eventos à sociedade.

Na tese foram realizadas análises de dados coletados de mídias sociais de grande circulação mundial e nacional e de especialidades dos temas abordados (REUTERS, 2019), (CNN, 2020), (FOLHA, 1996), (BUTLER, 1999), etc. Essas informações foram processadas conforme a metodologia desenvolvida para a tese, como apresentado no Capítulo 4, para a realização da predição do futuro de eventos extremos naturais e sociais. Como resultados da aplicação dessa metodologia, gerou-se três resultados na tese que foram: o portfólio com o grau de ameaças do evento extremo em questão, a validação do algoritmo "*P-Model*" como gerador de séries temporais endógenas e a predição do futuro dos eventos extremos sociais e naturais considerados na tese. Os temas de eventos extremos sociais e naturais abordados na tese foram os conflitos armados e a ocorrência de seca, queimadas e desmatamentos, respectivamente.

Para os dados de eventos extremos sociais, consideraram-se as informações contidas nas mídias sociais sobre os conflitos armados entre Síria e USA, Irã e USA e um conflito Global que envolve os países, China, Índia, Paquistão, Reino Unido, Japão, USA, Coreia do Norte, Coreia do Sul, Taiwan e Indonésia. Depois das notícias serem processadas usando as técnicas de KDD, análise de sentimentos, algoritmo "*P-Model*" e aprendizado profundo com rede LSTM, obteve-se como resultado da predição de ameaças futuras desses conflitos. Desses resultados, verificou-se que os valores de similaridades apresentadas para as predições de futuro dos eventos extremos sociais, apresentaram-se de forma a sugestionar os fatos que estavam sendo publicados nas notícias dos períodos de 3 meses preditos (DEJEVSKY, 2017), (MARCUS, 2020), (REUTERS, 2021).

Para os dados referentes a eventos extremos naturais, analisaram-se informações so-

bre a ocorrência de seca, queimadas e desmatamentos na região da floresta Amazônica para os anos de 2015, 2016, 2017, 2018, 2019 e 2020. Como resultado da análise de ameaça futura de um possível aumento do número de ocorrência de seca, queimadas e desmatamentos na região considerada, obteve-se uma precisão entre 46% e 71% para a predição de 3276 pontos para os anos sucessores dos anos de 2015, 2016, 2017, 2018, 2019 e 2020, como apresentado na Tabela 5.15. Na análise do ano de 2020, verifica-se que a ameça de ocorrência de seca, queimadas e desmatamentos continua constante para os 3 meses subsequentes. Isso foi comprovado, observando os dados, apresentados na Figura 5.40, do programa TerraBrasilis do INPE.

Para a verificação das séries endógenas e do método de predição gerados, realizou-se o cálculo dos parâmetros de média, variância, curtose, assimetria. Dessa investigação, observou-se que os parâmetros de curtose e assimetria, sugeriram que a série endógena de ameaças estaria fazendo a representação da variação de ameaças e que essa variação poderia indicar a ocorrência de algo maior que uma ameaça. O parâmetro de média não gerou informações que pudessem direcionar para alguma análise nesse sentido. O parâmetro de variância, no que lhe concerne, apresentou nas análises de todos os portfólios de eventos extremos sociais e naturais analisados, uma boa representação dos picos e vales das séries de ameaças. No entanto, considerando todas as populações ou modelos de dados analisados, esse parâmetro não permitiu gerar uma relação consistente entre essas informações.

Os resultados alcançados durante o desenvolvimento da tese foram apresentados e/ou publicados em conferências e revistas, como citados a seguir.

- publicação do artigo "*Sentiment Analysis Applied to Analyze Society's Emotion in Two Different Context of Social Media Data*" no "*journal Iberoamerican Society of Artificial Intelligence (IBERAMIA)*". O artigo publicado no "*journal*" da IBERAMIA é apresentado na íntegra na Seção *A.1*.

- submissão do artigo "*Modeling and Forecasting Interstate Armed Conflicts Assisted by Sentiment Analysis and LSTM-RNN*" no "*journal Frontiers in Big Data*", seção "*Machine Learning and Artificial Intelligence*".

- apresentação do pôster "*Forecasting Endogenous and Exogenous Time Series generated by P-model Algorithm using the Keras Machine Learning API*" no Congresso Nacional de Matemática Aplicada e Computacional - CNMAC 2019.

- apresentação do pôster "*A Social Media Portfolio of Threats in the Context of Armed Conflicts*" no Simpósio de Aplicações Operacionais em Áreas de Defesa - SIGE 2019.

Como trabalhos futuros sugerem-se,

- ampliar a análise e comparação dos dados de focos de queimadas e desmatamentos apresentados para os eventos extremos naturais para os outros anos analisados.
- expandir as análises realizadas para eventos extremos exógenos.
- aplicar os conceitos de ciência cidadã para a melhoria do texto base de análise das ameaças.
- aperfeiçoar a aplicação de análise de sentimento, considerando informações de emoções mais específicas das notícias como sarcasmos, ironia e a utilização de dados de imagens.

REFERÊNCIAS BIBLIOGRÁFICAS

ABBAS, O. M.; MUSA, M. E. M.; IBRAHIM, S. B. Using decision tree to predict armed conflicts in Sudan. **International Journal of Computer (IJC)**, v. 16, n. 1, p. 9–17, 2015. Disponível em: <https://www.ijcjournal.org/index.php/InternationalJournalOfComputer/article/view/388>. 17, 21, 22

ABBURI, H.; AKKIREDDY, E. S. A.; GANGASHETTY, S. V.; MAMIDI, R. Multimodal sentiment analysis of telugu songs. In: WORKSHOP ON SENTIMENT ANALYSIS WHERE AI MEETS PSYCHOLOGY(SAAIP 2016), Nagoya, Japan. **Proceedings...** New York City, USA: IJACAI, 2016. p. 48–52. Disponível em: <http://ceur-ws.org/Vol-1619/paper8.pdf>. 10, 14, 32

AHMADI, V. Deforestation prediction using neural networks and satellite imagery in a spatial information system. **CoRR**, abs/1803.02489, 2018. Disponível em: <http://arxiv.org/abs/1803.02489>. 20, 22, 24

ALI, A. E.; STRATMANN, T. C.; PARK, S.; SCHÖNING, J.; HEUTEN, W.; BOLL, S. C. Measuring, understanding, and classifying news media sympathy on twitter after crisis events. In: CHI '18: 2018 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS. **Proceedings...** Association for Computing Machinery, 2018. p. 1–13. ISSN 978-1-4503-5620-6. Disponível em: <https://dl.acm.org/doi/10.1145/3173574.3174130>. 148

ANTUNES, M. N.; SILVA, C. H.; AES, M. C. S. G.; CO, M. H. L. R. Monitoramento de informação em mídias sociais: o e-monitor dengue. **Transinformação**, v. 26, n. 1, p. 1–11, 2014. ISSN 0103-3786. Disponível em: <https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862014000100002>. 8, 14, 21, 32, 151

ARMENTERAS, D.; GIBBES, C.; ANAYA, J. A.; DÁVALOS, L. M. Integrating remotely sensed fires for predicting deforestation for redd. **Ecological Applications**, v. 27, n. 4, p. 1294–1304, 2017. DOI:10.1002/eap.1522. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/28208227/>. 19, 22, 24

ARUNACHALAM, R.; SARKAR, S. The new eye of government: citizen sentiment analysis in social media. In: IJCNLP 2013 WORKSHOP ON NATURAL LANGUAGE PROCESSING FOR SOCIAL MEDIA (SOCIALNLP. **Proceedings...** Nagoya, Japan: Asian Federation of Natural Language Processing,

2013. p. 23–28. Disponível em: <https://www.aclweb.org/anthology/W13-4204>. 6, 13, 32, 151

AYLIEN. **Leveraging deep learning for multilingual sentiment analysis**. 2016. Disponível em: <$http://blog.aylien.com/leveraging-deep-learning-for-multilingual/$>. Acesso em: 29 abr. 2017. 146, 153

BALABANTARAY, R. C.; MOHAMMAD, M.; SHARMA, N. Multi-class twitter emotion classification: a new approach. **International Journal of Applied Information Systems**, v. 4, n. 1, p. 48–53, Sept. 2012. ISSN 2249-0868. Disponível em: <https://research.ijais.org/volume4/number1/ijais12-450651.pdf>. 13, 32, 151

BALAHUR, A.; TURCHI, M. Comparative experiments for multilingual sentiment analysis using machine translation. In: CEUR WORKSHOP. **Proceedings...** 2012. v. 917, p. 75–86. ISSN 1613-0073. Disponível em: <http://ceur-ws.org/Vol-917/SDAD2012_8_Balahur.pdf>. 5

BAO, W.; YUE, J.; RAO, Y. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. **PLOS ONE**, v. 12, n. 7, p. 1–24, 07 2017. DOI:10.1371/journal.pone.0180944. Disponível em: <https://doi.org/10.1371/journal.pone.0180944>. 35

BARINO, F. O.; SANTOS, A. B. Rede neural convolucional 1d aplicada à previsão da vazão no rio madeira. In: SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS, 38., 2021. **Anais...** 2021. DOI:10.14209/SBRT.2020.1570640893. Disponível em: <https://biblioteca.sbrt.org.br/articles/2140>. 63

BIRD, S.; KLEIN, E.; LOPER, E. **Natural language processing with python: analyzing text with the natural language toolkit**. USA: O Reilly Media, 2009. 479 p. ISSN 978-0596516499. 2, 9, 56, 146, 150, 163

BONN, U. of. **Venema P-Model algorithm**. 2019. Disponível em: <http://www2.meteo.uni-bonn.de/staff/venema/themes/surrogates/pmodel/pmodel.m>. Acesso em: 06 mar. 2019. 42, 76

BROWN, S. **Measures of shape: skewness and kurtosis**. 2020. Disponível em: <https://brownmath.com/stat/shape.htm#Kurtosis>. Acesso em: 30 jan. 2021. 49

BROWNLEE, J. **A Gentle introduction to long short-term memory networks by the experts**. 2017. Disponível em: <https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/>. Acesso em: 20 jun. 2019. 35, 63

_____. **A tour of machine learning algorithms**. 2019. Disponível em: <http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>. Acesso em: 15 ago. 2019. 33

BUTLER, R. A. **Notícias ambientais - notícias sobre vida**. 1999. Disponível em: <https://brasil.mongabay.com/>. Acesso em: 13 out. 2020. 2, 27, 38, 53, 125

C.-K.PENG; S.V.BULDYREV; S.HAVTIN; M.SIMONS; H.E.STANLEY; A.L.GOLDBERGERZ. Mosaic organization of dna nucleotides. **Physical Review E**, v. 49, n. 2, 1994. Disponível em: <https://dbiom.org/files/publications/Peng_MosaicOrganizationDNANucleotides.pdf>. 44

CASTELLS, M. **A galáxia da internet reflexões sobre a internet, os negócios e a sociedade**. [S.l.]: Zahar, 2003. 244 p. ISSN 978-8571107403. 27, 51

CECI, F.; GONÇALVES, A. L.; WEBER, R. A model for sentiment analysis based on ontology and cases. **IEEE Latin America Transactions**, v. 14, n. 11, p. 4560–4566, 2016. DOI:10.1109/TLA.2016.7795829. Disponível em: <https://www.researchgate.net/publication/311777565_A_Model_for_Sentiment_Analysis_Based_on_Ontology_and_Cases>. 9, 15

CEDERMAN, L.-E.; WEIDMANN, N. B. Predicting armed conflict: time to adjust our expectations? **Science**, v. 355, n. 6324, p. 474–476, 2017. DOI:10.1126/science.aal4483. Disponível em: <https://science.sciencemag.org/content/355/6324/474.full?ijkey=OpXTDKtFRJiU6&keytype=ref&siteid=sci>. 17, 21, 22, 27, 51

CEPIK, M. A. C. **Espionagem e democracia**. Brasil: FGV, 2003. 936 p. ISBN 85-225-0437-7. 32, 151

CHAPELLE, O.; SCHÖLKOPT, B.; ZIEN, A. **Semi-supervised learning**. Londres: MIT Press, 2006. 524 p. ISBN 978-0-262-03358-9. 33

CHOLLET, F. **Keras: the Python deep learning library**. 2015. Disponível em: <https://keras.io/>. Acesso em: 1 ago. 2016. 3, 36, 37, 40, 56, 63, 146, 159

CLAUSET, A. Trends and fluctuations in the severity of interstate wars. **SciencesAdvances - Social Sciences**, v. 4, n. 2, 2018. DOI:10.1126/sciadv.aao3580. Disponível em: <https://advances.sciencemag.org/content/4/2/eaao3580>. 1, 18, 23

CNN. **CNN - breaking news, latest news and videos**. 2020. Disponível em: <https://edition.cnn.com/>. Acesso em: 08 dez. 2020. 2, 27, 52, 125, 146

CUADRADO, J. C. A.; GóMEZ-NAVARRO, D. P. G. **Un modelo lingüístico-semántico basado en emociones para la clasificación de textos según su polaridad e intensidad**. 290 p. Tese (Doutorado em Informática) — Departamento de Ingeniería del Software e Inteligencia Artificial Facultad de Informática Universidad Complutense de Madrid, Madri, 2011. 31, 150

DAHAL, B.; KUMAR, S. A. P.; LI, Z. Topic modeling and sentiment analysis of global climate change tweets. **Social Network Analysis and Mining**, v. 9, n. 1, 2019. ISSN 1869-5450. DOI:https://doi.org/10.1007/s13278-019-0568-8. Disponível em: <https://www.springerprofessional.de/en/topic-modeling-and-sentiment-analysis-of-global-climate-change-t/16789020>. 12, 16, 21, 67

DAVIS, A.; MARSHAK, A.; CAHALAN R.AND WISCOMBE, W. The landsat scale break in stratocumulus as a three-dimensional radiative transfereffect: implications for cloud remote sensing. **Journal of the Atmospheric Sciences**, v. 54, n. 2, p. 241–260, 1997. Disponível em: <https://journals.ametsoc.org/view/journals/atsc/54/2/1520-0469_1997_054_0241_tlsbis_2.0.co_2.xml>. 77

DAVIS, Z. S. Artificial intelligence on the battlefield: initial survey of potential implications for deterrence, stability, and strategic surprise. **Center for Global Security Research - Lawrence Livermore National Labora**, p. 1–26, 03 2019. Disponível em: <https://cgsr.llnl.gov/content/assets/docs/CGSR-AI_BattlefieldWEB.pdf>. 67

DEEKS, S.; LUBELL, N.; MURRAY, D. Machine learning, artificial intelligence, and the use of force by states. **Journal of National Security Law & Policy**, v. 10, n. 1, 2019. Disponível em: <https://jnslp.com/wp-content/uploads/2019/04/Machine_Learning_Artificial_Intelligence_2.pdf>. 23

DEJEVSKY, M. **As Syria's war enters its endgame, the risk of a US-Russia conflict escalates**. June 2017. Disponível em:

<https://www.theguardian.com/commentisfree/2017/jun/21/syria-war-endgame-us-russia-conflict-washington-moscow-accidental-war>. Acesso em: 29 abr. 2021. 95, 123, 125

DENG, L.; YU, D. Deep learning: methods and applications. **Foundations and Trends in Signal Processing**, v. 7, n. 3-4, p. 197–387, 2013. Disponível em: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/DeepLearningBook_RefsByLastFirstNames.pdf>. 34

DHAOUI, C.; WEBSTER, C.; TAN, L. Social media sentiment analysis: lexicon versus machine learning. **Journal of Consumer Marketing**, v. 34, 08 2017. DOI:10.1108/JCM-03-2017-2141. 11, 15

DHAWAN, V.; ZANINI, N. Big data and social media analytics. **Research Matters: A Cambridge Assessment**, v. 18, p. 36 – 41, 2014. Disponível em: <https://www.cambridgeassessment.org.uk/Images/465808-big-data-and-social-media-analytics.pdf>. 145

D'ORAZIO, V.; HONAKER, J.; PRASADY, R.; SHOEMATE, M. Modeling and forecasting armed conflict: automl with human-guided machine learning. In: 2019 IEEE INTERNATIONAL CONFERENCE ON BIG DATA. **Proceedings...** [S.l.], 2019. p. 4714–4723. 18, 23

EMMERT-STREIB, F.; YANG, Z.; FENG, H.; TRIPATHI, S.; DEHMER, M. An introductory review of deep learning for prediction models with big data. **Frontiers in Artificial Intelligence - Deep Learning in Computational Social Science**, v. 27, p. 16–32, 2018. DOI:https://doi.org/10.3389/frai.2020.00004. Disponível em: <https://www.frontiersin.org/articles/10.3389/frai.2020.00004/full>. 63

ESULI, A. **SentiWordNet**. 2002. Disponível em: <https://github.com/aesuli/SentiWordNet>. Acesso em: 09 out. 2018. 9

FOLHA, G. **UOL - o melhor conteúdo**. 1996. Disponível em: <www.uol.com.br>. Acesso em: 08 dez. 2020. 38, 53, 125

GABER, M. M.; COCEA, M.; WIRATUNGA, N.; GOKER, A. **Advances in social media analysis**. Switzerland: Springer, 2015. ISSN 1860-949X. 27

GAO, S. Deforestation prediction using time series and lstm. In: 2019 INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY AND

COMPUTER APPLICATION (ITCA). **Proceedings...** [S.l.], 2019. p. 95–99. DOI:10.1109/ITCA49981.2019.00029. 20, 25

GARVEY, C.; MASKAL, C. Sentiment analysis of the news media on artificial intelligence does not support claims of negative bias against artificial intelligence. **OMICS: A Journal of Integrative Biology**, v. 24, n. 2, p. 286–299, 2020. DOI:http://doi.org/10.1089/omi.2019.0078. Disponível em: <https://www.liebertpub.com/doi/10.1089/omi.2019.0078>. 148

GHOSH, R.; DAVI; K.; RAVI, V. A novel deep learning architecture for sentiment classification. In: IN RECENT ADVANCES IN INFORMATION TECHNOLOGY (RAIT) INTERNATIONAL CONFERENCE ON DATE OF CONFERENCE, 3., 2016, Dublin, Ireland. **Proceedings...** Dublin, Ireland, 2016. p. 1–5. ISBN 978-1-4799-8580-7. Disponível em: <https://ieeexplore.ieee.org/document/7507953>. 10, 14, 147

GIUSTI, R.; BATISTA, G. E. A. P. A. An empirical comparison of dissimilarity measures for time series classification. In: BRAZILIAN CONFERENCE ON INTELLIGENT SYSTEMS. **Proceedings...** 2013. p. 82–88. DOI:10.1109/BRACIS.2013.22. Disponível em: <https://sites.icmc.usp.br/gbatista/files/bracis2013_1.pdf>. 50

GLOBO. **G1 - o portal de notícias da Globo**. 2006. Disponível em: <https://g1.globo.com/>. Acesso em: 08 dez. 2020. 2, 27, 53

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning - adaptive computation and machine learning series**. The MIT Press, 2016. 800 p. ISBN 0262035618. Disponível em: <$http://www.deeplearningbook.org$>. 3, 34, 153, 154

GOOGLE. **TensorFlow wide & deep learning tutorial**. 2015. Disponível em: <$https://www.tensorflow.org/tutorials/wide_and_deep$>. Acesso em: 10 out. 2016. 56, 146, 150, 158, 159

_____. **API cloud natural language - google cloud**. 2016. Disponível em: <https://cloud.google.com/natural-language?hl=pt-br>. Acesso em: 25 fev. 2017. 2, 10

GROENEVELD, R. A. An influence function approach to describing the skewness of a distribution. **The American Statistician**, v. 45, n. 2, p. 97–102, may 1991. Disponível em: <http://www.jstor.org/stable/2684367>. 47

GUARDIAN, T. **News, sport and opinion from the guardian's US edition**. 2020. Disponível em: <https://www.theguardian.com/international>. Acesso em: 08 dez. 2020. 2, 27, 38, 52, 146

HALKIA, M.; FERRI, S.; SCHELLENS, M. K.; PAPAZOGLOU, M.; THOMAKOS, D. The global conflict risk index: a quantitative tool for policy support on conflict prevention. **Progress in Disaster Science**, v. 6, p. 100069, 2020. ISSN 2590-0617. DOI:https://doi.org/10.1016/j.pdisas.2020.100069. Disponível em: <http://www.sciencedirect.com/science/article/pii/S2590061720300065>. 19, 24

HALSEY, T. C.; JENSEN, M. H.; KADANOFF, L. P.; PROCACCIA, I.; SHRAIMAN, B. I. Fractal measures and their singularities: the characterization of strange sets. **Nuclear Physics B - Proceedings Supplements**, v. 2, p. 501–511, 1987. ISSN 0920-5632. Disponível em: <https://www.sciencedirect.com/science/article/pii/0920563287900363>. 41

HAO, J.-X.; FU, Y.; HSU, C.; LI, X. R.; CHEN, N. Introducing news media sentiment analytics to residents attitudes research. **Journal of Travel Research**, 2019. DOI:10.1177/0047287519884657. Disponível em: <https://doi.org/10.1177/0047287519884657>. 148

HARVARD, U. of. **Detrended fluctuation analysis (DFA)**. 2019. Disponível em: <http://reylab.bidmc.harvard.edu/download/DFA/intro/>. Acesso em: 16 nov. 2019. 45

HAYES, A.; JAMES, M. **Leptokurtic distributions**. 2021. Disponível em: <https://www.investopedia.com/terms/l/leptokurtic.asp>. Acesso em: 30 jan. 2021. 49

HAYKIN, S. O. **Neural networks and learning machines**. 3. ed. New York: Pearson, 2008. 936 p. ISBN 978-0131471399. 32, 33, 34, 81, 152, 153, 154

HEGRE, H.; ALLANSSON, M.; BASEDAU, M.; COLARESI, M.; CROICU, M.; FJELDE, H.; HOYLES, F.; HULTMAN, L.; HöGBLADH, S.; JANSEN, R.; MOUHLEB, N.; MUHAMMAD, S. A.; NILSSON, D.; NYGåRD, H. M.; OLAFSDOTTIR, G.; PETROVA, K.; RANDAHL, D.; RøD, E. G.; SCHNEIDER, G.; UEXKULL, N. von; VESTBY, J. Views: a political violence early-warning system. **Journal of Peace Research**, v. 56, n. 2, p. 474–476, 2018. Disponível em: <https://journals.sagepub.com/doi/full/10.1177/0022343319823860>. 17, 22

HELLE, V.; NEGUS, A.; NYBERG, J. **Improving armed conflict prediction using machine learning : views+**. 2018. Disponível em: <https://pdfs.semanticscholar.org/3008/beffb4496316bb1677253de89eb4b2a695c3.pdf>. 18, 23

HIDALGO, J. M. G. **Baseline sentiment analysis with WEKA**. 2013. Disponível em: <$http://jmgomezhidalgo.blogspot.com.br/2013/06/baseline-sentiment-analysis-with-weka.html$>. Acesso em: 20 jan. 2017. 6, 13

HINTON, G. E.; SALAKHUTDINOV, R. R. Reducing the dimensionality of data with neural networks. **Science**, v. 313, n. 5786, p. 504–507, jul. 2006. DOI:10.1126/science.1127647. Disponível em: <https://science.sciencemag.org/content/313/5786/504>. 154

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1997. DOI:1530-888X. Disponível em: <https://doi.org/10.1162/neco.1997.9.8.1735>. 3

HOLLANDER, J.; GRAVES, E.; RENSKI, H.; FOSTER-KARIM, C.; WILEY, A.; DAS, D. A (short) history of social media sentiment analysis. In: _____. **Urban social listening**. (Ed.). [S.l.]: Springer, 2016. p. 15–25. ISBN 978-1-137-59490-7. $DOI : 10.1057/978-1-137-59491-4_2$. Disponível em: <$https://link.springer.com/chapter/10.1057/978-1-137-59491-4_2$>. 67

HYNDMAN, R.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. Melbourne,: OTexts, 2018. 77

INSS CONFERENCE ON BIG DATA. **50 years of deep learning and beyond: an interview with Jürgen Schmidhuber**. 2015. Disponível em: <$https://innsbigdata.wordpress.com/2015/02/09/interview-with-juergen-schmidhuber/$>. Acesso em: 29 abr. 2017. 34

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Potencial e agressividade climática na Amazônia Legal**. 2019. Disponível em: <https://www.ibge.gov.br/geociencias/informacoes-ambientais/climatologia/15818-potencial-de-agressividade-climatica-na-amazonia-legal.html?=&t=o-que-e>. Acesso em: 06 jun. 2020. 67

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **TerraBrasilis**. 2019. Disponível em: <http://terrabrasilis.dpi.inpe.br/>. Acesso em: 10 fev. 2021. 110

JACKSON, P.; MOULINIER, I. **Natural language processing for online applications**: Textretrieval, extraction and categorization. Amsterdam / Philadelphia: John Benjamins B.V., 2002. 237 p. ISBN 90 272 49881. 29, 30, 31, 149

KEYLOCK, C. J. Multifractal surrogate-data generation algorithm that preserves pointwise hölder regularity structure, with initial applications to turbulence. **Physical Review E**, v. 95, p. 032123, Mar 2017. Disponível em: <https://link.aps.org/doi/10.1103/PhysRevE.95.032123>. 42

KOTO, F.; ADRIANI, M. Hbe: Hashtag-based emotion lexicons for twitter sentiment analysis. In: FORUM FOR INFORMATION RETRIEVAL EVALUATION, 2015. **Proceedings...** [S.l.], 2015. 9

KUMARI, S. Impact of big data and social media on society. **Global Journal for Research Analysis**, v. 5, p. 437–438, 03 2016. Disponível em: <https://www.researchgate.net/publication/306034329_impact_of_big_data_and_social_media_on_society>. 145

LABORATORY, M. M. **SenticNet**. 2009. Disponível em: <https://sentic.net/>. Acesso em: 15 set. 2018. 9

LE, Q. V. **A tutorial on deep learning part 2: autoencoders, convolutional neural networks and recurrent neural networks**. Mountain View, CA, 2015. 20p. Tutorial. Disponível em: <http://ai.stanford.edu/~quocle/tutorial2.pdf>. Acesso em: 12 out. 2017. 154, 155

LIMA, D. **Instituto socioambiental (ISA)**. 1994. Disponível em: <https://www.socioambiental.org/pt-br/o-isa>. Acesso em: 13 out. 2020. 53

LIMA, P. M.; GUEDES, E. B. Rainfall prediction for manaus, amazonas with artificial neural networks. In: LATIN AMERICA CONGRESS ON COMPUTATIONAL INTELLIGENCE (LA-CCI). **Proceedings...** 2015. p. 1–6. DOI:10.1109/LA-CCI.2015.7435934. Disponível em: <https://ieeexplore.ieee.org/document/7435934>. 19, 24

LIU, B. Sentiment lexicon generation. In: _____. **Sentiment analysis: mining opinions, sentiments, and emotions**. [S.l.]: Cambridge University Press, 2015. p. 189–201. 9

MALANDRAKIS, N.; KAZEMZADEH, A.; POTAMIANOS, A.; NARAYANAN, S. Sail: a hybrid approach to sentiment analysis. **Second Joint Conference on Lexical and Computational Semantics (SEM)**, v. 2, p. 438–442, june 2016. Disponível em: <https://sail.usc.edu/publications/files/Nikos-SemEval-S13-2072.pdf>. 9

MANNING, C. D.; CARPENTER, B. Probabilistic parsing using left corner language models. In: INTERNATIONAL WORKSHOP ON PARSING TECHNOLOGIES - MIT, 5., 1997, Boston, MA. **Proceedings...** Boston, MA, 1997. p. 105–124. ISSN 978-94-015-9470-7. Disponível em: <https://link.springer.com/chapter/10.1007/978-94-015-9470-7_6>. 6

MARCUS, J. **US-Iran war of words raises fresh fears of Gulf clash**. April 2020. Disponível em: <https://www.bbc.com/news/world-middle-east-52399283>. Acesso em: 25 ago. 2020. 96, 123, 125

MENEVEAU, C.; SREENIVASAN, K. R. Simple multifractal cascade model for fully developed turbulence. **Physical Review Letters**, v. 59, n. 13, p. 1424–1427, 1987. DOI:10.1103/PhysRevLett.59.1424. Disponível em: <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.59.1424>. 40

MICROSOFT CORPORATION. **Machine learning algorithm cheat sheet for microsoft azure machine learning studio**. 2017. Disponível em: <https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-algorithm-cheat-sheet>. Acesso em: 15 mar. 2017. 34

MITCHELL, T. **Machine learning**. USA: McGraw-Hill, 1997. 432 p. ISSN 978-0070428072. 32, 152

MORARIU, V. V.; BUIMAGA-IARINCA, L.; VAMOS, C.; SOLTUZ, S. **Detrended fluctuation analysis of autoregressive processes**. 2007. Disponível em: <https://arxiv.org/pdf/0707.1437.pdf>. 44

MORELLO, T. F.; RAMOS, R. M.; ANDERSON, L. O.; ROSAN, T. M.; STEIL, L. Predicting amazon fires for policy making. **Encontro Nacional de Economia**

- Economia Agrícola e do Meio Ambiente, v. 44, 11 2016. Disponível em: <https://www.anpec.org.br/encontro/2016/submissao/files_I/i11-3b68242e7c3a5a3a7f24ce256c5d517c.pdf>. 19, 24

MOYER, J. D.; KAPLAN, O. **Will the Coronavirus fuel conflict? projections based on economic and development data show an increased risk of internal violence in fragile states driven by rising prices and falling incomes**. Foreign Policy − the Global Magazine of News and Ideas, 2020. Disponível em: <https://foreignpolicy.com/2020/07/06/coronavirus-pandemic-fuel-conflict-fragile-states-economy-food-prices/>. Acesso em: 20 set. 2020. 27

MUCHLINSKI, D.; SIROKY, D.; HE, J.; KOCHER, M. Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data. **Political Analysis**, v. 24, n. 1, p. 87–103, 2016. ISSN 10471987, 14764989. Disponível em: <http://www.jstor.org/stable/24573207>. 17, 22

MUELLER, H.; RAUH, C. **The hard problem of prediction for conflict prevention**. Faculty of Economics, University of Cambridge - Cambridge Working Papers in Economics, Mar 2020. Disponível em: <https://ideas.repec.org/p/cam/camdae/2015.html>. 12, 16, 21, 32

MÄNTYLÄ, M. V.; GRAZIOTIN, D.; KUUTILA, M. The evolution of sentiment analysis − a review of research topics, venues, and top cited papers. **Computer Science Review**, v. 27, p. 16–32, 02 2018. ISSN 1574-0137. DOI:https://doi.org/10.1016/j.cosrev.2017.10.002. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1574013717300606>. 67

NEELAKSHI, J.; ROSA, R. R.; SAVIO, S.; MENESES, F. C. de; STEPHANY, S.; FORNARI, G.; MURALIKRISHNA, P. Spectral fluctuation analysis of ionospheric inhomogeneities over brazilian territory part ii: E-f valley region plasma instabilities. **Advances in Space Research**, v. 64, n. 8, p. 1592–1599, 2019. ISSN 0273-1177. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0273117719305058>. 44

NEMES, L.; KISS, A. Social media sentiment analysis based on covid19. **Journal of Information and Telecommunication**, v. 5, n. 1, p. 1–15, 2021. DOI:10.1080/24751839.2020.1790793. Disponível em: <https://doi.org/10.1080/24751839.2020.1790793>. 12, 16, 21

NEWSBOT. **Related news at the click of a button**. 2019. Disponível em: <https://getnewsbot.com/>. Acesso em: 05 jan. 2019. 2, 162

OLIVEIRA, C. A. **IDEAL - uma interface dialógica em linguagem natural para sistemas especialistas**. 131 p. Tese (Doutorado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 1990. Disponível em: <http://urlib.net/rep/6qtX3pFwXQZ3r59YCT/GUpqq>. Acesso em: 04 maio 2021. 29

PADILHA, A. **Gírias atuais mais usadas na internet**. 2020. Disponível em: <$https://www.dicionariopopular.com/girias-atuais-internet/$>. Acesso em: 20 maio 2020. 158

PENNEBAKER, J. W.; FRANCIS, M. E. **Linguistic inquiry and word count - LIWC**. 1996. Disponível em: <http://liwc.wpengine.com/>. Acesso em: 20 jan. 2020. 9

POZZI, F. A.; FERSINI, E.; MESSINA, E.; LIU, B. **Sentiment analysis in social networks**. Cambridge, MA: Morgan Kaufmann, 2016. 284 p. ISBN 9780128044384. 11, 15, 32, 33

RALEIGH, C. **Armed conflict location and event data project - ACLED**. 2010. Disponível em: <https://www.acleddata.com/>. Acesso em: 25 ago. 2020. 67

RAMOS, F. M.; BOLZAN, M. J. A.; Sá, L. D. A.; ROSA, R. R. Atmospheric turbulence within and above an amazon forest. **Physica D: Nonlinear Phenomena**, v. 193, n. 1, p. 278–291, 2004. ISSN 0167-2789. Anomalous distributions, nonlinear dynamics, and nonextensivity. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167278904000442>. 44

REIS, J.; GONÇALVES, P.; ARAÚJO, M.; PEREIRA, A.; BENEVENUTO, F. Uma abordagem multilíngue para análise de sentimentos. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING. **Proceedings...** Porto Alegre, RS, Brasil: SBC, 2015. ISSN 2595-6094. Disponível em: <https://sol.sbc.org.br/index.php/brasnam/article/view/6767>. 9, 147

REUTERS. **Reuters news agency:world's largest news agency**. 2019. Disponível em: <https://www.reuters.com/>. Acesso em: 08 dez. 2020. 2, 27, 38, 52, 125, 146

_____. **Iran and China sign 25-year cooperation agreement**. March 2021. Disponível em: <https://www.reuters.com/article/us-iran-china/iran-and-china-sign-25-year-cooperation-agreement-idUSKBN2BJ0AD?il=0>. Acesso em: 30 abr. 2021. 97, 123, 125

REYES-MENENDEZ, A.; SAURA, J.; ALVAREZ-ALONSO, C. Understanding #worldenvironmentday user opinions in twitter: a topic-based sentiment analysis approach. **International Journal of Environmental Research and Public Health**, v. 15, 11 2018. DOI:10.3390/ijerph15112537. 11, 15

ROSA, R. R.; NEELAKSHI, J.; PINHEIRO, G. A. L. L.; BARCHI, P. H.; SHIGUEMORI, H. Modeling social and geopolitical disasters as extreme events: a case study considering the complex dynamics of international armed; conflicts. In: SANTOS, L.; NEGRI R. G.AND CARVALHO, T. J. (Ed.). **Towards mathematics, computers and environment: a disasters perspective**. [S.l.: s.n.], 2019. p. 233–254. 1, 2, 18, 23, 40, 41, 42, 43, 44, 51, 76

SAGHA, H.; CUMMINSY, N.; SCHULLER, B. Stacked denoising autoencoders for sentiment analysis:a review. **WIREs Data Mining Knowl Discovy**, v. 7, n. 5, 2017. Doi: 10.1002/widm.1212. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1212>. 147

SAHA, S.; SAHA, M.; MUKHERJEE, K.; ARABAMERI, A.; THINGO, P. T.; PAUL, G. C. Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, reptree: a case study at the gumani river basin, india. **Science of The Total Environment**, v. 730, p. 139–197, 2020. DOI:https://doi.org/10.1016/j.scitotenv.2020.139197. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0048969720327145>. 21, 22, 25

SANTOS, C. dos; GATTI, M. Deep convolutional neural networks for sentiment analysis of short texts. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS: TECHNICAL PAPERS. **Proceedings...** Dublin, Ireland: Dublin City University and Association for Computational Linguistics, 2014. p. 69–78. Disponível em: <https://www.aclweb.org/anthology/C14-1008>. 7, 14

SANTOS, T.; FILHO, V. S. A.; ROCHA, V.; MENEZES, J. Os impactos do desmatamento e queimadas de origem antrópica sobre o clima da amazônia

brasileira: um estudo de revisão. **Revista Geográfica Acadêmica**, v. 11, p. 157–181, 11 2017. DOI:10.18227/1678-7226rga.v11i2.4430. Disponível em: <https://revista.ufrr.br/rga/article/view/4430>. 1

SCHMIDHUBER, J. Deep learning in neural networks: an overview. **Neural Networks**, v. 61, p. 85–117, Jan 2015. ISSN 0893-6080. Disponível em: <http://dx.doi.org/10.1016/j.neunet.2014.09.003>. 34

SINGH, K.; SHAKYA, H.; BISWAS, B. Happiness index in social network. **Advanced Informatics for Computing Research**, p. 261–270, 07 2017. 9

SIVARAJAH, U.; IRANI, Z.; GUPTA, S.; MAHROOF, K. Role of big data and social media analytics for business to business sustainability: a participatory web context. **Industrial Marketing Management**, v. 86, p. 163 – 179, 2020. ISSN 0019-8501. DOI:https://doi.org/10.1016/j.indmarman.2019.04.005. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0019850118305236>. 145

SOCHER, R.; BAUER, J.; MANNING, C. D.; NG, A. Y. Parsing with compositional vector grammars. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings...** Sofia, Bulgaria: Association for Computational Linguistics, 2013. v. 51, p. 455–465. Disponível em: <https://www.aclweb.org/anthology/P13-1045>. 5, 8, 13

SORNETTE, D. Endogenous versus exogenous origins of crises. In: ALBEVERIO, S.; JENTSCH, V.; KANTZ, H. **Extremes events in nature and society**. [S.l.]: Springer, 2006. p. 107–131. 2, 43, 61

SPACY. **Industrial-strength natural language processing**. 2019. Disponível em: <https://spacy.io/>. Acesso em: 20 maio 2019. 146

SUTTON, C. S.; BARTO, A. G. **Reinforcement learning: an introduction**. 2. ed. Londres: The MIT Press, 2012. 389 p. ISBN 9780262332767. 34

TABAK, F. S.; EVRIM, V. Comparison of emotion lexicons. In: 2016 HONET-ICT. **Proceedings...** [S.l.], 2016. p. 154–158. 9

TAVENARD, R.; FAOUZI, J.; VANDEWIELE, G.; DIVO, F.; ANDROZ, G.; HOLTZ, C.; PAYNE, M.; YURCHAK, R.; RUSSWURM, M.; KOLAR, K.; WOODS, E. Tslearn, a machine learning toolkit for time series data. **Journal of Machine Learning Research**, v. 21, n. 118, p. 1–6, 2020. Disponível em: <http://jmlr.org/papers/v21/20-091.html>. 50

THE UNIVERSITY OF WAIKATO. **Weka 3: data mining software in Java**. 2006. Disponível em: <$http://www.cs.waikato.ac.nz/ml/weka/index.html$>. Acesso em: 20 jan. 2017. 6

THELWALL, M.; BUCKLEY, K.; PALTOGLOU, G.; CAI, D.; KAPPAS, A. **SentiStrength**. 2010. Disponível em: <http://sentistrength.wlv.ac.uk/>. Acesso em: 02 mar. 2019. 9

TWITTER. **Twitter developer documentation - API**. 2017. Disponível em: <$https://dev.twitter.com/overview/api$>. Acesso em: 28 abr. 2017. 157

UNIVERSITY OF SOUTH ALABAMA. **Positive and negative skewness**. 2014. Disponível em: <https://www.ijamhrjournal.org/viewimage.asp?img=IntJAdvMedRes_2014_1_1_30_134449_u5.jpg>. Acesso em: 30 jan. 2021. 48

UOL. **EUA ataca síria com mais de 50 mísseis**. 2017. Disponível em: <https://noticias.uol.com.br/ultimas-noticias/ansa/2017/04/06/eua-ataca-siria-com-mais-de-50-misseis.htm>. Acesso em: 05 mar. 2019. 52

UPPSALA CONFLICT DATA PROGRAM - UCDP. **Uppsala Conflict Data Program**. 1980. Disponível em: <https://ucdp.uu.se/>. Acesso em: 06 ago. 2020. 67

VELHO, H. F. C.; ROSA, R. R.; RAMOS, F. M.; PIELKE, R. A.; DEGRAZIA, G. A.; NETO, C. R.; ZANANDREA, A. Multifractal model for eddy diffusivity and counter-gradient term in atmospheric turbulence. **Physica A: Statistical Mechanics and its Applications**, v. 295, n. 1, p. 219–223, 2001. ISSN 0378-4371. Proceedings of the IUPAP International Conference on New Trends in the Fractal Aspects of Complex Systems. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0378437101000772>. 44

VERONESE, T.; ROSA, R.; BOLZAN, M.; FERNANDES, F. R.; SAWANT, H.; KARLICKY‘, M. Fluctuation analysis of solar radio bursts associated with geoeffective X-class flares. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 73, n. 11, p. 1311–1316, 2011. ISSN 1364-6826. Influence of Solar Activity on Interplanetary and Geophysical Phenomena. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1364682610002907>. 45

VIEIRA, S. **O que é curtose e como se mede?** 2018. Disponível em: <http://soniavieira.blogspot.com/2018/11/curtose.html>. Acesso em: 30 jan. 2021. 49

VILARES, D.; ALONSO, M. A.; GÓMEZ-RODRíGUEZ, C. Supervised sentiment analysis in multilingual environments. **Information Processing & Management**, v. 53, n. 3, p. 595–607, 2017. ISSN 0306-4573. Disponível em: <$https://www.sciencedirect.com/science/article/pii/S0306457316302540$>. 15

WHEELER, D. J. **Problems with skewness and kurtosis, part two**. 2011. Disponível em: <https://www.qualitydigest.com/inside/quality-insider-article/problems-skewness-and-kurtosis-part-two-080111.html>. Acesso em: 30 jan. 2021. 47, 48

ZAFARANI, R.; ABBASI, M. A.; LIU, H. **Social media mining - an introduction**. USA: Cambridge University Press, 2014. 344 p. 28

ZAVATTARO, S. M.; FRENCH, P. E.; MOHANTY, S. D. A sentiment analysis of u.s. local government tweets: the connection between tone and citizen involvement. **Government Information Quarterly**, v. 32, n. 3, p. 333–341, 2015. ISSN 0740-624X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0740624X15000337>. 32, 151

ZHANG, Y.; SHANG, L.; JIA, X. Sentiment analysis on microblogging by integrating text and image features. In: CAO, T.; LIM, E.-P.; ZHOU, Z.-H.; HO, T.-B.; CHEUNG, D.; MOTODA, H. (Ed.). **Advances in knowledge discovery and data mining**. Cham: Springer, 2015. p. 52–63. ISBN 978-3-319-18032-8. 8, 14, 32, 151

ANEXO A - APLICAÇÕES DE ANÁLISE DE SENTIMENTO

A.1 *Sentiment analysis applied to analyze society's emotion in two different context of social media data*

Abstract

In the last few decades, the growth in the use of the Internet has generated a substantial increase in the circulation of information on social media. Due to the high interest of several areas of society in the analysis of these data, a study of better techniques for the manipulation and understanding of this type of data is of great importance so that this enormous volume of information can be interpreted quickly and accurately. Based on this context, this study shows two approaches of sentiment analysis to verify the emotion of the population in different context. The first approach analyses the positive and negative sentiment about 2018 presidential elections in Brazil considering data from the Twitter social network. The second approach performs analysis of data from social media to identify threats sentiment level of armed conflicts considering data off the conflict between Syria and the USA in 2017. To achieve this goal, machine learning techniques such as auto-encoder and deep learning will be considered in conjunction with NLP text analysis techniques. The results obtained show the effectiveness of the approaches used in the classification of sentiment within the domains used according to the methodology developed for this work.

Keywords: Machine Learning, Deep Learning, Auto-encoder, Natural Language Processing, Sentiment Analysis, Social Media

A.1.1 Introduction

The evolution of the Internet has enabled the advent of social media as one of the main means of circulation of personal, political and dissemination information. As the consequence, the amount of information generated daily in these means of communication gradually increases annually. This excessive amount of information has drawn the attention from different areas of knowledge who have realized the importance of using this information in a suitable way to verify the opinions and sentiment of the population in a certain subject. In this way, the need to use modern techniques, as machine learning and sentiment analysis, is identified to assist in the precise verification of specific information among these huge volumes of data (DHAWAN; ZANINI, 2014), (KUMARI, 2016), (SIVARAJAH et al., 2020).

In accordance with this need, this paper presents two approach of applying sentiment analysis to verify the opinion of the population on two main themes: the presidential

elections in Brazil in 2018 considering data from social networks and an analysis of social media data for the identification of threats of armed conflicts considering data on social media about the conflict between Syria and the USA in 2017. Emphasizing that sentiment analysis is a machine learning area that allows the identification of human emotions in texts, images, sounds, etc. considering a given domain (AYLIEN, 2016).

For the analysis of the election for president of Brazil, data were collected in Portuguese from the social network Twitter candidates Bolsonaro and Haddad. For the analysis of tweets, a vocabulary was built with words related to the choice of the topic. Each vocabulary word was polarized at 0 (negative word) and 1 (positive word) considering the theme of the election as a domain of analysis. In the analysis of the armed conflict, from the news collected, the one that best represented the idea of threats among government officials in the countries of the analyzed conflict was selected and which served as a basis for calculating the similarity with the others analyzed news. Syrian conflict data was collected on social networks around the world and in English, such as Reuters (REUTERS, 2019), CNN (CNN, 2020), The Guardian (GUARDIAN, 2020), etc.

The analysis performed on the data collected on the social network Twitter generated as results the identification of the percentage of positive and negative messages (tweets) for each candidate analyzed. This analysis allowed to identify which candidate was more favorable to win the elections. The analysis of information from social media, on the other hand, allowed to identify the percentage or level of threat among those involved in the conflict. This result can be used to identify how these threats favored the start of the armed conflict.

To achieve the results in the work, applied sentiment analysis together with Natural Language Processing (NLP) and machine learning techniques. For the application of NLP, the word embedding APIs NLTK (BIRD et al., 2009), Tensorflow (GOOGLE, 2015) and SpaCy (SPACY, 2019) were used. For the application of machine learning techniques such as Auto-encoder, Deep Auto-encoder and Deep Learning, was used the Deep Learning Keras API (CHOLLET, 2015).

The objective of this work is to show two sentiment analysis approaches using two different public information sources, as social networks and social media. These approaches identify the positive/negative sentiment and threat degree emotions expressed in that information sources by the considered population.

A.1.2 Related works

This section presents the state of the art of some of the main work related to the emotion analysis in social network and social media. There is no intention in this work to make a detailed bibliographic review of the published articles.

In 2015, the article A Multilingual Approach for Sentiment Analysis, (REIS et al., 2015), presented a comparison between different sentiment analysis tools using nine different languages: Portuguese, French, Spanish, Italian, Turkish, Russian, Arabic, Dutch, German and English. The database was initially in English and than was translated into the other languages using the Python Goslate API for translation. The sentiment analysis tools used in the comparison were: Linguistic Inquiry and Word Count (LIWC), SentiStrength, SentiWordNet, SenticNet, SASA - SailAil Sentiment Analyze, Happiness Index, PANAS-t - Positive Affect Negative Affect Scale, NRC Emotion Lexicon, NRC Hashtag Sentiment Lexicon, Sentiment140 Lexicon, OpinionLexicon, VADER - Valence Aware Dictionary for sEntiment Reasoning and Emoticons. In the analysis it was found that the best accuracy of the tools was for the English language.

In 2016, (GHOSH et al., 2016) published a work entitled A novel deep learning architecture for sentiment classification that proposed a hybrid deep learning architecture with a two-layer Boltzmann Restricted Machine (RBM) and a Probabilistic Neural Network (PNN) for the classification of sentiments.
In the first stage, RBM performs the dimensionality reduction. In the next step, the PNN performs the classification of sentiments. The work tested five different sets of data and compared the results with current works. The proposed method showed the best precision of 93.3 %, 92.7 %, 93.1 %, 94.9 % and 93.2 % for the Movies, Books, DVD, Electronics and Utensils approaches kitchen, respectively.

In 2017, (SAGHA et al., 2017) published the article Stacked Denoising Auto-encoders for Sentiment Analysis: A review in which different types, topologies and learning methods that use Auto-encoders in the analysis of sentiment in texts multi-domain and multilingual. The main subjects approach in the article are presented as follows:

- overview of the state of the art of Sentiment Analysis, highlighting the application Auto-encoders (AEs), Denoising Auto-encoder (DAs) and Stacked Denoising Auto-encoders (SDAs).

- use of marginalized SDAs (mSDAs) and their variants, such as: Heterogeneous Hybrid Transfer Learning (HHTL) and Stacked Instance Denoising Auto-encoders.
- comparison between the models presented.

The author concludes that SDAs can be further deepened and that cross-cultural analysis is an area to be further explored in sentiment analysis.

In 2018, (ALI et al., 2018) show a study that analyzes the different approaches that Eastern and Arab media apply to crisis events. In the work, the messages published on Twitter about the November 2015 terrorist attack in Beirut and Paris were used as a case study. In the analysis, 2390 tweets were used with the sentiment of sympathy classified for the training of a regression model of a deep convolutional neural network. This model was also used to predict the sentiment of sympathy for the upcoming crisis events. Three forms of analysis were used: bias coverage, which verifies if there was any difference in the coverage or volume of news in both countries; the News media sympathy bias that analyzes how the degree of sympathy of the messages could affect the population and information propagation that verified if the sentiment of positive sympathy spread in the messages. As a result, it was found that both countries had similar coverage, 79% accuracy in predicting the sentiment of sympathy was also achieved, and it was observed that the retweets were impartial in terms of whether a tweet was sympathetic or not.

In 2019, (HAO et al., 2019) show a study which present on how the avoidance of Hong Kong residents' actions in relation to tourism and how these actions are affected by socioeconomic factors. This analysis was carried out using 72,755 newspaper news from 2003 to 2015. In order to carry out the analysis, a sentiment analysis framework was developed in the work in news published in the media in Chinese language. This framework uses SVM and NB techniques to classify news. The prediction results of this analysis are: for the SVM (Support Vector Machine) classifier (accuracy = 0.913; F-measure = 0.914); for the classifier NB (Naive Bayes) (accuracy = 0.839, F-measure = 0.840).

In 2020, (GARVEY; MASKAL, 2020) presents a study on how the negative influence of the news published in the news media influences the perception of society in relation to the understanding of the use of Artificial Intelligence in society, specifically in the area of health. The work uses quantitative and qualitative approaches to analyze news published on AI in the period from 1956 to 2018. For this purpose, the Google

Sentiment analysis tool, the Cloud Natural Language API, was used. As a result, the work concluded that the data analyzed failed to support the theory that the negative sentiment of the media about AI had any influence on its use.

A.1.3 Theoretical foundation

This section shows briefly the theory used in developing this work. The section A.1.3.1 explains how Natural Language Processing (NLP) applies to this work. The section A.1.3.2 presents the sentiment analysis technique related to this work. The section A.1.3.3 shows some important definition of machine learning, related to this work, with the techniques: MultiLayer Perceptron, Auto-encoder and Deep Auto-encoder.

A.1.3.1 Natural language processing - NLP

Natural Language Processing or NLP is a sub-area of artificial intelligence and enables the development of systems that allow computer-human iteration using human natural language whether by text or by speech. ems that allow computer-human iteration using human natural language whether. The NLP can be divided into the following stages (JACKSON; MOULINIER, 2002) which are detailed below.

- Tokenization is very well characterized in artificial languages as programming languages, since they do not have much ambiguity. However, in natural language the same character may have several meanings depending on its context. In this way, the tokenization was used considering the approach **Languages Delimited by Spaces** - like the European languages, word boundaries are only indicated by the insertion of blank spaces. However, the symbols between the spaces are not necessarily the tokens needed for the next processing. This is due to the ambiguous nature of writing systems as to the range of tokenization conventions that exist.

- Lexical Analysis performs text analysis at the level of the word. A basic task of logical analysis is to relate morphological variants to their respective keywords. These words are found in a keyword dictionary grouped together with their semantic and syntactic information. In the case of the analysis of natural language, logical analysis dismantles the words of a sentence in its grammatical components (noun, adjective, pronoun, etc.).

- Syntactic Analysis is the task of recognizing a sentence and assigning it a syntactic structure. These syntactic structures are attributed by Context

Free Grammar (CFG). The application of CFG by means of specific algorithms generates a representation in a tree structure. These trees analyze an important intermediate state of representation for semantic analysis.

- Semantic Analysis is the representation of intermediate meanings. They are only composed of linguistic expressions. These representations of meanings are only attributed to phrases based on the knowledge acquired with the logical and grammatical phases. Thus, this type of analysis is used in the understanding of the meaning of a sentence. It is also widely used for elimination of ambiguities.

- Pragmatic Analysis is the last stage of the analysis of natural language. In this phase, the meaning is elaborated on the basis of contextual knowledge and logical form and is mapped to the final language of representation of the subject matter. This type of analysis is also used to validate the semantic analysis. In this analysis it is considered that words can be associated by meanings (water, swimming) or subject proximity (water, well).

In this work, the tokenization stage was used to eliminate unwanted characters from social media news and Twitter messages (tweets), along with the machine learning techniques present in the NLTK (BIRD et al., 2009) and Tensorflow (GOOGLE, 2015) APIs. NLTK is an open source platform for creating Python pro-grams to work with human language data. The platform has an easy to use interface with the word processing library defined for classification, tokenization, derivation, markup, semantic reasoning and analysis, and libraries related to Natural Language Processing (NLP) (BIRD et al., 2009).

A.1.3.2 Sentiment analysis

The Sentiment Analysis area refers to the tasks of analysis, identification and classification of all information that is characterized in an emotional way, a subjective way, or an opinion generating information, be it information in text, image or sound format (CUADRADO; GóMEZ-NAVARRO, 2011). For the accomplishment of these characterization tasks, it is usually used the Natural Language Processes statistics and/or machine learning methods. According to (CUADRADO; GóMEZ-NAVARRO, 2011), these tasks are usually divided as follow.

- Subjectivity Classification: deals with the identification of parts of the texts that have a sense of subjectivity.

- Polarity Classification: determines that fragments of texts are classified as positive or negative sentiments.

- Intensity Classification: works with the emotional intensity expressed in the text. This type of approach is usually divided into classes: strongly positive, positive, neutral, negative or strongly negative sentiments.

- Sentimental Analysis Based on Topics or Features: is the verification of existing features related to sentiments about the subject matter.

- Opinion Mining: is related to the retrieval of information from a query. Thus, it allows to consult a specific topic and to classify it in a certain category.

In recent years, Sentiment Analysis has been applied in different ways of expressing sentiments in alternative textual forms, mainly through information from social networks. In the current literature of the area, there are works that identify sentiments through sounds, using emoticons from social networks (BALABANTARAY et al., 2012), and images (ZHANG et al., 2015). In addition, Sentiment Analysis is also being used in the most diverse sectors of society. In the last few years, the network and the social media sector has turned its attention to the study of the sentiments of the population and the government. In many countries, the sentiments of the population are being studied in relation to the policies and programs launched by the different areas of their governments (municipal (county), state and federal) (ARUNACHALAM; SARKAR, 2013) (ZAVATTARO et al., 2015).

In the work of (CEPIK, 2003) it is emphasized that a way of controlling intelligence activities is the knowledge of the public opinion. In Brazil, research is already being carried out in the area involving the view of society in relation to the public programs and government policies. One of these surveys is the work (ANTUNES et al., 2014) that analyzes the hashtags of social networks to assess the population's adherence, by state, to the campaign of the Aedes Aegypti mosquitoes that transmits the dengue virus among other diseases.

This work uses the sentiment analysis Polarity Classification to analyze election and sentiment analysis based on topics or features to analyze armed conflict data.

A.1.3.3 Machine learning

The Machine Learning deals with the computational algorithms that allow learning and, consequently, its improvement through the repetition of experiments. Within the Machine Learning there are several applications ranging from Data Mining that allows the discovery of general rules in a large volume of data, to systems that automatically learn the needs of one user (HAYKIN, 2008).

a) **MLP - MultiLayer perceptron**

In the Multilayer Perceptron network each unit performs a weighted sum of their inputs and transmit that level of activation through a transfer function to produce an output. The network therefore has a simple interpretation as an input-output model, with weights and biases as free model parameters. Such networks can model functions of arbitrary complexity with the number of layers and the number of units in each layer determining the complexity of the function (MITCHELL, 1997) (HAYKIN, 2008). The learning of the MLP network is performed using the backpropagation algorithm. This algorithm is based on two basic steps.

- Propagation, in this step an input pattern is presented and its result is propagated, layer by layer. The synaptic weights are fixed and at the end a set of network output is released.
- Backpropagation, in this step the network output is compared to the output that is desired to calculate the error correction parameter. The weights are adjusted according to the result of the error correction parameter calculation. This adjustment is applied, layer by layer - from the output layer to the input layer (HAYKIN, 2008).

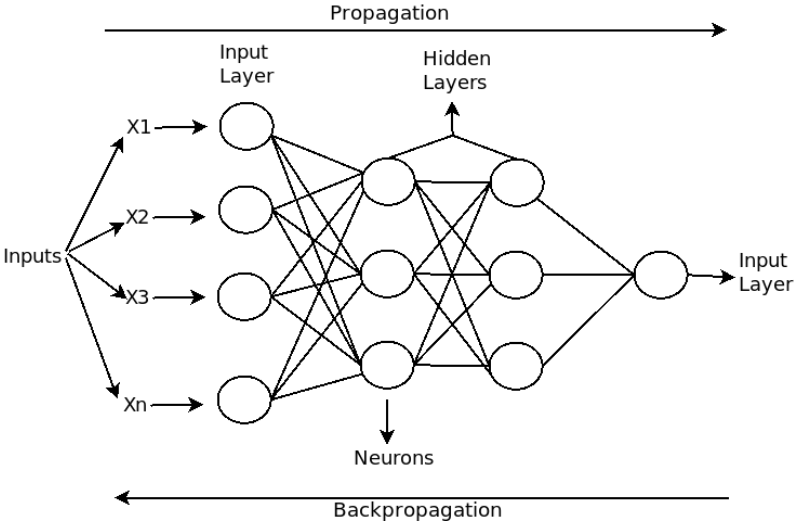
The figure A.1 shows an example of a computational model for the MLP network.

The MLP neural network was applied in this research to classify the positives and negatives sentiments in tweets about running candidates to the 2018 Brazilian presidential election.

b) **Deep learning**

Deep Learning consists of a feed forward neural network that has a large depth related to the number of layers between the input of the network and its output. Feed forward neural networks aim to map a function

Figura A.1 - MultiLayer Perceptron Representation.



Source:Adapted of Haykin (2008).

$y = f(x, \theta)$ and learn the value of the parameter that represents the best fitting of the function (GOODFELLOW et al., 2016). In the Deep Learning model the first layers are used in an unsupervised way, and later values in the next layers are used as initial values for supervised learning. Several knowledge areas that are using Deep Learning, Natural Language Processing (NLP) are making a significant increase in its use. The use of Deep Learning in the NLP area made a great optimization of information processing possible, as can be seen in (AYLIEN, 2016).

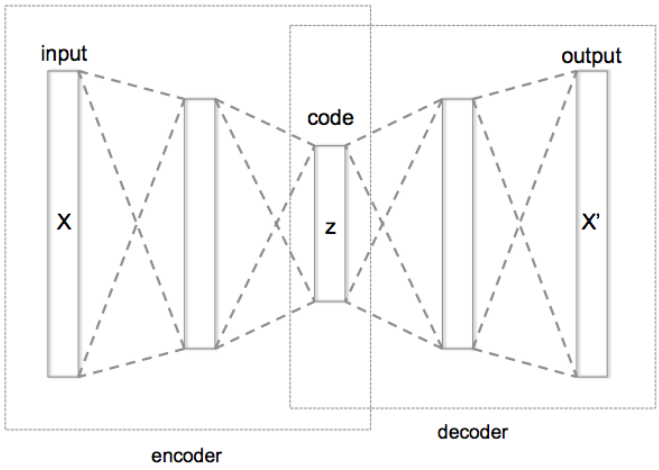
In this work, the Deep Learning is used by SpaCY API to calculate similarity levels in news about armed conflict.

c) **Auto-encoder** Auto-encoder is a learning algorithm that uses a neural network to represent the dimensionality reduction of its input in its output. According to (GOODFELLOW et al., 2016), this algorithm internally has a hidden layer z that describes the code used to represent data entry. Self-encryption consists of two parts.

- The Encoder function is defined by $z = f(x)$ and compresses the input into a latent space representation.
- The Decoder function is defined by $r = g(h)$ and which reconstructs the output of this representation.

The Figure A.2 shows an example of the neural network structure Auto-encoder.

Figura A.2 - Structure of Auto-encoder neural network.



Source: Adapted of Le (2015).

The auto-encoder is typically used for dimensional reduction or feature learning (HINTON; SALAKHUTDINOV, 2006). For training one can use both the Feed-forward neural networks and the Backpropagation (HAYKIN, 2008). According to (GOODFELLOW et al., 2016) there are various types of Auto-encoder.

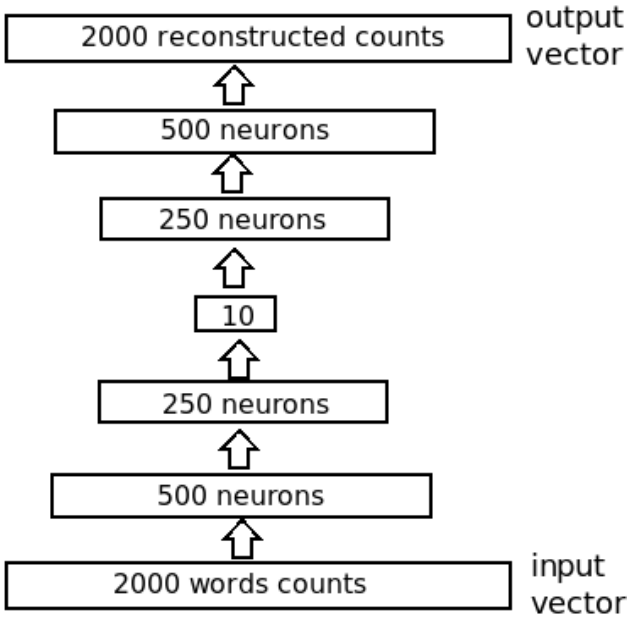
- Sparse Auto-encoder, it has a training criterion involving a scattered penalty (h) on the code z layer. In which we have, $L(x, g(f(x))) + h$, where $g(h)$ is the decoder output and $h = f(x)$ is the encoder output. Sparse Auto-encoder is typically used for sorting tasks (LE, 2015).
- Denoising Auto-encoder, it receives corrupted data as input and is trained to reproduce the original data and not corrupted as a result. This Auto-encoder minimizes the equation $L(x, g(f(x)))$, where x represents a copy of data that has been corrupted by some form of noise. Denoising Auto-encoder is commonly used for the retrieval of information with noise (LE, 2015).
- Convolutional Auto-encoder, it differs from conventional Auto-encoders, since their weights are shared between places at the en-

trance, preserving the spatial location. The reconstruction is therefore due to a linear combination of small parts of the data based on the initial code. Convolutional Auto-encoder is typically used for image pattern recognition (LE, 2015).

- Deep Auto-encoder, it is an algorithm consisting of an input layer, multiple layers of encoding and multiple layers of decoding. It can be pre-trained as a stack of single-layer Auto-encoders. It is commonly used for the dimensionality reduction. The Figure A.3 presents a example of the Deep Auto-encoder structure.

The Figure A.3 presents a example of the Deep Auto-encoder structure.

Figura A.3 - Structure of Deep Auto-encoder neural network.



Source: Adapted of Le (2015).

The Auto-encoder and the its variation Deep Auto-encoder were used in this work to detect features in tweets. These features are applied in the MLP neural network to classification of positive and negative sentiments about candidates to the Brazilian presidential elections.

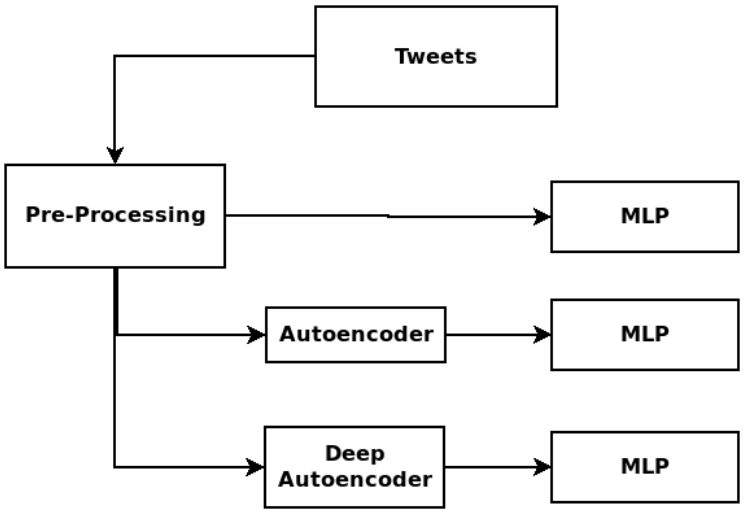
A.1.4 Methodology

The methodology applied in this work was divided in two phases. The first phase, showed in the Figure A.4, was developed to analyze the Brazilian president 2018 election to identify positives and negatives sentiments of the population. The second phase, showed in Figure A.5, was to analyze data reference to Syrian and the USA armed conflict to identify threats degree about this conflict in the public media news. These methodologies are detailed in the next sections.

A.1.4.1 Phase 1: Methodology applied to analyze the brazilian president 2018 election

The methodology developed for analyzing the Brazilian president 2018 election data, was collected from Twitter. It consists of the following steps: collect from tweets (Twitter messages) about Bolsonaro and Haddad candidates, during the period from June to September, 2018. Pre-process these tweets to eliminate characters and symbols without meanings. Then, processing of the resulting data applying the MLP network, Auto-encoder - MLP and Deep Auto-encoder - MLP. These processing result in the percentage of positive and negative sentiment found in tweets published about each candidates. Figure A.4 presents the scheme of this methodology.

Figura A.4 - Methodology for analyzing the data collected from Twitter about Brazil's presidential election in 2018.



a) **Tweets** The data used in this work were collected from the Twitter social network. Twitter has an API that allows you to easily and freely collect the

tweets (TWITTER, 2017). To validate the developed methodology in this work, 35,000 tweets were collected for each presidential candidate who participated in the second round of the Brazilian election in the period of June to September, 2018.

An example of the tweets referring to candidate Bolsonaro used in this work is shown as follows.

```
#Brasil #BolsonaroPresidente
Agora e a hs pra varrer essa quadrilha do poder,
nem fome
#goBolsonarogo #Bolsonaro @jairbolsonaro #BrasilDecide
Doria mostrando seu apoio a #BolsonaroPresidente
falou por todos nos!
#Bolsonaro17  #ELESIM
#eleicoes2018
#BrasilAcimaDeTudo #DeusAcimaDeTodo
Quem estiver ao lado de Lula sera derrotado.
#PTNuncaMais
#PTNAO
#ForaPT
CORRUPTO - SAFADO - PRESO
```

An example of the tweets referring to candidate Haddad used in this work is shown as follows.

```
#HaddadPresidente
#Haddad13
#eleicoes2018 #elenao
TUDO, MENOS O BOLSONARO!!!!!!
Vamos divulgar as PROPOSTAS do Haddad.
#AgoraEHaddad
So #Haddad13 pode nos livrar de
tao mitologica tragedia
Professores contam por que querem Haddad
presidente do Brasil
```

```
Rejeição das mulheres:
Haddad 49\% x 41\% Bolsonaro.
```

---

b) **Pre-Processing** The tweets were collected in the Portuguese language. To pre-process this text this phase was divided in two steps, as follows.

- Tokenization, in which one eliminates characters that have no meaning in the text, such as α, #, http,:, ?, !.
- Word Embedding, in which one generates the vector embedded words that contains each word of the set collected tweets. In this vector each word is presented in the form of an integer, which represents the number of times a word is repeated in the set of analyzed tweets.

  To generate embedding words vector, one uses the Embedding Tensorflow API function, which comes from an open source library for machine learning for a wide variety of tasks (GOOGLE, 2015), rather than using ready-made techniques such Word2vec, Sense2vec, etc. Such choice was made with the intention to obtain the representation of the word set in the form that were collected without modifications in its meaning. After Embedding words, the values in the vector were normalized considering the maximum value existing in this vector.

c) **Processing** In this phase one uses machine learning techniques MLP, Auto-encoder and Deep Auto-encoder. Also, one uses $1 - gram$ method to classify sentiment in each word of tweets. The $1 - gram$ method that considers one word to do the analysis of the sentiment involved in the classification. A Portuguese vocabulary of 1000 words labelled in positive sentiment equal one and negative sentiment equal zero, showed a sample in Table A.1, generated by a work team and related with election, it was also used to verify if the sentiment in all the analyzed tweets were positive or negative for a given presidential candidate. In this vocabulary the most used slang in social networks (PADILHA, 2020) were considered, and information about emoticons were not considered. The next sections show details about the use of this learning techniques.

d) **The MLP** In this work, the Keras API was used to generated the multilayer perceptron neural network to obtain the classification of the collected tweets. Keras is an open source API that works with high-level neural

Tabela A.1 - Sample of Vocabulary with Words Classified in Sentiment Positive or Negative.

Word	Sentiment
abraçar	1
acabou	0
aceita	1
briga	0
brincadeira	0
bugada	0
fracasso	0
fraudes	0
furto	0
oportunistas	0
oposição	0
opressão	0
organizar	1
orgulho	1
sancionar	1
sangrou	0
sapão	1
saúde	1
segurança	1
votos	1
zelar	1
zoar	0

networks, written in Python and allows to develop applications together with the TensorFlow API (an end-to-end open source machine learning platform) (GOOGLE, 2015). It allows a quick experimentation, that is, to focus on the result in the shortest possible time (CHOLLET, 2015). This data set was divided in training, validation and test, considering the percentage of 60%, 20%, and 20%, respectively. As truth of the MLP the Portuguese vocabulary of 1000 words specific of the election domain was used. So each tweet is processed and its words are classified in positive or negative words considering the vocabulary created for this work. After, the twitter is classified according with how many positive or negative words were found in the message. The architecture of the MLP is presented in Table A.2.

Tabela A.2 - The Architecture of MLP Neural Network applied in the work.

Activation Function	sigmoid
Epochs	100
Learning Rate	0.0001
Hidden Layers	2
Hidden Layers Neuron Numbers	3

e) **The Auto-encoder - MLP**

The Keras API again was used to constructed one auto-encoder to dimensionality reduce the collected tweets vector. The result of the auto-encoder application was a word vector with similar features. The word vector was applied in the MLP to classify the sentiment towards a given candidate. The MLP Network is the same one used in the classification with the original data. The architecture of the Auto-encoder is presented in Table A.3.

Tabela A.3 - The Architecture of Auto-encoder - MLP Neural Network applied in the work

Activation Function	encoder=RELU decoder=sigmoid
Epochs	100
Learning Rate	0.0001
Hidden Layers	1
Hidden Layers Neuron Numbers	encoder=3000 decoder=35000

f) **The Deep Auto-encoder - MLP**

The Keras API was used to constructed one deep auto-encoder to dimensionality reduce the collected tweets vector. The intention of using deep auto-encoder was to present how the increase of layers in the auto-enconder can improve its performance. The result of the deep auto-encoder application was a word vector with similar features. The word vector also was applied in the MLP to classify the sentiment towards a given candidate. The MLP Network is also the same one used in the classification with the original data. The architecture of the deep auto-encoder is presented in Table A.4.

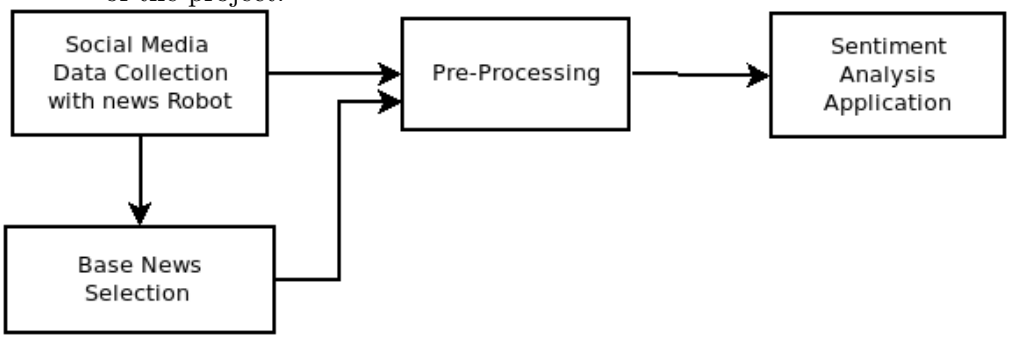
Tabela A.4 - The Architecture of Deep Auto-encoder - MLP Neural Network applied in the work

Activation Function	encoder=RELU decoder=sigmoid
Epochs	100
Learning Rate	0.0001
Hidden Layers	encoder = 3 decoder = 3
Hidden Layers Neuron Numbers	encoder=3000,1000,500 decoder=1000,3000,35000

A.1.4.2 Phase 2: Methodology applied to analyze the armed conflict social media data

The methodology developed for analyzing the armed conflict Social Media data collected consists of the following steps: collect news about armed conflict between Syrian and the USA from media news, pre-processing of the news to eliminate characters and symbols without meanings, processing of these data with the MLP network, Auto-encoder and Deep Auto-encoder techniques and and result degree of threats found in public media news exchanged between rulers of Syria and the USA. Figure A.5 presents the scheme of this methodology.

Figura A.5 - Methodology flow chart of the phases sequence followed for the development of the project.



a) **Social Media Data collection with News Robot** The social media are responsible for publishing the events that occur daily in the world. Indeed, news are published of Heads of States referring to each other in an aggressive manner. Many of these verbal aggression may often culminate in some form of conflict. In this work, data is collected from the most

relevant official social media in the world supported by news agencies (e.g. Reuters) in the English language. It is stored in a .csv file, generated for this work, considering the information of date and the internet access address (URL). For the speed-up of the collection process, news trackers are freely available, such as the News-bot (NEWSBOT, 2019). It is considered to search for words related to the subject such as threats, conflicts, weapons, death, among others.The robot collected 40 news related to the domain of threat of armed conflict. All those related to the conflict between Syria and the USA in the period from January 2016 to April 2017. Thus, each url of news stored in the .csv file are read, and pre-process to calculate the threats level. The threat level is a value between 0 and 1, and corresponds to the similarity percent of the text with base news. Table A.5 show one sample of .csv file structure of collected news in social media.

Tabela A.5 - Sample of .csv file structure with Date and URL of the Collected News.

Date	**URL**
01/03/16	reuters.com/article/us-mideast-crisis-syria-israel..
14/03/16	theguardian.com/world/2016/mar/14/syria-chemical-weapons ..
11/08/16	nytimes.com/2016/08/12/world/middleeast/syria-chlorine..
11/08/16	amnesty.org/en/latest/news/2016/08/syria-fresh-chemical..
25/08/16	csis.org/analysis/unpacking-syrias-chemical-weapons-problem..
13/09/16	bellingcat.com/news/mena/2016/09/13/chemical-attacks-syria..
16/09/16	foreignpolicy.com/2016/09/16/chemical-weapons-watchdog..
13/03/17	time.com/4699178/us-troop-increase-syria-raqqa-isis..
15/03/17	theguardian.com/world/2017/mar/15/syria-conflict-study..
05/04/17	politifact.com/truth-o-meter/article/2017/apr/05/revisiting..

b) **The Base news selection** In this work, a text file is generated from the news collected from social media by the chat bot. It is used to calculate the similarity. This domain is a base news selected by the News-bot robot, considering words that make some kind of reference to armed conflicts. Some of the words, like arms, attack, force, threat, weapons, tanks and conflict were presented as a reference for the robot news search. This text contains terms that express the emotion of threat among the leaders of the countries and was constructed using some of the collected news. Following part of the text used in the Syria and the USA armed conflict work is shown.

```
Even though Moscow is in fact targeting Syria opposition and
rebels killing thousands of civilians and destroying
infrastructure such as hospitals, water plants, bakeries,
and schools it can justify hitting areas where Nusra is
present because the groups forces cannot be separated from
other rebel factions. The US Secretary of State John Kerry,
at that time, pursued a joint command and control centre to
co-ordinate US and Russian aerial operations.
Instead of working with long-established groups inside
Syria, the US military and CIA have tried a series of
programmers of vetting,training and equipping moderate
fighters outside the country, all of which have failed.
Meanwhile, the US blocked any assistance to anti-Assad
factions in southern Syria, and detached from the critical
battlefront in the north-west, where Russia and the Assad
regime have laid siege to opposition-held areas of Aleppo
city. The power of the blended rebel forces was made clear
when a rebel offensive turned the tide in the battle
for Aleppo. The US could belatedly recognise the folly of
its artificial labels and establish lines of co-operation
with the groups inside Syria.
```

---

c) **The Pre-processing** In the pre-processing phase of data collection from social media on Syria/US armed conflict, the tokenization technique was applied to eliminate characters that have no meaning for the text, such as accent and punctuation characters (α, #, http,: ,?,!,.). The tokenization process is applied to the base text and social media news using NLTK (BIRD et al., 2009) API.

d) **The Sentiment analysis application**

After the creation of the .csv file with the news collected and organized, the process of sentiment analysis the news begins. The URL information contained in the .csv file is used to read each news item in real time. As this reading is done via Web using the BS4 library of the Python 3.7 programming language, one access a .html (Hypertext Markup Language) file, which contains the news information within the $< p >< /p >$ paragraphs tags. The result of this process is a text containing the news information to be analyzed.

The treatment of the information contained in the text is processed using the concepts of Natural Language Processing using the tokenization step described in Section A.1.3.1 for the elimination of symbols and characters that have no meaning representation for the text. After the tokenization of the information, the news analysis phase begins by applying the concepts of Sentiment Analysis through the library SpaCy (Industrial-Strength Natural Language). SpaCy is a free open source library for advanced Natural Language Processing in Python. SpaCy was designed specifically for use in production and helps create applications that process and understand large volumes of text. It can be used to build information extraction or natural language understanding systems, or to pre-process text for Deep Learning.

The project uses Sentiment Analysis based on topics or resources for extracting information from selected news. In this process of extracting information, a basic text was initially defined, which is considered to represent 100% of threats to be verified. This basic text is defined empirically, considering the knowledge of the people on the subject addressed. After the base news selection, an analysis is made of the percentage of similarity of the new data news to be analyzed within the base news using the SpaCy library. The result of this process is the degree of threat that each news item has in relation to the extreme event analyzed. This degree of threat is calculated and stored for each news item accessed and stored in the Similarity (%) field of the .csv file. The described process is illustrated in Figure A.6.

The results this methodology's application are shown in the Section A.1.5.2.

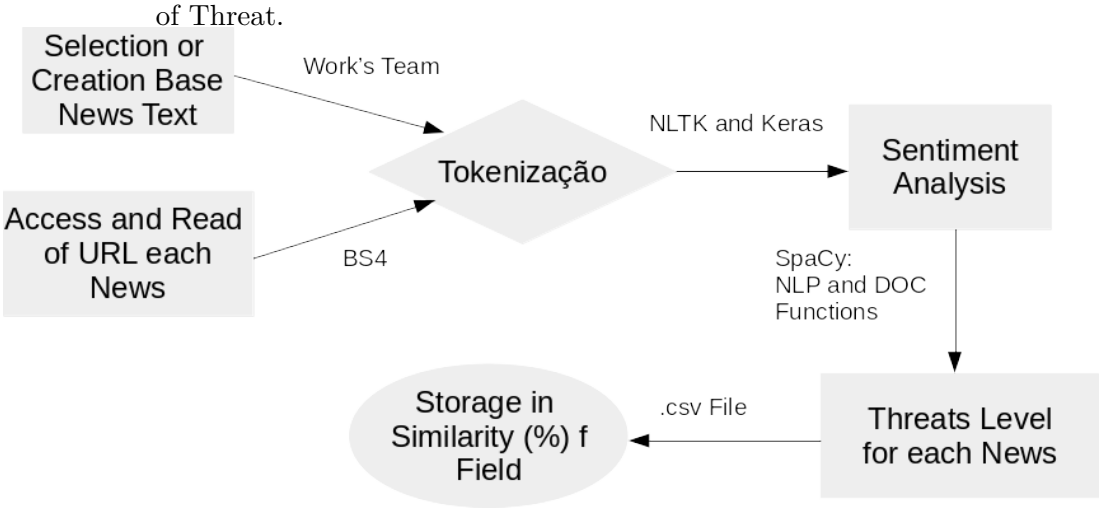
A.1.5 Results

The results of the analysis of Twitter data for the 2018 presidential election are presented in the section A.1.5.1. The results of the analysis of social media data on armed conflicts are presented in the section A.1.5.2

A.1.5.1 Data analysis results of the twitter for the 2018 presidential election

The results obtained in this work about the analysis of the unbalanced Twitter data were related to the application of MLP machine learning techniques, Auto-

Figura A.6 - Methodology to Application of Sentiment Analysis for Calculating the Level of Threat.



encoder and Deep Auto-encoder in the same database of the candidates Bolsonaro and Haddad, from the second round of the 2018 Brazilian presidential elections. The combination of these analysis made it possible to compare such techniques and confirm the population's sentiment in relation to the candidates in the election. The results of this comparison are presented separately by candidates and groups of techniques.

In Table A.6 the results of the processing of collected data for the candidate Bolsonaro are shown. It is observed that the percentage of positive sentiment, which represents the number of words expressing positive information, in the text about the candidate Bolsonaro is always greater than the negative sentiment for the three techniques analyzed. It is also verified that the technique of Auto-encoder combined with MLP has presented a better classification of the positive sentiment.

The Table A.7 presents the results of the processing of the data collected for the Haddad candidate. It is observed that the percentage of negative sentiment, which represents the number of words expressing negative information, is always greater than the positive sentiment for the three techniques analyzed. It is also verified that the technique of Auto-encoder combined with MLP was also the one that presented a better classification of the negative sentiment.

The results of the Table A.6 demonstrate that the candidate Bolsonaro has a higher percentage of positive than negative sentiments in the analysis using the three machine learning techniques: MLP (50,199% positive and 49,801% negative), Auto-

Tabela A.6 - Comparison of the results of the application of machine learning techniques MLP, Auto-encoder-MLP and Deep Auto-encoder-MLP of the candidate Bolsonaro data.

	BOLSONARO	
Machine Learning Techniques	**Sentiment Positive (%)**	**Sentiment Negative (%)**
MLP Representation	50.199	49.801
Auto-encoder-MLP Representation	53.339	46.661
Deep Auto-encoder-MLP Representation	51.394	48.606

Tabela A.7 - Comparison of the results of the application of machine learning techniques MLP, Auto-encoder-MLP and Deep Auto-encoder-MLP of the candidate Haddad data.

	HADDAD	
Machine Learning Techniques	**Sentiment Positive (%)**	**Sentiment Negative (%)**
MLP Representation	46.614	53.386
Auto-encoder-MLP Representation	44.622	55.378
Deep Auto-encoder-MLP Representation	45.618	54.382

econder-MLP (53,339% positive and 46,661% negative) and Deep Auto-encoder-MLP (51,394% positive and 48,606% negative). The result of the Table A.7 demonstrate that the candidate Haddad has a higher concentration of negative than positive sentiments in the analysis using the three machine learning techniques: MLP (46,614% positive and 53,386% negative), Auto-encoder-MLP (44,622% positive and 55,378% negative) and Deep Auto-encoder-MLP (45,618% positive and 54,382% negative). Thus, it is noteworthy that the results presented are consistent with the true result of the election, which had the candidate Bolsonaro as winner. In the actual analysis the sentiments of sarcasm and irony in the text were not considered. In the results, it was observed that the increase number of words in the vocabulary can be better for the classification of tweets either in positive or in negative. To verify the results obtained in the classification, the accuracy was calculated for data training of each neural network model used in this work. These results are presented in

Table A.8. It is observed that the Deep Auto-encoder-MLP technique had a better accuracy in the classification of the information with the value of 78.5 %. While the Auto-encoder-MLP techniques had 74.3 % and MLP had 59.2 % for unbalanced data inside of the presented context.

Tabela A.8 - Showing of the accuracy results of the application of machine learning techniques MLP, Auto-encoder-MLP and Deep Auto-encoder-MLP.

Machine Learning Techniques	**ACCURACY**
MLP Representation	0.592 (59.2%)
Auto-encoder-MLP Representation	0.743 (74.3%)
Deep Auto-encoder-MLP Representation	0.785 (78.5%)

A.1.5.2 Data analysis results of the social media about armed conflict

This section presents the results obtained by analyzing the news from social media according to the methodology presented in the section A.1.4.2. The news were collected from January 2016 to April 2017. The end date of the collection refers to the eve of the launch of the 59 Tomahawk missiles by the USA with destination to Syria. The launch day was not considered in the analysis to give an idea of the degree of threat that preceded the attack.

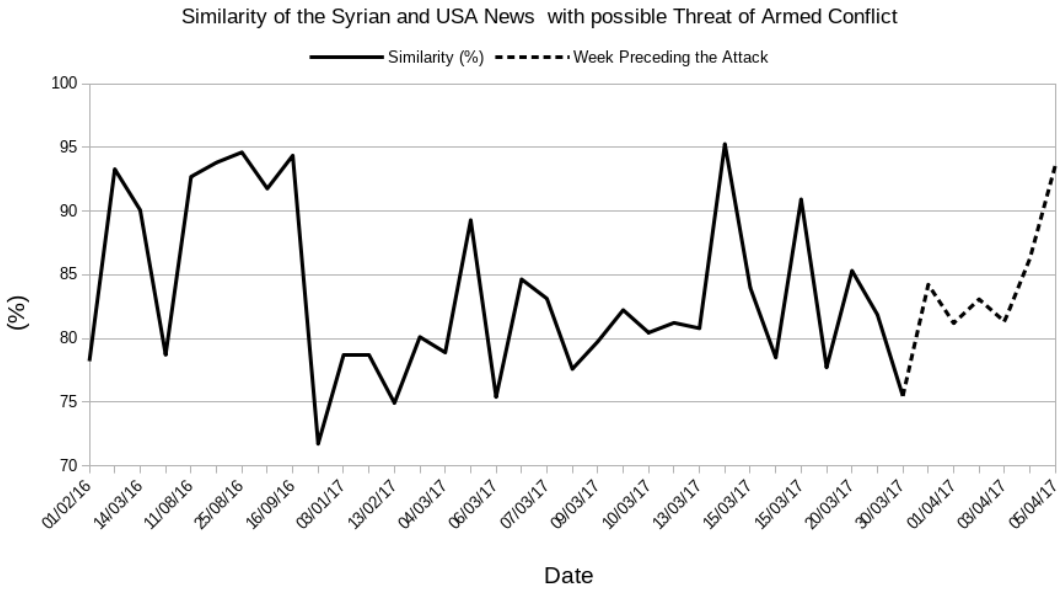
The Table A.9 presents the results with values greater than or equal to 90% of the similarity. The Date column shows the publication of the news date, the URL column contains the information of which social media the news was published in and the column Similarity(%) shows the similarity's percentage of the news analyzed with the base news with high level of threat.

It was analysed 40 news about threat between Syrian and the USA until the launch of Tomahawk missiles from the Mediterranean Sea to Syria. The Figure A.7 presents the time evolution of the news threats during the analysed period. The spikes in the graphic represent the biggest values of similarity. Observe that graphic values of threats precede the missile attack by one week, in the period between March 30, 2017 and April 5, 2017 (dashed line), were increasing until the last analysed news. That is the signature for hostilities commence and the launching of the 59 Tomahawks.

Tabela A.9 - Sample of Data analysis of the result of the armed conflict between Syrian and the USA to determine the probability of threats.

Date	URL	Similarity (%)
01/03/16	reuters.com/us-mideast-crisis-syria-israel..	93.30
14/03/16	theguardian.com/syria-chemical-weapons ..	90.08
11/08/16	nytimes.com/middleeast/syria-chlorine..	92.71
11/08/16	amnesty.org/syria-fresh-chemical..	93.82
25/08/16	csis.org/syrias-chemical-weapons...	94.62
13/09/16	bellingcat.com/chemical-attacks-syria..	91.78
16/09/16	foreignpolicy.com/chemical-weapons-watchdog..	94.37
13/03/17	time.com/us-troop-increase-syria-raqqa-isis..	95.27
15/03/17	theguardian.com/syria-conflict-study..	90.93
05/04/17	politifact.com/apr/05/revisiting..	93.62

Figura A.7 - Time series of the threat level in the analyzed news of the conflict between Syria and the USA.



A.1.6 Conclusion

This work shows two approaches for sentiment analysis application to help understand emotional opinions of a population. The first approach data from the second round of the 2018 Brazilian presidential election collected from the social network Twitter was compared with analysis produced by the machine learning techniques MLP, Auto-encoder and Deep Auto-encoder. The concept of polarized classification was used in this work to analyze the sentiments of the population in relation to

the candidates Bolsonaro and Haddad. The results showed that although a direct analysis of the texts was carried out, without considering the sarcasm and irony commonly found in the Portuguese language, satisfactory results were obtained in relation to the real performance of candidates in the election. The use of the $1-gram$ method allowed an analysis of the context of the information used and not only of the isolated tweets. This first approach was applied in Portuguese language to verify presidential election results in Brazil using social network Twitter data and polarization technique of NLP and obtained the accuracy of 59,2% for MLP, 74,3% for Auto-encoder MLP and 78,5% for Deep Auto-encoder MLP.

The second approach, was applied in social media data in English language to verify the level of threats in the armed conflict between Syrian and the USA. This analysis use Sentimental Analysis Based on Topics or Features by choice of the one news defined as base news. The analysis of these data used a news base to compare with others news collected by a robot news collector and applying in spaCy NLP Python API. The result shows that the level of threats increased in the days before the attack of April 5, 2017 with a percent of similarity by sentiment threats between 70% and 95%.

For future work, we intend to extend the application of $N - gram$ concepts, to use the identification of sarcasms and irony in the Portuguese language texts, to increase vocabulary to improve the result of classification, and to use balanced Twitter data to better classification of information in social media. Also, it is intended to generated a public armed conflict portfolio with content of the .csv file, to apply citizen science to help the selection of news base, and to apply this methodology to predict future armed conflicts.

ANEXO B - TEXTO BASE PARA ANÁLISE DA EMOÇÃO DE AMEAÇAS

Na aplicação da técnica de análise de sentimento, construiu-se um texto base para a definição do conceito de ameaça que se gostaria que se encontrasse nos textos das notícias analisadas. Para cada evento extremo abordado, foi gerado um texto base. Assim, na Seção B.0.1, apresenta-se o texto base para cada evento extremo social analisado, considerando os conflitos armados entre a Síria e USA, Irã e USA e Global e na Seção B.0.2, apresenta-se um texto base para os eventos extremos naturais, considerando a análise de ocorrência de secas, queimadas e desmatamento para os anos de 2015, 2016, 2017, 2018, 2019 e 2020.

B.0.1 Texto base para os eventos extremos sociais

Para os eventos extremos sociais foram gerados um texto base diferente para cada conflito armado analisado. Assim, na Seção B.0.1.1, apresenta-se o texto base do conflito armado entre a Síria e USA. Na Seção B.0.1.2, apresenta-se o texto base do conflito armado entre Irã e USA e na Seção B.0.1.3, apresenta-se o texto com a definição da emoção de ameaça, para o conflito armado global.

B.0.1.1 Texto base para o conflito armado entre Síria e USA

As fighting over Syria s last rebel-held stronghold intensifies and puts US ally Turkey in direct conflict with Russia there is a growing chorus for the US to do something about the dire humanitarian crisis Nearly one million people have been displaced since December by Syrian President Bashar al Assad s offensive into the Idlib province backed by Russian air power and Iranian--commanded forces Relief organizations are struggling to respond to the overwhelming need amid freezing emperatures and a lack of basic resources like tents But for the four million civilians in Idlib there is no where to go as Assad s forces back them closer to the border with Turkey which remains shut as Turkey struggles to host nearly four million Syrian refugees already Turkey and the Syrian forces it arms and backs have been pushing to hold Assad s offensive at bay with direct clashes on Thursday But while the Trump administration has condemned the offensive and vocalized support to Turkey advocates including the last US ambassador to Syria are urging for US intervention

What s happening in Idlib is the worst case scenario we have worried about in Syria since 2011 We never wanted it to come we hoped it wouldn t come and it s here said Ambassador Robert Ford who was forced to leave Syria amid threats from Assad when he served as US envoy from 2011 to 2014 He continued This is not just another problem in Syria The scale is much greater than anything the world has seen in recent decades More than 900000 people have fled their homes or shelters in Idlib according to the United Nations most of whom were previously displaced by Syria s now nine-year old civil war Approximately 80% of the displaced are women and children facing severe winter weather conditions like snow and freezing temperatures The military offensive has killed hundreds of civilians according to war monitoring groups But the freezing temperatures have killed seven children according to the humanitarian group Save the Children with the Associated Press reporting at least 10 people have died We are striving to save lives but the space for these efforts is shrinking Filippo Grandi the high commissioner for refugees at the U N said Thursday That s in part because aid organizations themselves have also been forced to flee including Huzayfa al Khateeb a Syrian radio reporter and volunteer relief worker He was forced to flee from his home in western Ghouta outside Damascus three years ago but in Idlib in recent days he s been forced to live in his car while so many sleep outside or in the rubble of schools hospitals and other buildings bombed by Assad s forces and Russia There is no place safe in Idlib and in any place Every day they are bombing us al Khateeb said They cant live here because the Assad regime is bombing us Russian regime and they can t there is no single town no single area you can live For Joumane Mohamad and her two young children they are lucky enough to have a house to rent But the psychological toll of the bombing has deepened even after nine years of war have forced them to move to 10 different houses across the country I hope that you will let the world know what sufferings what pain what frustration and what inhuman conditions that we are living here now she said during a briefing call with al Khateeb organized by Refugees

International an advocacy organization We feel as if the world has betrayed us as if the world has abandoned us Mothers are suffering the most because of all the burdens and responsibilities and fears they have to deal with Mohamad said adding every morning feels as if it could be our last meeting She said she tried to smuggle her family including her nine-year old and her nursery-school age son across the border into Turkey once but was scared away by gunfire Turkey has all but shut down its border in recent months overwhelmed by the 3 6 million Syrians it already hosts according to U N data Instead Turkey President Recep Tayyip Erdogan has threatened to use military force to repel Assad s offensive if it isn t halted Erdogan has armed trained and backed rebel forces throughout Syria s civil war and entered into agreements with Russia and Iran to create safe zones in areas including Idlib which Assad Russia and Iran have later seized US special envoy for Syria James Jeffrey traveled to Ankara last Wednesday to meet with senior Turkish officials and offer US assistance for its NATO ally Before the trip Jeffrey told reporters in Washington the US was looking at the various things we can do to halt the offensive including more sanctions but hinted at no immediate action That s not enough according to activists who rallied on Capitol Hill Thursday What s also more outrageous is the lack of outrage that we re seeing across the United States and in the Western world said Mouaz Moustafa executive director of the Syrian Emergency taskforce an advocacy group No one is talking about what s unfolding there no one is speaking out for children some of whom have been burned to a cinder He and Ford called on the Trump administration and Congress to take some prompt steps to halt the offensive including finding funds for humanitarian groups to deal with the unprecedented need press Russian officials to halt their support and back Turkey in what way they can President Donald Trump once claimed credit for saving Idlib by tweeting his opposition to a previous Assad offensive in 2018 But while he told Assad and Russia Don t do it in December his tweet was undercut by his clear unwillingness

to take action to stop any fighting that he does not see as
Let Syria and Assad protect the Kurds and fight Turkey for
their own land he tweeted in October I hope they all do
great we are 7000 miles away But Ford warned that whatever
comes next the crisis will spill out of Syria just as Assad
and Russia s assault on Aleppo led to mass refugee flows into
Europe that bolstered far-right politics across the continent
While the Turkish-Syrian border is closed for now al Khateeb
and Mohamad both said there are plans to storm the border
because they say it may be Syrians only option to survive
Even if it kills 1000 people the rest will save their lives
and cross the border and be in safety al Khateeb said He
added This is what people are thinking about
because there is no other chance no other choice for us

B.0.1.2 Texto base para o conflito armado entre Irã e USA

The United States Army has released a bleak assessment of its 2003
invasion of Iraq and subsequent attempts to defeat a Sunni Muslim
insurgency until a 2011 withdrawal claiming that neighboring Iran
was the only true winner of the operation the decision to attack
was a preemptive response to accusations that Iraqi President
Saddam Hussein possessed weapons of mass destruction and offered
tacit support to the AlQaeda militant group that conducted the
attacks of 2001 These charges later proved to be false and the
Army has now admitted that Husseins fall managed to empower a
mutual foe of both the Iraqi leader and the US President Donald
Trumps view that a conflict with Iran would be a short war was an
illusion and that his threat of obliteration amounted to
threatening genocide Citing a number of troubling and escalatory
indications from Iran Bolton said the US was deploying warships
to the Middle East to send a clear and unmistakable message that
it would meet any Iranian attacks on US interests with unrelenting
force He did not offer evidence of the threat from Iran which
promptly dismissed his warning as psychological warfare He also
threatened to resume highlevel uranium enrichment in 60 days if
the pacts remaining signatories ? the United Kingdom Russia
France China Germany and the European Union failed to protect

Irans oil and banking industries from US sanctions People here worry about a military confrontation and that means the money the government spends on the military is very much justified because there is a serious threat he said People understand that the Rouhani government has been trying to reduce tensions with the US People here worry about a military confrontation and that means the money the government spends on the military is very much justified because there is a serious threat he said People understand that the Rouhani government has been trying to reduce tensions with the US As tensions rise a group of European thinktank leaders said the US must rejoin the nuclear deal if it wants to achieve its stated aim of reining in Irans regional activities and its missile programme Tensions between Iran and the United States could force the US military to keep up a costly deployment of troops and hardware in the Middle East for years to come to protect oil shipping and Persian Gulf allie Scenarios in which Iran is able to cause damage to the global oil market that lasts years and fundamentally reshapes the environment are highly unlikely ut given Irans demonstrated ability to hit Saudi oil infrastructure with precision the US military could be forced to maintain a large force in the region with air and missile defense systems and US naval ships at the ready the report said Such a deployment over a period of years would carry a high price tag and derail plans by the Pentagon to shift its focus away from the Middle East to countering China A spate of attacks on tankers and on a key oil hub in Saudi Arabia since May already has prompted the US to deploy an additional 14000 troops to the Middle East with two squadrons of fighter aircraft and air and missile defense systems sent to Saudi Arabia.

B.0.1.3 Texto base para o conflito armado global

India?s top military commander has said a tense border standoff with Chinese forces in the western Himalayas could spark a larger conflict, even as senior commanders from both sides met near the front line for their eighth round of talks in recent months. Chief of Defence Staff Bipin Rawat said the situation was tense at the Line of Actual Control, the de facto border

in eastern Ladakh where thousands of Indian and Chinese troops are locked in a months-long confrontation. Indian officials say China is assisting rebel groups that have stepped up attacks on its border with Myanmar in recent months, opening another front in the conflict between two nations already engaged in a deadly standoff in the Himalayas. Armed groups in Myanmar -- including the United Wa State Army and the Arakan Army, which was designated a terrorist organization this year -- are acting as Beijing?s proxies by supplying weapons and providing hideouts to insurgent groups in India?s northeastern states, according to Indian officials with knowledge of the situation, who asked not to be identified due to rules for speaking with the media. The British mission?s spokesperson described the Indo-Pacific as ?increasingly important for the UK, as it is at the centre of global economic growth and a region of increasing geostrategic importance?. The UK has arange of enduring security interests in the region?, the spokesperson said. The findings reveal that while Indian security analysts give serious attention to China?s nuclear policy and capabilities, Chinese analysts maintain a dismissive attitude about the relevance of nuclear weapons in China-India relations. The attitude stems from a widely held view that India?s indigenous military technologies are significantly behind China?s and that China will continue widening the gulf between the two countries? conventional and nuclear capabilities. However, Chinese analysts do not appear to fully appreciate the long-term destabilizing implications of this growing gap. India may feel pressure to build out its nuclear arsenal, and this could further threaten the fragile stability between India and Pakistan. Chinese experts tend to underestimate the role Beijing may have in shaping New Delhi?s threat perception and nuclear strategy. China has deployed large numbers of troops and weapons along a disputed Himalayan border in violation of bilateral agreements, India?s foreign ministry has said, accusing Beijing of escalating tensions and triggering a deadly clash last week. China has blamed

India for the clash ? the deadliest between the two nuclear-armed neighbours for at least five decades ? and said Indian troops attacked Chinese officers and soldiers. China blamed India for the violent clashes and insisted it did not want the situation to escalate. Chinese Foreign Ministry spokesman Zhao Lijian said on Wednesday the border situation was ?stable and controllable?. Indian fighter jets have roared over a flashpoint Himalayan region as part of a show of strength even as China again accused India of provoking the border clash earlier this month that killed at least 20 soldiers. Chinese troops had taken ?defensive measures and determinedly counterattacked against the Indian side?s violent actions, successfully protecting national sovereignty and territorial integrity?, the Chinese ministry said on its social media account. The increased activity along the Myanmar border has sparked concern in New Delhi that India?s military is becoming stretched as tensions remain with China and Pakistan on other parts of its land border, which runs for roughly 14,000 kilometers (8,700 miles). The officials said India moved several battalions consisting of about 1,000 troops each into the Myanmar border area after a soldier was killed in an ambush on Oct. 21. India has accused China of violating the border consensus reached between them during recent peace talks. At least 20 Indian soldiers died in clashes with Chinese troops in the region in June. China has not said if its soldiers also died. The two nuclear powers accused each other of crossing the poorly demarcated border and provoking the fight. China has denied that its troops violated the status quo. "Chinese border troops have always strictly observed the Line of Actual Control and have never crossed the line. The border troops of the two countries have been in communication over territory issues," China’s foreign ministry spokesman Zhao Lijian said. Indian and Chinese soldiers have engaged in skirmishes in which shots have been fired for the first time in over four decades.

Some 20 Indian and an unknown number of Chinese troops]were killed in a particularly violent clash along the border in June.

B.0.2 Texto base para os eventos extremos naturais

A temporada de incêndios de 2020 na floresta amazônica pode ser muito mais rigorosa do que em 2019 e um dos motivos dessa piora são as mesmas condições climáticas que intensificam a temporada de furacões no hemisfério norte segundo pesquisadores Em agosto passado uma série de grandes incêndios provocados pelo homem na Amazônia lançou nuvens de fumaça sobre a cidade de São Paulo transformando o dia em noite e gerou protestos internacionais Mas embora esses incêndios tenham sido incomuns e alarmantes a situação poderia ter sido ainda pior se a Amazônia estivesse em época de estiagem Mas neste ano infelizmente condições mais secas do que a média são exatamente o que está previsto para o sul da Amazônia e um dos motivos é o aumento extraordinário de calor no Atlântico Tropical Norte a milhares de quilômetros de distância Esse calor oceânico também fez com que a temporada de furacões no Atlântico batesse recordes logo de início um prenúncio das previsões para uma temporada extraordinariamente tumultuada Algumas pesquisas sugerem existir uma relação causal entre os próprios furacões e os piores anos de incêndios na Amazônia embora isso seja assunto de maior debate Acredito que o oceano esteja acentuando ambos os fenômenos afirma Chris Landsea pesquisador meteorologista do Centro Nacional de Furacões da Administração Nacional Oceânica e Atmosférica dos Estados Unidos Está provocando anos de furacões no Atlântico e ao mesmo tempo aumentando a vulnerabilidade a incêndios na Amazônia A temporada de incêndios de 2020 na floresta amazônica pode ser muito mais rigorosa do que em 2019 e um dos motivos dessa piora são as mesmas condições climáticas que intensificam a

temporada de furacões no hemisfério norte segundo
pesquisadores Em agosto passado uma série de grandes
incêndios provocados pelo homem na Amazônia lançou
nuvens de fumaça sobre a cidade de São Paulo
transformando o dia em noite e gerou protestos
internacionais Mas embora esses incêndios tenham sido
incomuns e alarmantes a situação poderia ter sido ainda
pior se a Amazônia estivesse em época de estiagem Mas
neste ano infelizmente condições mais secas do que a
média são exatamente o que está previsto para o sul da
Amazônia e um dos motivos é o aumento extraordinário de
calor no Atlântico Tropical Norte a milhares de
quilômetros de distância Esse calor oceânico também fez
com que a temporada de furacões no Atlântico batesse
recordes logo de início um prenúncio das previsões
para uma temporada extraordinariamente tumultuada
Algumas pesquisas sugerem existir uma relação causal
entre os próprios furacões e os piores anos de
incêndios na Amazônia embora isso seja assunto de
maior debate Acredito que o oceano esteja acentuando
ambos os fenômenos afirma Chris Landsea pesquisador
meteorologista do Centro Nacional de Furacões da
Administração Nacional Oceânica e Atmosférica dos
Estados Unidos Está provocando anos de furacões no
Atlântico e ao mesmo tempo aumentando a
vulnerabilidade a incêndios na Amazônia Geralmente
queimadas são um recurso utilizado por produtores
rurais na Amazônia para limpar a terra para uso na
agropecuária embora muitos incêndios também sejam
ateados por invasores em florestas públicas em uma
tentativa de se apossar de novas terras Infelizmente
tudo indica que este será mais um péssimo ano para o
desmatamento escreveu por e-mail Barlow E ao
contrário de 2019 essas queimadas para limpeza de
terra provavelmente serão agravadas por um clima
mais árido do que o habitual o que significa
que podem se espalhar mais rápido tornar-se

incontroláveis e até mesmo se alastrar para a mata virgem Aliás essas previsões de sazonalidade indicam que grandes extensões da Amazônia podem ser tornar áridas com o avanço da estação seca compreendida entre junho e novembro Um dos motivos disso são as temperaturas oceânicas distantes ao norte componentes essenciais que fundamentam a previsão de incêndios desenvolvida por Morton Segundo Yang Chen cientista de Geociências da Universidade da Califórnia em Irvine que desenvolveu a previsão juntamente com Morton as temperaturas no Atlântico Tropical Norte estão atualmente muito acima da média Quando essa região do oceano está especialmente quente é deslocada para o norte na Zona de Convergência Intertropical uma massa de ar de baixa pressão que gera tempestades intensas e muita precipitação nos trópicos Se essa massa de chuva se afastar mais em direção ao norte antes da chegada da estação seca no sul da Amazônia ela antecipará a estação seca e provocará uma estiagem acima da média Em anos anteriores quando o Oceano Atlântico Tropical Norte estava quente em 2005 e 2010 foram registrados recordes de estiagem na Amazônia explica Morton E com essas estiagens vieram os incêndios As águas quentes do Atlântico Tropical Norte também alimentam furacões que transferem umidade para o oeste e depois para o norte nos ventos predominantes em vez do sul Aliás uma pesquisa publicada por Morton e Chen em 2015 revela que temporadas ativas de furacões no Atlântico e temporadas intensas de incêndios na Amazônia andam lado a lado Embora haja uma correlação de ambos os fenômenos com o calor no Atlântico Tropical Norte eles possuem uma correlação mais nítida entre si Morton acredita que isso indica uma relação causal entre os dois fenômenos Quando tempestades tropicais e furacões

são formados captam a umidade que de outra forma circularia em direção ao continente sul-americano e desviam essa umidade à Costa do Golfo e costa leste dos Estados Unidos Em outras palavras retiram a umidade da Amazônia Chen está menos convencido de que os furacões do Atlântico desencadeiem diretamente uma estiagem na Amazônia embora concorde que ambos os fenômenos compartilham a mesma causa ou seja o excesso de calor no Atlântico Tropical Norte e seu impacto nos padrões climáticos Landsea do Centro Nacional de Furacões dos Estados Unidos também não está convencido de uma relação causal direta entre uma maior quantidade de furacões no Atlântico e a seca amazônica Ele ressalta que furacões são fenômenos muito transitórios Duram apenas alguns dias e respondem apenas por um pequeno porcentual de precipitação no Caribe Entretanto ele concorda que há certamente uma associação entre os dois fenômenos De qualquer modo a temporada de furacões de 2020 deve servir como alerta para a Amazônia já ocorreram seis tempestades tropicais dignas de receber nomes no Atlântico a essa altura um recorde para a temporada iniciada apenas em 1o de junho E é esperada uma intensificação na atividade de furacões à medida que o verão avança e o calor aumenta em todo Atlântico Tropical Nossa expectativa é que será uma temporada bastante agitada afirma Landsea Após uma estação de chuvas com índices bem abaixo de níveis históricos, grandes partes da Amazônia estão mais secas que o normal, gerando preocupações sobre um aumento ainda maior de incêndios e desmatamento na estação seca que se aproxima. Dados da NASA e do Instituto Nacional de Pesquisas Espaciais (Inpe) apontam para condições climáticas adversas: o solo está mais seco, as temperaturas estão mais altas, e a as águas subterrâneas estão escassas. A temporada de chuvas deste ano, que ocorreu entre dezembro e fevereiro, está entre as 10 piores já

registradas, com apenas 75% da chuva usual
para a estação. Dados dos satélites Grace, da NASA,
mostram que o leste da Amazônia e o Cerrado têm muito
menos águas subterrâneas armazenadas do que o normal.
Dados dos aquíferos são usados para ajudar a prever
secas globalmente. Imagem cortesia da NASA e do Centro
Nacional dos Estados Unidos para Mitigação de Secas.
No ano passado, desmatamento e incêndios florestais na
Amazônia atingiram os níveis mais altos em uma década.
Mas as chamas foram contidas pela umidade natural da
floresta após diversos meses de chuvas intensas.
Este ano, a floresta pode não ter a mesma sorte. ?É
muito preocupante. Pode ser que haja mais ocorrências
de incêndios do que vimos no ano passado, e mais cedo?,
diz Daniel Nepstad, ecologista florestal e presidente
do Earth Innovation Institute que estudou resistência
florestal a secas, à Mongabay. Desde que chegou à
Presidência, em janeiro de 2019, Jair Bolsonaro tem
reduzindo proteções ambientais e tentado mudar a
legislação do país para abrir terras indígenas
para mineração e outras atividades que são o principal
motor de destruição da Amazônia.
O governo também diminuiu operações de repressão a
crimes ambientais em meio à pandemia do novo
coronavírus, abrindo as portas para o aumento do
desmatamento, da mineração ilegal e da grilagem. Como
resultado, o desmatamento aumentou 51% no primeiro
trimestre, comparado com o mesmo período de 2019, de
acordo com o Inpe. A taxa de desflorestamento de
agosto de 2019 até o final de março de 2020 é quase
duas vezes maior em relação ao mesmo período no ano
anterior, segundo o Inpe. Dados mostram todo o bioma
amazônico com temperaturas mais altas do que a média de
janeiro a março de 2020. Mapa produzido pelo Centro
Nacional dos Estados Unidos para Informação Ambiental.
Especialistas alertam que um clima mais seco pode ser
letal à região, intensificando uma retração florestal

e permitindo que áreas intocadas peguem fogo no longo prazo. Embora as previsões climáticas para a Amazônia sejam imprevisíveis, o que pode reverter o déficit atual, modelos climáticos do Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) não indicam precipitações acima da média para os próximos meses. ?Estamos entrando em um período de estiagem em que as chuvas [geralmente] não têm intensidade suficiente para impactar a chance de fogo e outros fatores relacionados à estação seca?, diz o climatologista Diego Jatobá, do CPTEC, à Mongabay. A escassez de água no solo logo após um período de chuvas também sinaliza problemas nos próximos meses, diz o cientista da NASA e hidrólogo Matthew Rodell, que trabalha no projeto de monitoramento por satélite Gravity Recovery and Climate Experiment Follow-On (Grace-Fo).
Os dados de níveis da água acumulados abaixo da superfície da Terra são usados para prever secas em todo o mundo. Imagens recentes da América do Sul feitas pelo projeto mostram grandes áreas da Amazônia e do Cerrado na zona vermelha, o que significa que os lençóis freáticos estão severamente impactados. ?Se virmos uma precipitação que varia de normal a média neste ano, então há potencial para seca?, diz Rodell à Mongabay em uma entrevista por telefone. ?Eu estaria preocupado?. Para Ane Alencar, diretora de ciência do Instituto de Pesquisa Ambiental da Amazônia (Ipam) e especialista em incêndios florestais, um aumento em incêndios este ano depende muito mais de decisões políticas do que de condições climáticas, as quais ela espera que fiquem dentro da média. ?Minha expectativa é a pior possível em relação ao descontrole do desmatamento e do fogo?, diz Alencar à Mongabay. ?Autocombustão não existe na Amazônia?. O El Niño, evento climático cíclico que causa intensas secas na Amazônia, é muito improvável de ocorrer este ano,

acrescentou Alencar. ?Em um ano de El Niño, um
incêndio induzido vira 100. Não é isso que vamos
ver este ano?, diz. Uma nota técnica publicada pelo
Ipam esta semana também antecipa um salto
em incêndios causados pelo aumento do desmatamento
em 2019, tendo em vista que o modus operandi de
grileiros consiste em desmatar em um ano e atear
fogo no ano seguinte para transformar a floresta
em fazendas e pastos.
As condições de solos mais secos se sobrepõem a
altas taxas de desmatamento. No Pará, palco de 40%
de todo o desmatamento na Amazônia no ano passado,
incêndios precoces podem ocorrer, indicam mapas
climáticos. Em uma área em torno da cidade de
Santarém, que já tem as taxas mais altas de
desmatamento na Amazônia, incêndios e queimadas
para limpeza de terras também podem aumentar após
a recente pavimentação da rodovia BR-163, que
corta a floresta amazônica. A localização da obra
foi descrita no Twitter por Bolsonaro como ?no
coração do agronegócio?. ?Concluiu-se o
asfaltamento da BR-163 na região. Com isso,
talvez haja uma movimentação de especuladores de
terra. Por ter menos chuva, fica
mais fácil desmatar. O maquinário entra com mais
facilidade?, diz Marcos Heil Costa, um dos principais
cientistas climáticos do Brasil e pesquisador do
Painel Intergovernamental das Nações Unidas sobre
Mudanças Climáticas (IPCC). A seca gradual da
Amazônia tem preocupado cientistas há décadas, tendo
em visa a ameaça do início de um ciclo de retração
que pode levar a floresta para um caminho sem volta,
transformando-a em uma savana. Nos últimos anos,
eventos climáticos extremos, como secas e enchentes,
aumentaram e dados publicados mostram estações mais
secas cada vez mais longas e intensas. Mas números
envolvendo o período chuvoso da Amazônia são novos.

Dados do CPTEC enviados à Mongabay mostram que oito dos 10 períodos de chuva mais secos na região norte registrados desde 1962 ocorreram após os anos 2000. ?Temos observado uma tendência negativa, uma redução significativa de chuvas ao longo dos anos?, diz Jatobá à Mongabay. Apesar disso, ainda há esperança para reverter o cenário atual.

De acordo com Nepstad, alguns passos já estão sendo dados por uma parte do agronegócio para estimular agricultores a plantarem árvores. ?Agricultores se comprometeram a plantar (árvores) em 6 milhões de hectares nos próximos 20 anos, mas frequentemente não têm dinheiro ou incentivos para isso.

Essa é parte da solução?, diz. Segundo ele, há intervenções administrativas que podem impedir esse grave cenário de retração florestal?.

ANEXO C - ANÁLISES ESTATÍSTICAS PARA OS PARÂMETROS DE MÉDIA, VARIÂNCIA, CURTOSE E ASSIMETRIA PARA A QUANTIDADE DE 8 E 32 SÉRIES

C.1 Eventos extremos sociais - conflitos armados entre Síria e USA, Irã e USA e Global

Nesta seção são apresentados as análises realizadas para os parâmetros de média, variância, curtose e assimetria para os conflitos armados entre a Síria e USA, Irã e USA e Global, considerando 32 séries de 512 elementos e 8 séries com 2048 elementos. Essas análises são apresentadas nas Figuras C.1, C.2, C.3, C.4, C.5 e C.6, respecitvamente.

Figura C.1 - Análise dos parâmetros de média, variância, curtose e assimetria considerando 32 séries com 512 elementos para o conflito armado entre a Síria e USA.

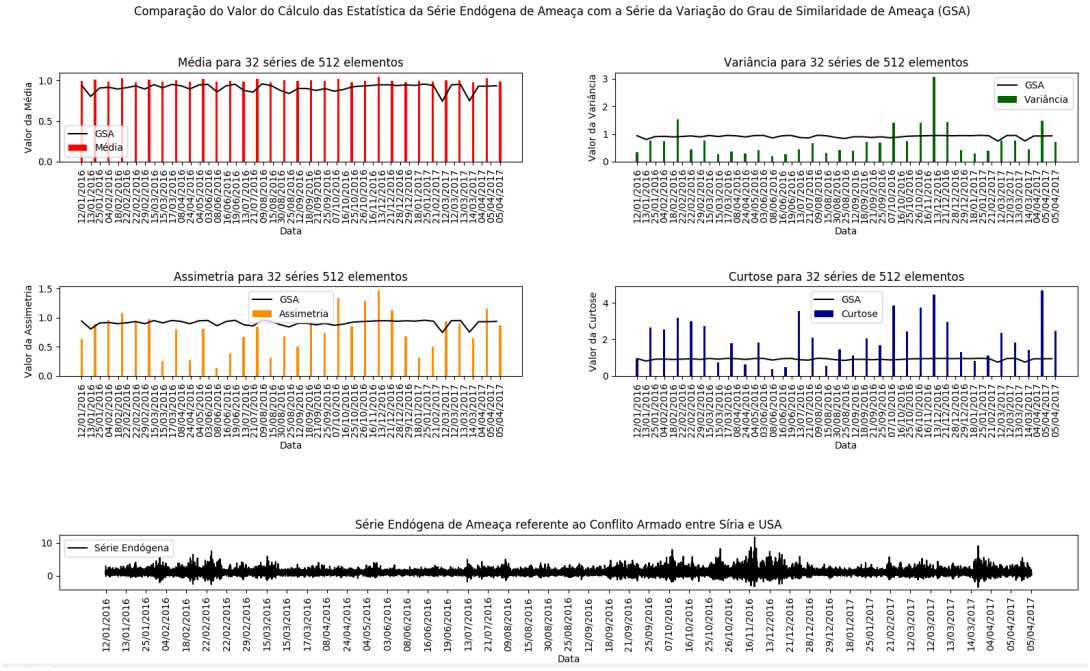


Figura C.2 - Análise dos parâmetros de média, variância, curtose e assimetria considerando 32 séries com 512 elementos para o conflito armado entre Irã e USA.

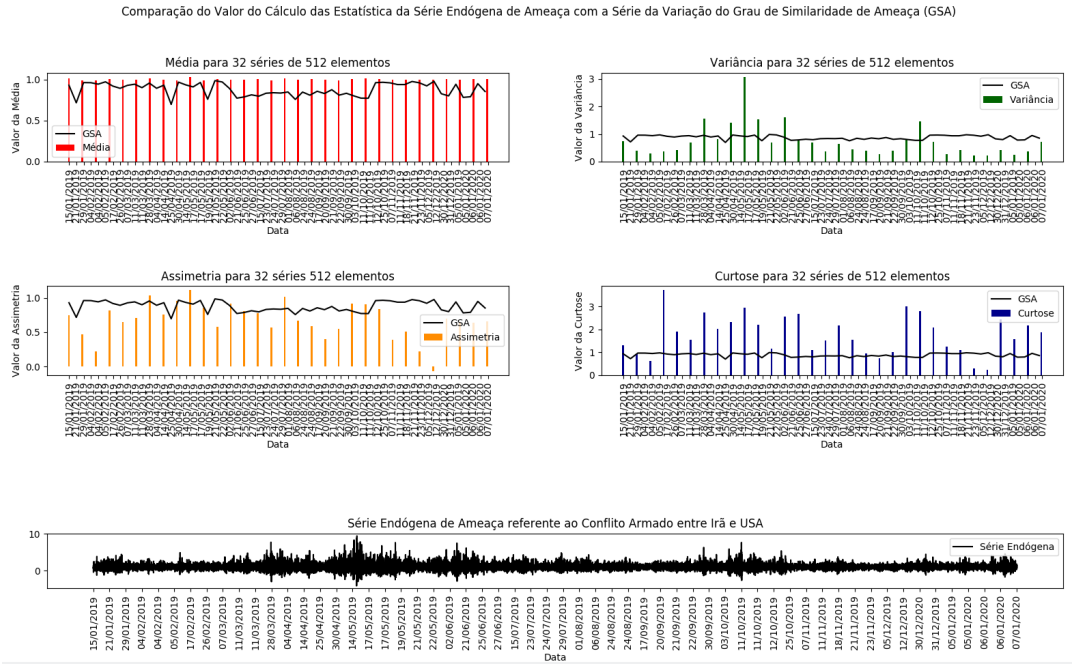


Figura C.3 - Análise dos parâmetros de média, variância, curtose e assimetria considerando 32 séries com 512 elementos para o conflito armado Global.

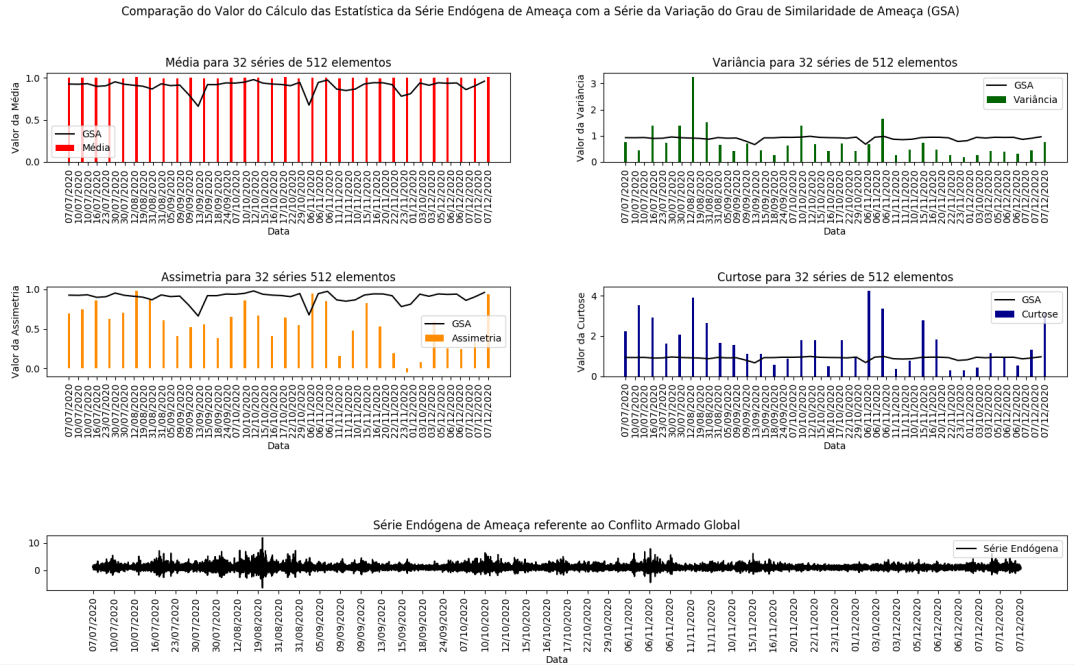


Figura C.4 - Análise dos parâmetros de média, variância, curtose e assimetria considerando 8 séries com 2048 elementos para o conflito armado entre a Síria e USA.

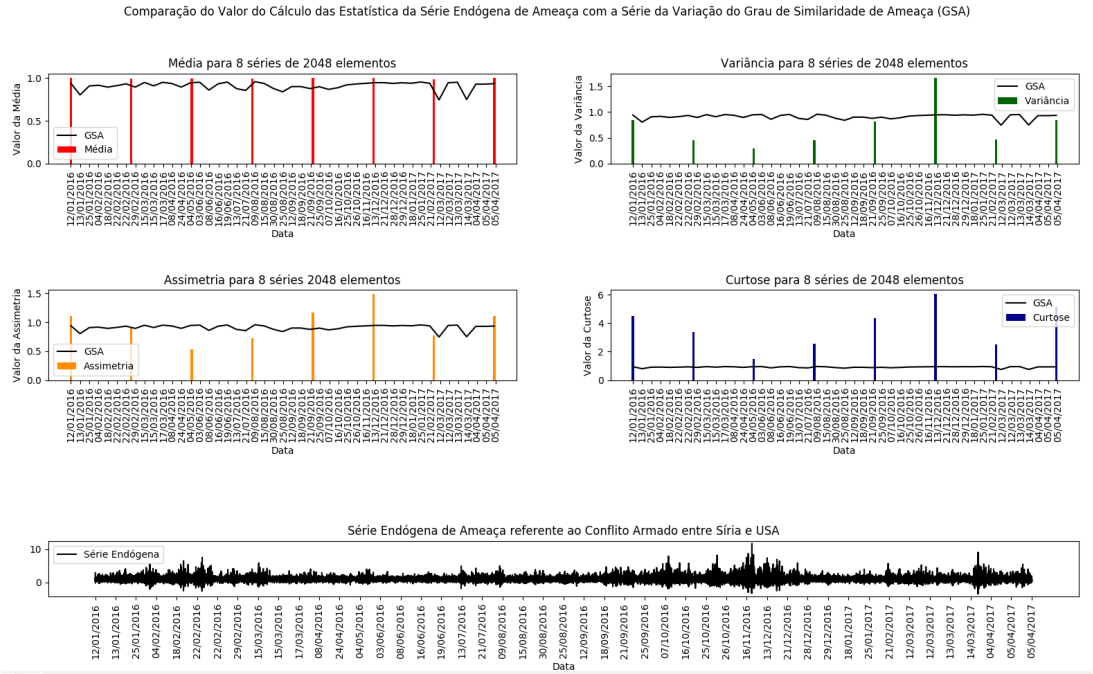


Figura C.5 - Análise dos parâmetros de média, variância, curtose e assimetria considerando 8 séries com 2048 elementos para o conflito armado entre Irã e USA.

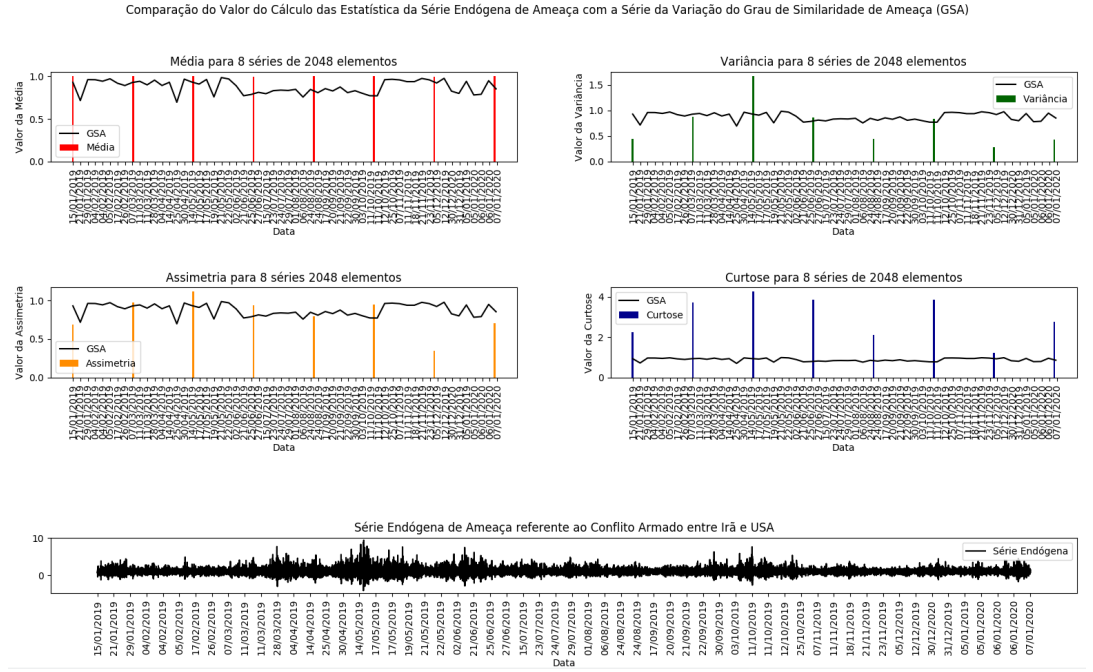
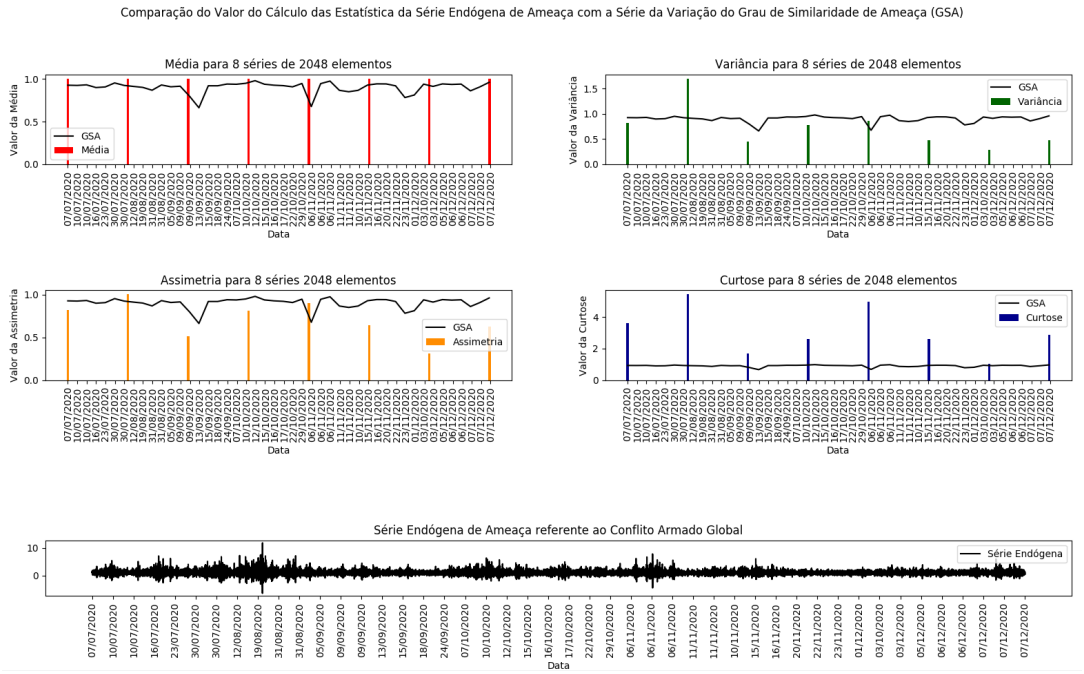


Figura C.6 - Análise dos parâmetros de média, variância, curtose e assimetria considerando 8 séries com 2048 elementos para o conflito armado Global.



C.2 Eventos extremos naturais - ocorrência de seca, queimadas e desmatamentos para o ano de 2020

Nesta seção são apresentados as análises realizadas para os parâmetros de média, variância, curtose e assimetria para a ocorrência de seca, queimadas e desmatamentos para o ano de 2020, considerando 32 séries de 512 elementos e 8 séries com 2048 elementos. Essas análises são apresentadas nas Figuras C.7 e C.8, respectivamente.

Figura C.7 - Análise dos parâmetros de média, variância, curtose e assimetria considerando 32 séries com 512 elementos para a ocorrência de seca, queimadas e desmatamentos para o ano de 2020.

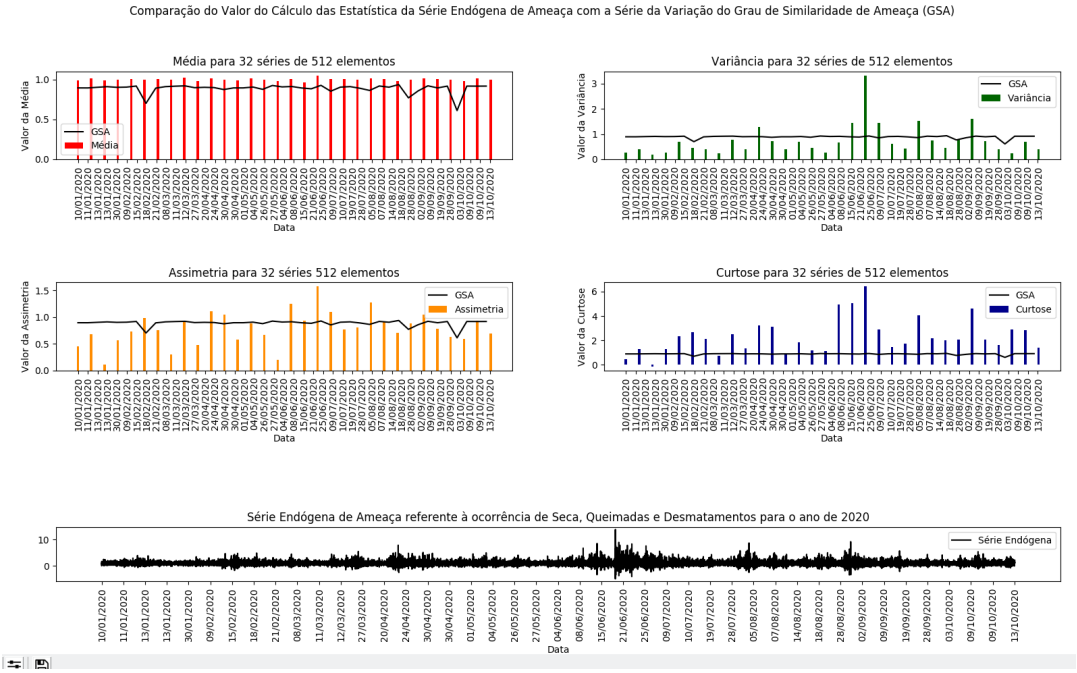
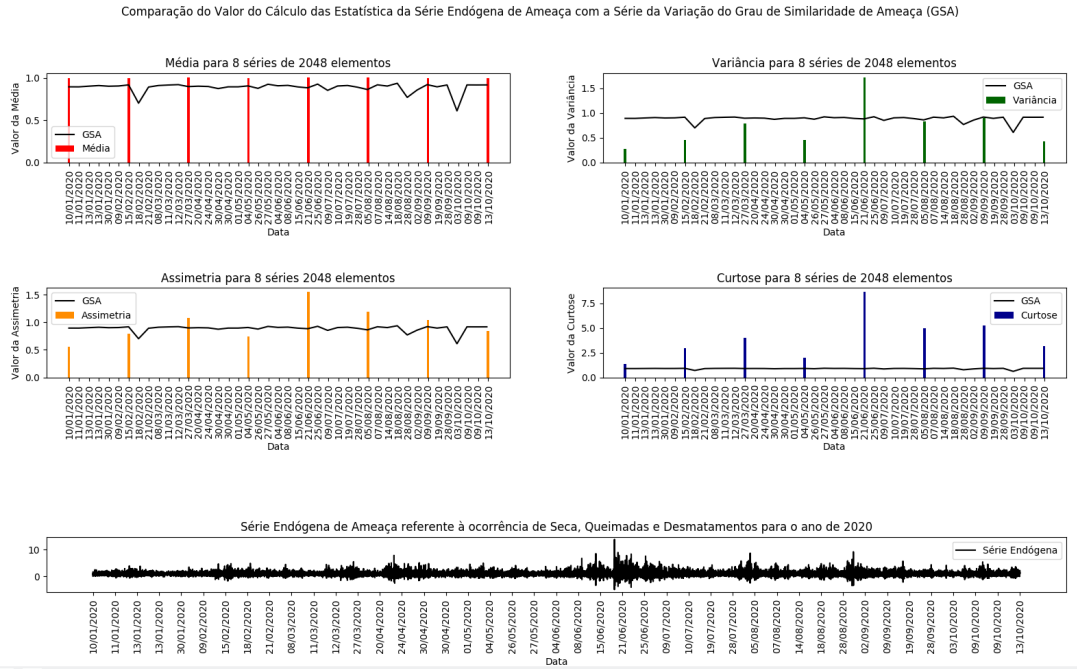


Figura C.8 - Análise dos parâmetros de média, variância, curtose e assimetria considerando 8 séries com 2048 elementos para a ocorrência de seca, queimadas e desmatamentos para o ano de 2020.



ANEXO C - PREDIÇÃO DOS EVENTOS EXTREMOS SOCIAIS PARA A QUANTIDADE DE ELEMENTOS DAS SÉRIES ENDÓGENAS DE 1024, 2048, 4096 E 8192

C.1 Predição do conflito armado entre Síria e USA

Nesta seção são apresentados os testes realizados para a predição da série temporal endógena de 1024, 2048, 4096 e 8192 elementos, para o conflito armado entre Síria e USA. Esses resultados são apresentados nas Figuras C.1, C.2, C.3 e C.4.

Figura C.1 - Predição de parte da série temporal endógena de ameaças do conflito entre a Síria e USA utilizando a série endógena de 1024 elementos.

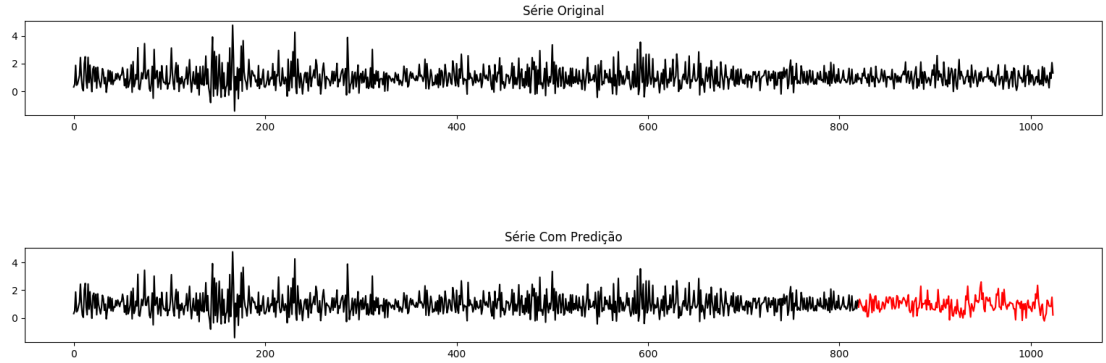


Figura C.2 - Predição de parte da série temporal endógena de ameaças do conflito entre a Síria e USA utilizando a série endógena de 2048 elementos.

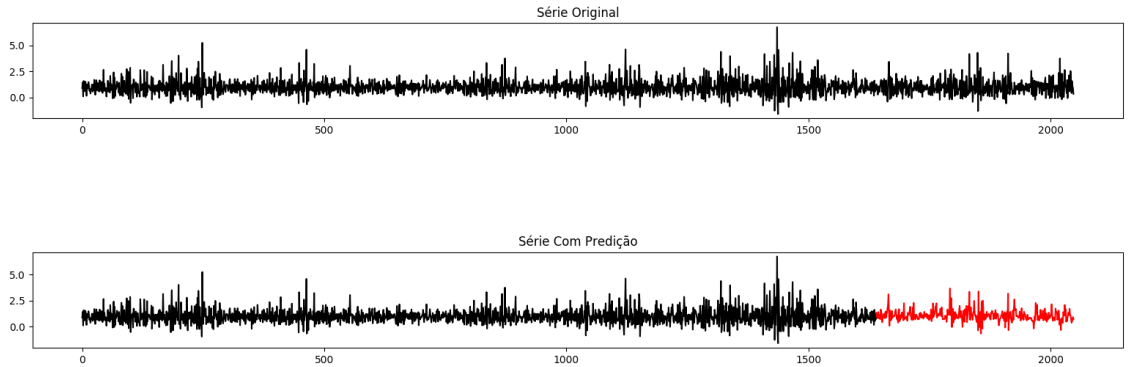


Figura C.3 - Predição de parte da série temporal endógena de ameaças do conflito entre a Síria e USA utilizando a série endógena de 4096 elementos.

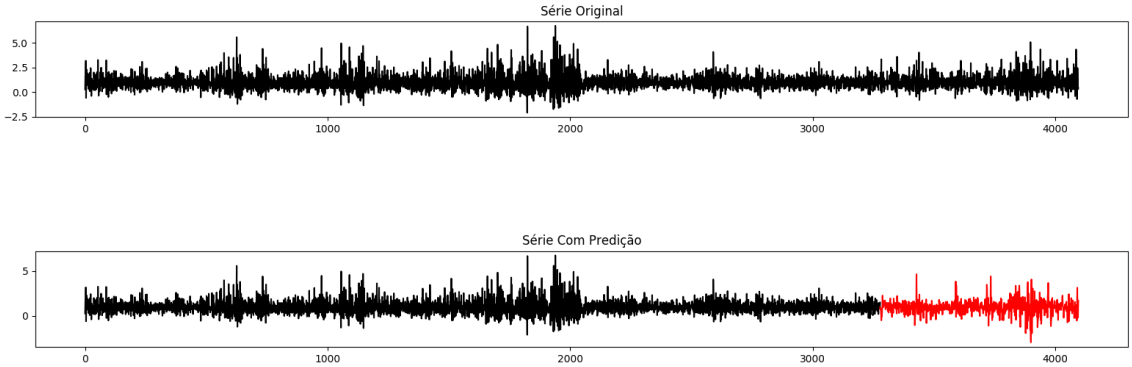
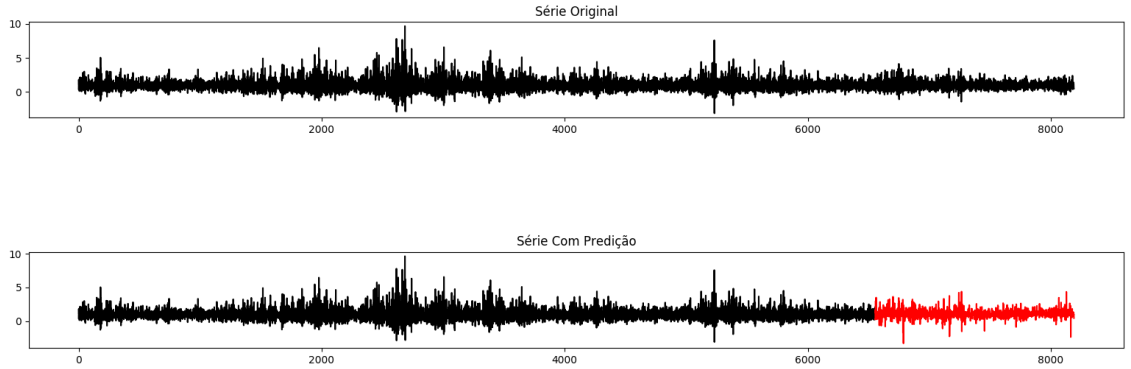


Figura C.4 - Predição de parte da série temporal endógena de ameaças do conflito entre a Síria e USA utilizando a série endógena de 8192 elementos.



PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

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

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.

Propostas e Relatórios de Projetos (PRP)

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

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.

Pré-publicações (PRE)

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

Manuais Técnicos (MAN)

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

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.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

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.