

sid.inpe.br/mtc-mxx/aaaa/00.00.00.00-XXX

SARA – SISTEMA PARA ANÁLISE DE REGIÕES ATIVAS A PARTIR DOS ATRIBUTOS ENVOLVIDOS EM EXPLOSÕES SOLARES

Jéssica de Farias Pereira¹, Ana Estela Antunes da Silva¹, André Leon Sampaio
Gradvohl¹, Guilherme Palermo Coelho¹, José Roberto Cecatto²

¹ Faculdade de Tecnologia (FT) – Universidade Estadual de Campinas

² Divisão de Astrofísica (DIDAS) – Instituto Nacional de Pesquisas Espaciais

Relatório Técnico

URL do documento original:
<<http://urlib.net/xx/yy>>

INPE
São José dos Campos
Fev. / 2019

Esta ficha será revisada pelo SID.

Cutter Sobrenome, Prenome(s) Completo(s) do(s) Autor(es).
Título da publicação / Nome(s) Completo(s) do(s) Autor(es). - São José dos Campos: INPE, ano da publicação.

Grau (Mestrado ou Doutorado em Nome do Curso) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, ano de defesa.

Orientador: Nome completo do orientador(es).

1. Assunto. 2. Assunto. 3. Assunto. 4. Assunto. 5. Assunto.
I. Título

CDU

Dados Internacionais de Catalogação na Publicação



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

AGRADECIMENTOS

Autores agradecem o apoio financeiro da FAPESP por meio do projeto “Exploração de regras de associação para atributos de explosões solares via agrupamento hierárquico aglomerativo”, processo 17/03362-1. Agradecem também às respectivas instituições pelas condições oferecidas para a realização deste trabalho.

RESUMO

A estimativa da probabilidade de ocorrência de explosão solar, dentro de uma janela temporal futura (24 h, 48 h ou 72 h), é muito importante. Isto porque tem a capacidade de servir de base para: (a) a publicação de alerta sobre a possível ocorrência futura do fenômeno, (b) permitir uma análise prévia de impacto do fenômeno nos serviços / sistemas tecnológicos e (c) subsidiar a tomada de decisões referentes a ações para evitar ou mitigar as influências, perturbações, danos e riscos que o fenômeno poderá produzir aos serviços / sistemas tecnológicos e ao ser humano. Neste sentido, a probabilidade estimada opera como uma expectativa – ou meio preditivo - da ocorrência futura do fenômeno. Entretanto, para que possa ser estimada essa probabilidade são necessários: (1) uma análise do histórico e das condições correspondentes que provocaram os fenômenos observados anteriormente e (2) uma ferramenta que permita, a partir de determinados parâmetros ou atributos observados num determinado momento, estimar a probabilidade de ocorrência do fenômeno dentro de uma janela temporal futura de determinada duração. Além disso, a criação dessa ferramenta não é algo simples e direto, pois exige certas etapas intermediárias. Como dois exemplos podemos citar: o acesso a bases históricas tanto de dados quanto de fenômenos energéticos solares e um sistema computacional que forneça o suporte necessário. O sistema computacional designado por “Sistema para Análise de Regiões Ativas” (SARA) foi desenvolvido para permitir a consulta e análise dos dados relacionados a regiões ativas do Sol e explosões solares, tornando as informações mais acessíveis e auxiliando o estudo de seus atributos, e bem como possibilitando a exploração de regras de associação desses atributos para diversos períodos. Com o uso do sistema é possível ter acesso direto aos dados dos arquivos SRS e SGAS do Space Weather Prediction Center - SWPC (<https://www.swpc.noaa.gov>) que são combinados pelo sistema em um arquivo apenas pelo sistema. Isso não era possível anteriormente, pois, os dados estavam distribuídos em dois arquivos e sua consulta se limitava ao site do SWPC. Por meio do uso do SARA, também é possível exportar os dados utilizados e submetê-los a outras técnicas em outras ferramentas. Acredita-se que o sistema se mostra relevante para o estudo e análise de regiões ativas e das prováveis condições existentes nas mesmas que estejam associadas à ocorrência futura de explosões solares. Nesta Nota Técnica, como resultado da colaboração entre a DIDAS-INPE e a FT-Unicamp, apresentamos o sistema que foi criado, suas características, sua operação e os resultados preliminares obtidos.

Palavras-chave: Explosões solares. Regiões ativas. Mineração de dados. Regras de associação.

SYSTEM FOR THE ANALYSIS OF ACTIVE REGIONS SINCE RELATED ATTRIBUTES OF OBSERVED SOLAR BURST / FLARES (SARA)

ABSTRACT

It is very important to estimate the solar flare occurrence-probability within a future time window – 24 h, 48 h or 72 h. This is due to the fact that it can support; (a) alerts release on the possibility of a future phenomenon occurrence, (b) it allows the previous analysis about the impacts of the phenomenon on technological systems/services, and (c) to be subsidiary on taking decisions related to the actions to avoid or mitigate those influences, disturbances, damages and risks that the phenomenon can yield to technological systems/services as well as to the human. Thus, the estimated probability functions as an expectation – or a forecast method – to the phenomenon occurrence. However, as a purpose to estimate this probability the following are necessary: (1) an analysis about the historical database and corresponding conditions that caused the previously observed phenomena and (2) since these parameters which can be either determined or observed attributes for a specific time, a tool which permits to estimate the phenomenon occurrence probability within a chosen future time window. The production of the mentioned tool is difficult, because it requires some steps to be accomplished. As examples, we can cite: the access to historical databases of solar parameters as well as of energetic phenomena and a computing system to a proper support. So, it has been developed the computing-system designed as "Sistema para Análise de Regiões Ativas" (SARA) to allow the consulting and analysis of data related to solar active regions and solar bursts. So, the information became more accessible and auxiliary on the attribute studies. Also, it permitted the exploration of association rules using these attributes for several periods. SARA allowed a direct access to the SRS and SGAS data files from the Space Weather Prediction Center – SWPC (<https://www.swpc.noaa.gov>), These data files are combined for simplicity. By this way, the consulting to the data of interest became quicker. Using the SARA system, it became also possible to work with the data, and applying other techniques available in a number of tools. It is believed the system presents itself as relevant to the studies and analysis of active regions as well as their more probable internal conditions associated to the future occurrence of solar bursts. In this Technical Report, as a result of the collaboration between the DIDAS-INPE and the FT-Unicamp, we present the developed system, its characteristics, functional possibilities and the preliminary results obtained up to now.

Keywords: Solar bursts. Active regions. Data mining. Association rules.

LISTA DE FIGURAS

	<u>Pág.</u>
Figura 1 - Arquivo de dados SRS do dia 12/02/2018.....	12
Figura 2 – Arquivo de dados SGAS do dia 12/02/2018.....	12
Figura 3 – Representação alto nível da arquitetura do sistema.....	15
Figura 4- Tela inicial/Menu Base de dados/Base de dados atualizada.....	17
Figura 5 - Base de dados desatualizada.....	18
Figura 6 - Consulta à base de dados no formato original.....	19
Figura 7 – Consulta à base de dados original, alguns atributos e período.....	19
Figura 8 - Base de dados classificada.....	20
Figura 9 - Consulta à base de dados classificada, sem número de manchas. .	21
Figura 10 - Menu Regras de Associação.....	21
Figura 11 - Regras de associação geradas.....	22
Figura 12 - Menu Sobre.....	23

LISTA DE TABELAS

	<u>Pág.</u>
Tabela 1 - Classificação das Explosões solares a partir do fluxo de raios-X.....	6
Tabela 2 - Base de dados proposta.....	13
Tabela 3 - Classificação do fluxo de rádio 10.7 cm discretizada.....	14
Tabela 4 - Classificação da área do grupo de manchas discretizada.....	14

SUMÁRIO

	<u>Pág.</u>
1 Motivação ao desenvolvimento do sistema SARA.....	1
1.1 Objetivos	3
2 Conceitos envolvidos no desenvolvimento do sistema SARA	4
2.1 Atributos envolvidos em explosões solares e regiões ativas	4
2.2 Regras de Associação	7
3 Métodos e dados utilizados	9
3.1 Elicitação dos requisitos	10
3.2 Base de dados proposta	10
3.3 Discretização dos atributos	13
3.4 Ambiente de desenvolvimento	15
4 Descrição das telas do sistema	17
Referências Bibliográficas	24

1 INTRODUÇÃO – MOTIVAÇÃO AO DESENVOLVIMENTO DO SISTEMA

SARA

O Sol é uma estrela que apresenta comportamento dinâmico ao longo do tempo. O principal fato associado a esse comportamento é o ciclo de atividade solar. Durante um ciclo solar, a atividade solar oscila entre um período de baixa atividade, definido como mínimo solar, ascendendo para um período de alta atividade, chamado de máximo solar. Após o máximo solar, a atividade decai até atingir novamente o mínimo caracterizando o início de um novo ciclo (YOUNG, 2010).

A partir do mínimo é observado um aumento gradual do número de manchas solares, acompanhado dos aumentos de irradiância total, fluxo rádio medido em 10.7cm e frequência de ocorrência de explosões solares. acompanhadas do aumento do fluxo de raio-X na forma de explosões solares,. Tais explosões, quando ocorrem na área visível do Sol, observado a partir da Terra, e com uma intensidade de moderada a forte ou extrema, podem afetar temporariamente a atmosfera superior do nosso planeta, perturbando ou interrompendo sinais de sistemas tecnológicos, e influenciando seu desempenho e confiabilidade (SCHWENN, 2006).

Além disso, geralmente essas explosões emitem radiação eletromagnética num amplo espectro, desde ondas de rádio até raios-X e raios-gama. Por este motivo, são observadas em solo por instrumentos nas bandas visível e rádio e, em outras bandas por instrumentos à bordo de satélites. A série de satélites GOES, que monitora o Sol continuamente, possui um medidor do fluxo de raios-X solares que registra o nível de emissão desde o Sol calmo até as explosões solares. A longa (décadas) série temporal dessas medidas permite vários tipos de abordagem para análise dos dados.

Mas, o aspecto mais importante é que a radiação emitida por uma explosão solar leva pouco mais de 8 minutos para chegar à Terra. Outra coisa importante

é que, mesmo para um instrumento a bordo de um satélite geo-estacionário, o registro da explosão solar ocorre com um décimo de segundo de antecedência de sua chegada em solo terrestre, o que na prática é imediato. Portanto, a partir do registro desse fenômeno, não existe qualquer tempo hábil para: tomar conhecimento do fenômeno, emitir alertas, analisar a situação, e definir e executar ações de prevenção ou mitigação dos efeitos provocados no ambiente terrestre e em sistemas tecnológicos. Sendo assim, uma vez que ainda não existe um modelo definitivo que represente adequadamente o fenômeno, a única maneira de lidar com o problema é realizar a previsão de sua ocorrência com a devida antecedência – pelo menos um dia, ou mais.

Na literatura, o estudo do clima espacial ainda é muito recente e desafiador. O grande desafio associado às explosões solares é a possibilidade de previsão antecipada a fim de minimizar os danos causados. Entretanto, para que seja possível prever adequadamente as explosões solares, ainda é preciso entender quais atributos estão associados a elas e como eles influenciam a ocorrência de uma explosão. Alguns desses atributos são: o número e área de manchas solares, fluxo de rádio e fluxo de raio-X integrados sobre o disco solar visível (DIERCKXSENS et al., 2015).

Uma forma de auxiliar o estudo e entendimento desses fenômenos é por meio da análise dos dados envolvidos em explosões solares a partir de suas respectivas regiões ativas de origem, e da aplicação de algoritmos de mineração de dados. Regiões ativas são regiões da atmosfera solar que possuem um forte (de centenas a poucos milhares de Gauss de intensidade) e dinâmico campo magnético onde ocorrem os eventos (explosões) associados à atividade solar, com uma frequência muito maior durante o período de máximo do ciclo de atividade (BOULDER, 2014).

Uma das dificuldades encontradas no estudo das explosões solares é conseguir analisar os fenômenos a partir das regiões ativas que os produziram. Tais regiões ativas possuem características diferentes umas das outras e a

dificuldade está justamente em entender quais características produziram determinados eventos. Além disso, com o conhecimento que se tem até o presente, uma explosão solar pode ocorrer a qualquer momento, embora existam vários parâmetros solares que podem fornecer indícios de uma maior probabilidade de ocorrência do fenômeno em determinados períodos de tempo, como já mencionado acima. O Space Weather Prediction Center (SWPC), é uma instituição que disponibiliza dados referentes à atividade solar a partir das medições realizadas pelo satélite GOES-15, por exemplo, e que provê informações referentes às características das regiões ativas e dos eventos produzidos pelo Sol. Porém, não há um relatório que mostre todas essas informações em conjunto, o que dificulta a análise diária (SPACE, 2018). A união dessas informações num relatório ou arquivo único facilita muito o entendimento e trabalho com esses dados.

Em adição a essa questão de disponibilização dos dados, uma das técnicas de mineração de dados capaz de auxiliar no entendimento dos mesmos é a técnica de associação, que tem como premissa encontrar elementos que implicam na presença de outros elementos em uma mesma transação. Além da união das informações referentes às regiões ativas e explosões solares, essa técnica é capaz de descobrir a existência de relações entre esses dados (VASCONCELOS; CARVALHO, 2004).

Com base no contexto do clima espacial apresentado, o SARA auxilia a análise de regiões ativas ao permitir a consulta às características de eventos associados a uma região específica, e a consulta às regras de associação criadas a partir desses dados. Com essas informações, espera-se que o sistema possa auxiliar no entendimento das regiões ativas e como as mesmas estão relacionadas às explosões solares.

1.1 Objetivos do Sistema para Análise de Regiões Ativas (SARA)

O objetivo desse sistema é consultar e analisar dados relacionados a regiões ativas e explosões solares.

Os objetivos específicos são sintetizados por:

- Consultar os dados de regiões ativas a partir da captação dos mesmos disponibilizados no site Space Weather Prediction Center (SWPC) (SPACE, 2018) para construção de uma base de dados.
- Reunir os dados referentes às características das regiões ativas, presentes no arquivo Solar Region Summary (SRS), e os eventos produzidos pelo Sol, do arquivo Solar and Geophysical Activity Summary (SGAS), ambos disponibilizados pelo SWPC.
- Possibilitar a consulta às regiões ativas associadas aos eventos que produziram.
- Possibilitar a criação e análise de regras de associação utilizando a base de dados construída que está fundamentada em parâmetros (atributos) de regiões ativas, por meio dos índices de suporte e confiança.

2 CONCEITOS ENVOLVIDOS NO DESENVOLVIMENTO DO SISTEMA SARA

Nas sub-seções a seguir são apresentados os principais conceitos envolvidos no desenvolvimento do sistema SARA, relacionados ao clima espacial e a regras de associação.

2.1 Atributos envolvidos em explosões solares e regiões ativas

A atividade que ocorre no Sol costuma ser dividida em períodos de tempo chamados ciclos solares, onde cada ciclo tem em média onze anos de duração (LOSKUTOV et al., 2001). Durante um ciclo solar, muitas (milhares) explosões solares – a explosão solar é um repentino aumento da emissão de radiação eletromagnética solar originada de uma região ativa da atmosfera solar – são registradas. Essas explosões podem atingir a Terra cerca de oito minutos depois de seu início.

Por outro lado, manchas solares são áreas mais escuras e frias observadas na “superfície” do Sol, a fotosfera. A fotosfera tem uma temperatura de cerca 5,800 K enquanto as manchas têm temperaturas na faixa de 3000 a 4000 K (Sunspots, 2018). Isto ocorre porque as manchas se constituem de intensos campos magnéticos – centenas a poucos milhares de Gauss - os quais inibem o processo convectivo de transporte de energia das camadas sub-superficiais até a fotosfera. Essas áreas parecem escuras, em comparação com as regiões mais quentes da fotosfera brilhante que predomina (BOULDER, 2011), por efeito de contraste pois, como o gás está mais frio conseqüentemente se torna bem menos brilhante.

Existem três sistemas de classificação magnética para as manchas solares: Mt. Wilson, Zurich e McIntosh. A classificação de Mt. Wilson baseia-se na distribuição de polaridades magnéticas dentro de cada grupo de manchas solares, e possui oito classes: alfa, beta, beta-gamma, beta-delta, gamma, delta, beta-gamma-delta e gamma-delta (COLAK; QAHWAJI, 2007). A classificação de Zurich considera o sinal do campo magnético das manchas no grupo, quantas manchas individuais constituem o grupo, as características da penumbra e o tamanho da região de manchas (NGUYEN et al., 2016; ALEXANDER, 2009). Já a classificação de McIntosh é uma versão modificada da classificação de Zurich. A forma geral da classificação de McIntosh é Zpc , onde Z é a classe de Zurich modificada, classificando a mancha de acordo com a polaridade, existência de penumbra e tamanho; p descreve a penumbra da maior mancha do grupo; e c descreve a distribuição de manchas no interior do grupo - ou o quão compactas as manchas estão dispostas no interior do grupo (COLAK; QAHWAJI, 2007).

O fluxo de raio-X integrado ao longo de todo o disco solar é um dos parâmetros usados para classificar as explosões solares, já que, durante uma explosão, a energia magnética é convertida em radiação eletromagnética num amplo

espectro, incluindo os raios-X. Elas então são classificadas de acordo com o seu fluxo, com base na magnitude do pico da explosão.

As explosões podem ser categorizadas em cinco classes (A, B, C, M e X), de acordo com o nível de fluxo máximo de raio-X registrado em Watt por metro quadrado, em uma escala logarítmica, sendo A a menor e X a maior década de intensidade das explosões. A Tabela 1 apresenta os níveis de fluxo correspondentes a cada uma das classes de explosões solares registradas na banda de raios-X pelo sensor do satélite GOES que registra continuamente essa emissão solar.

Tabela 1 - Classificação das Explosões solares a partir do fluxo de raios-X.

Fluxo de Raios-X – 1-8Å (W/m ²)	Classe
$10^{-8} - 10^{-7}$	A
$10^{-7} - 10^{-6}$	B
$10^{-6} - 10^{-5}$	C
$10^{-5} - 10^{-4}$	M
$> 10^{-4}$	X

Fonte: Produção do autor.

Dependendo de características como tamanho, forma e densidade das manchas solares, bem como principalmente da dinâmica de seus respectivos campos magnéticos, elas podem produzir explosões solares de classe M ou X.

Para o sistema SARA foi utilizada a base de dados do SWPC para capturar os dados relacionados às regiões ativas e aos eventos associados. A base de dados do SWPC provê, entre outros atributos, a classificação magnética Mt. Wilson do grupo de manchas, a área e o número de manchas e a localização da região ativa. Em relação ao fluxo de raios-X, a base disponibiliza o valor do

fluxo relacionado às classes das explosões, além do valor do fluxo de rádio (SPACE, 2015).

2.2 Regras de Associação

Mineração de dados é uma área da computação que tem por objetivo encontrar padrões, ou seja, ocorrências de dados com características ou comportamento similares durante um evento específico, em um conjunto de dados. É uma das principais técnicas da mineração de dados que podem ser executadas em conjuntos de dados é a técnica de associação (HAN; KAMBER; PEI, 2012).

A tarefa de associação é uma ferramenta importante de mineração ao analisar grandes volumes de dados. É usada para descobrir relacionamentos interessantes entre os atributos ao procurar por relações recorrentes em uma base. Chama-se regra o resultado da análise dessas relações recorrentes entre os atributos (LEE et al., 2016).

Esses tipos de relações entre elementos podem ser encontrados através das regras de associação, que são apresentadas, nos conjuntos de dados, na forma: SE atributo X ENTÃO atributo Y, onde um conjunto de atributos X implica outro conjunto de atributos Y, podendo ser representadas pela expressão $\{X\} \rightarrow \{Y\}$.

As regras criadas a partir do algoritmo de associação, APRIORI (HAN; KAMBER; PEI, 2012), mostram as associações possíveis entre os atributos, bem como a frequência com que essas associações aparecem na base de dados e quão promissoras elas são – a partir de índices que ajudam a validar tais regras (HAN; KAMBER; PEI, 2012). Dessa forma, o algoritmo de associação pode ajudar a compreender melhor como diferentes elementos interagem uns com os outros (TAN; STEINBACH; KUMAR, 2005).

Para que uma regra de associação seja criada, os atributos presentes em uma transação devem ser analisados. Uma transação corresponde aos itens presentes em uma determinada operação, e trazendo este conceito à análise

proposta baseada em regiões ativas, uma transação diz respeito às características de uma região relacionadas a produção de eventos da atividade solar – sendo o foco da análise o evento das explosões solares (VASCONCELOS & CARVALHO, 2004).

Dessa forma, o algoritmo de associação procura nas transações da base de dados quais atributos estão presentes, sendo um atributo de uma determinada transação, assim como um conjunto de atributos, chamados de itemsets. Os itemsets definidos de uma transação são procurados em toda a base de dados medindo a frequência com que eles aparecem. A partir da frequência calculada para esses itemsets, serão considerados regras de associação dos itemsets que sustentarem uma frequência mínima estabelecida, chamada de suporte da regra. Se tal itemset for frequente, todos os seus subconjuntos de itens também serão considerados frequentes, e se a condição de suporte ainda for atendida, os itemsets derivados também resultarão em uma regra de associação (HAN; KAMBER; PEI, 2012).

Além do índice de suporte da regra, o algoritmo analisa outro índice para definir quais regras serão criadas a partir das transações da base de dados, a confiança da regra. Após a definição dos itemsets frequentes da base de dados, são validados quais itemsets – atributos ou conjunto de atributos – aparecem com maior frequência em conjunto com outros itemsets em uma mesma transação. A confiança da regra mede a probabilidade de que um itemset Y esteja presente em uma transação dado a presença do itemset X, definindo assim a regra de associação $\{X\} \rightarrow \{Y\}$ tendo Y como conseqüente de X.

Sendo assim, a força de uma regra de associação, ou o quão interessante ela pode ser, é medida através dos níveis de suporte e confiança, sendo que o suporte mede o quão frequente os itens de uma transação aparecem na base de dados, sendo descrito pela fórmula $suporte(X) = |\{t \in T; X \subseteq t\}| / |T|$ onde T é o conjunto de transações, t é uma transação da base de dados e X é um itemset

contido em t (TAN; STEINBACH; KUMAR, 2005). Já a confiança, fornece a probabilidade estimada de Y dado X , mostrando o quão verdadeiro é a relação entre o conseqüente e antecedente: $confiança(X \rightarrow Y) = suporte(X \cup Y) / suporte(X)$. Quanto maior a confiança, maior a probabilidade de que Y esteja presente nas transações que contém X (TAN; STEINBACH; KUMAR, 2005).

Para que seja possível utilizar o algoritmo APRIORI e descobrir regras de associação, é preciso que os dados utilizados sejam discretos, ou seja, apresentem classes que os identifiquem. Isso se dá ao fato do algoritmo calcular a frequência dos itemsets através da presença ou não dos determinados atributos, algo que só se torna viável quando os mesmos correspondem a classes específicas.

Normalmente, as regras de associação são consideradas interessantes se satisfizerem tanto um limite de suporte mínimo quanto um limite de confiança mínima (HAN; KAMBER; PEI, 2012). Esses valores mínimos são dados que são submetidos ao algoritmo com limite para criação das regras, além de parâmetros de análise.

Entretanto, as regras de associação devem ser analisadas com cuidado, Tan, Steinbach e Kumar (2005) explicam que uma regra não necessariamente implica em uma relação de causa e efeito, e que sim, indicam uma forte relação entre os itens presentes no antecedente e conseqüente da regra.

3 MÉTODOS E DADOS UTILIZADOS

Nesta seção são descritos: os requisitos funcionais do sistema, os dados utilizados pelo mesmo e as ferramentas e bibliotecas utilizadas no seu desenvolvimento.

3.1 Requisitos Funcionais

Os requisitos funcionais levantados foram os apresentados a seguir.

- Possibilidade de consultar diariamente os dados relacionados às regiões ativas e os eventos produzidos por elas, definindo: tipo da base de dados a ser consultada, em seu formato original ou categorizada; tipos de atributos escolhidos para consulta; e o período, podendo ser selecionado de 1997 até o dia atual da consulta.
- Possibilidade de criar regras de associação, considerando as mesmas funcionalidades presentes na consulta à base de dados, podendo definir também os índices de suporte e confiança mínimos a serem utilizados.
- Exportar a base de dados consultada e as regras de associação geradas para o formato CSV, permitindo assim utilizar esses dados em outras ferramentas.

3.2 Base de dados

Os dados utilizados pelo sistema SARA estão, de alguma forma, relacionados às explosões solares a partir das regiões ativas. Tais dados são disponibilizados pelo SWPC em diferentes formatos e fontes. Dessa forma, foi efetuada uma análise sobre os atributos envolvidos em explosões solares e a que tipo de relatório eles pertencem. Com isso, foram identificados dois conjuntos de dados diferentes, um com as informações referentes às características das regiões ativas solares, e outro com os eventos produzidos por elas. Esses conjuntos de dados são disponibilizados no SWPC através de arquivos texto com as informações associadas às regiões ativas e eventos produzidos. Tais arquivos são disponibilizados diariamente desde 1997.

Para os dados relacionados às características das regiões ativas solares, foi utilizado o arquivo de dados Solar Region Summary (SRS). Nele, diariamente são divulgadas informações sobre: identificador da região ativa, localização, área das manchas (medida em Milionésimos da área do Disco Solar – MDS),

número de manchas observadas, classificação magnética de grupos de manchas de Zurich e classificação magnética de Mt. Wilson (SWPC, 2018).

As informações sobre a atividade solar produzida pelas regiões ativas foram extraídas dos arquivos de dados Solar and Geophysical Activity Summary (SGAS) e “events”. Neles, constam as informações sobre os eventos produzidos, que podem ser: explosões solares medidas a partir do fluxo de raios-X, explosões ópticas, fluxo de rádio medido em 10.7 cm, registrando quando possível, o início, máximo e fim do evento (SWPC, 2018).

No SGAS, apenas alguns dos eventos são registrados, e são os que atendem aos seguintes requisitos: explosões de raios-X com intensidade de classe M ou maior; explosão óptica de importância >2b; fluxo de rádio > 100 solar flux unit (sfu) medido na frequência 245 MHz; fluxo de rádio 100% acima da medida de fundo na frequência 10.7 cm (SWPC, 2018). Dentre os atributos disponibilizados pelos arquivos de dados do SWPC, considerando SRS e SGAS, foram selecionados: área, número de manchas, classificação magnética da mancha de Mt. Wilson, fluxo de raios-X e fluxo de rádio.

As Figuras 1 e 2 mostram os arquivos de dados originais extraídos do SWPC. O arquivo SRS, representado na Figura 1, traz as características da região ativa solar observada no dia 12 de fevereiro de 2018. O arquivo SGAS, representado na Figura 2, traz os eventos produzidos por determinadas regiões, referentes ao mesmo dia.

Essas informações aparecem destacadas nas figuras por um retângulo vermelho. Observando os arquivos, é possível perceber que uma mesma região produziu quatro eventos nesse dia.

Na Figura 1, no arquivo SRS, foi observada uma única região de identificador 2699 com classificação magnética Beta. Na Figura 2, nos eventos produzidos no dia, pode-se ver que esta região produziu quatro eventos, sendo esses eventos explosões solares de classe B.

Figura 1 - Arquivo de dados SRS do dia 12/02/2018.

```

:Product: 0212SRS.txt
:Issued: 2018 Feb 12 0030 UTC
# Prepared jointly by the U.S. Dept. of Commerce, NOAA,
# Space Weather Prediction Center and the U.S. Air Force.
#
Joint USAF/NOAA Solar Region Summary
SRS Number 43 Issued at 0030Z on 12 Feb 2018
Report compiled from data received at SWO on 11 Feb
I. Regions with Sunspots. Locations Valid at 11/2400Z
Nmbr Location Lo Area Z LL NN Mag Type
2699 S07W14 167 0230 Dai 10 14 Beta
IA. H-alpha Plages without Spots. Locations Valid at 11/2400Z Feb
Nmbr Location Lo
2698 S03W58 211
II. Regions Due to Return 12 Feb to 14 Feb
Nmbr Lat Lo
None

```

Fonte: Adaptada do Space Weather Prediction Center.

Figura 2 – Arquivo de dados SGAS do dia 12/02/2018.

```

:Product: 0212SGAS.txt
:Issued: 2018 Feb 12 0245 UTC
# Prepared jointly by the U.S. Dept. of Commerce, NOAA,
# Space Weather Prediction Center and the U.S. Air Force.
#
Joint USAF/NOAA Solar and Geophysical Activity Summary
SGAS Number 043 Issued at 0245Z on 12 Feb 2018
This report is compiled from data received at SWO on 11 Feb
A. Energetic Events
Begin Max End Rgn Loc Xray Op 245MHz 10cm Sweep
0104 0109 0111 2699 084.8 1200
0819 0824 0827 2699 081.7 390
2143 2147 2149 2699 081.6 100
2310 2313 2316 2699 081.0 150
B. Proton Events: None
C. Geomagnetic Activity Summary: Field conditions were at quiet
levels.
D. Stratwarm: Not Available
E. Daily Indices: (real-time preliminary/estimated values)
10 cm 078 SSN 024 Afr/Ap 002/001 X-ray Background A6.3
Daily Proton Fluence (flux accumulation over 24 hrs)
GT 1 MeV 3.9e+05 GT 10 MeV 1.7e+04 p/(cm2-ster-day)
(GOES-15 satellite synchronous orbit W135 degrees)
Daily Electron Fluence
GT 2 MeV 7.30e+04 e/(cm2-ster-day)
(GOES-15 satellite synchronous orbit W135 degrees)
3 Hour K-indices:
Boulder 0 0 0 0 1 2 1 0 Planetary 1 0 0 0 0 1 0 1
F. Comments: None

```

Fonte: Adaptada do Space Weather Prediction Center.

A partir desses arquivos de dados e dos atributos selecionados para análise, foi definida a base de dados proposta, representada pela Tabela 2.

Cada linha da base representa um evento de uma região em um determinado dia, se tornando uma transação da base de dados. Para que fosse possível montar a base dessa maneira, foi necessário pré-processamento dos arquivos para a extração das informações necessárias.

Tabela 2 - Base de dados proposta.

Data			Região (N#)	Área (MDS)	No. de manchs	Class. Magn. (Mt. Wilson)	Fluxo raios-X 1-8 Å (W/m ²)	Fluxo rádio diário * 10.7cm (s.f.u.)
Ano	Mês	Dia						
2014	12	11	2234	30	11	BG	$3,8 \times 10^{-6}$	150
2014	12	12	2234	110	9	BG	$4,0 \times 10^{-6}$	148
2014	12	11	2235	220	3	B	$1,4 \times 10^{-6}$	150
2014	12	12	2236	180	5	B	$2,8 \times 10^{-6}$	148

Fonte: Adaptada do Space Weather Prediction Center – arquivos SRS e SGAS.

* Integradado sobre todo o disco solar visível.

3.3 Discretização dos atributos

Para submissão da base de dados proposta à criação das regras de associação, se faz necessária a discretização dos atributos. Como descrito na seção 2.2, o algoritmo de associação descobre os relacionamentos através da presença ou não dos atributos em determinadas transações. Portanto, os atributos utilizados na base de dados que não são categóricos, ou que ainda não possuem uma classificação conhecida na literatura, precisaram ser discretizados. Deve-se ressaltar que o critério de discretização aplicado foi subjetivo tendo como intuito primeiro uma estratificação aproximadamente uniforme e, em seguida, a distribuição quantitativa dos valores dos atributos baseado nos valores históricos registrados.

O atributo fluxo de raios-X já possui discretização conhecida na literatura que, conforme descrito, divide o fluxo em 5 categorias de acordo com a intensidade, sendo as classes A, B, C, M e X. Para o atributo de classificação de mancha foi utilizada a classificação de Mt. Wilson que pode assumir os valores alfa, beta, gamma, beta-gamma, beta-gamma-delta, gamma-delta, beta-delta e delta.

Para os demais atributos, foi necessária sua discretização. Para discretização do fluxo de rádio, os valores foram discretizados conforme a Tabela 3.

Tabela 3 - Classificação do fluxo de rádio 10.7cm discretizada.

Fluxo de rádio (10.7cm)	
< 80	Baixo
80 < x < 120	Médio
120 < x < 160	Alto
> 160	Altíssimo

Fonte: Produção do autor.

Para a utilização da área do grupo de manchas dada em milionésimos da área do disco solar visível, a discretização dos dados foi realizada conforme a Tabela 4.

Tabela 4 - Classificação da área do grupo de manchas discretizada

Área da mancha (MADS)	
0 a 200	Área pequena
200 a 500	Área moderada
500 a 1000	Área grande
> 1000	Área gigantesca

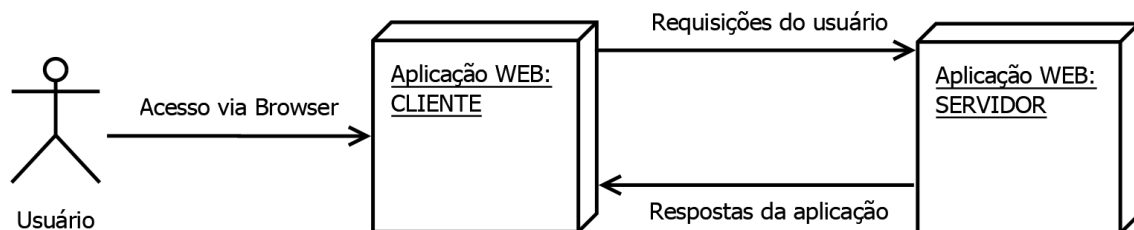
Fonte: Produção do autor.

Com isso, o sistema é capaz de fornecer acesso diário às informações referentes às regiões ativas solares e análise entre os atributos através da técnica de associação.

3.4 Ambiente de desenvolvimento

Para o desenvolvimento da aplicação foram utilizadas diversas tecnologias em diferentes partes do sistema. De modo a garantir a facilidade de expansão e manutenção, foi pensado numa arquitetura que modularizasse a aplicação em uma API (“Application Programming Interface”), hospedada em um servidor web, e uma aplicação cliente que a consumisse, podendo assim desacoplar as aplicações. A Figura 3 mostra uma representação alto nível da aplicação de acordo com a arquitetura escolhida.

Figura 3 – Representação alto nível da arquitetura do sistema.



Fonte: Produção do autor.

A aplicação WEB do lado do servidor, ou comumente conhecida como aplicação “back-end”, foi desenvolvida em Python e é a parte da aplicação que possui as regras da parte que envolve os conceitos da Física e a execução do algoritmo do sistema: o acesso aos dados do SWPC, criação das bases de dados e regras de associação. A linguagem Python foi escolhida por ser uma linguagem amplamente utilizada para aplicações relacionadas a ciência de dados. Ela fornece uma ampla gama de bibliotecas relacionadas a mineração de dados, além de ser uma linguagem de fácil aprendizado (PYTHON, 2018).

Duas bibliotecas Python principais foram utilizadas no “back-end” da aplicação: MLXtend e Flask. A biblioteca MLXtend (Machine Learning Extensions) forneceu suporte para a criação das regras de associação através da utilização do algoritmo APRIORI e de sua análise a partir dos índices de suporte e confiança (MLXTEND, 2018). A biblioteca Flask permitiu o suporte a requisições HTTP, como métodos GET e POST, transformando assim o sistema em uma API (FLASK, 2018).

Não foi utilizado no sistema um banco de dados relacional. A base de dados proposta se dá pela união de dois arquivos texto em um terceiro e, ao ser feita uma requisição, o sistema carrega a base de dados em memória e manipula os dados conforme solicitado. Dessa forma, é possível que a API manipule as informações sem maiores custos respondendo às requisições com mais agilidade.

Para o desenvolvimento da parte cliente da aplicação, também comumente conhecido como “front-end”, foi utilizado o framework JavaScript ReactJS. ReactJS é um “framework” novo, porém muito completo e popular, sendo atualmente o “framework” JavaScript para aplicações cliente mais utilizado (MIGLANI, 2018). Ele utiliza a tecnologia “Single Page Application”, que permite manipular toda a aplicação em uma única página, melhorando assim a experiência do usuário, desempenho e facilidade com a manutenção da aplicação.

As principais bibliotecas utilizadas para o desenvolvimento do “front-end” da aplicação foram create-react-app, que permite uma instalação fácil e rápida do ambiente; e react-router, que permite o mapeamento de rotas e URLs da aplicação (NPM, 2018).

Para controle e versionamento da aplicação, foi utilizado o Git juntamente com o repositório do GitHub (GITHUB, 2018) como plataforma de hospedagem, fornecendo assim uma forma segura e prática para armazenamento do código do sistema.

4 DESCRIÇÃO DAS TELAS DO SISTEMA

O sistema foi construído para plataforma web e possui uma interface gráfica de forma a garantir o acesso dos possíveis usuários interessados. Dessa forma, esta seção mostrará as telas do sistema explicando suas respectivas funcionalidades de acordo com a solução proposta.

O sistema possui três menus principais: Base de Dados, Regras de Associação e Sobre. A tela inicial do sistema é o menu Base de Dados conforme mostrado na Figura 4.

Figura 4- Tela inicial/Menu Base de dados/Base de dados atualizada.

A imagem mostra a interface web do sistema SARA. No topo, há um menu de navegação com as opções 'SARA', 'Base de dados', 'Regras de Associação' e 'Sobre'. Abaixo, há uma seção de informações da base de dados que indica 'Último dia na base de dados 2018-11-04' com um botão 'Atualizada' em verde. A interface é dividida em duas colunas de opções de filtragem. A coluna da esquerda contém 'Tipo de Base de dados' com radio buttons para 'Original' (selecionado) e 'Classificada', e 'Atributos' com checkboxes para 'Área', 'Número de manchas solares', 'Classificação magnética da mancha', 'Explosões por Raio-X' e 'Fluxo de rádio'. A coluna da direita contém 'Período' com campos para 'Começo' (01/01/1997) e 'Fim' (04/11/2018), além de um checkbox 'Use toda a base de dados' selecionado. Um botão azul 'Pesquisar' está localizado na base da seção de filtros. Na parte inferior da interface, há um link azul 'Esconder'.

Fonte: Produção do autor.

Logo num primeiro momento, o usuário tem acesso ao status atual da base de dados mostrando a data da última atualização realizada, ou seja, o último dia em que constam registros da atividade solar na base de dados. Caso essa data seja a data corrente, o usuário terá a informação que a base de dados está atualizada, se não, mostrará o status da base como desatualizada e permitirá a opção de atualizá-la conforme Figura 5.

Figura 5 - Base de dados desatualizada.

SARA Base de dados Regras de Associação Sobre

Informações da base de dados: Último dia na base de dados 2018-11-04 Desatualizada [Atualizar](#)

Tipo de Base de dados

Original Classificada

Atributos

Área Explosões por Ralo-X

Número de manchas solares Fluxo de rádio

Classificação magnética da mancha

Período

Começo: 01/01/1997

Fim: 04/11/2018

Use toda a base de dados

[Pesquisar](#)

[Esconder](#)

Fonte: Produção do autor.

Independentemente da atualização da base, o usuário poderá efetuar consultas considerando toda a base de dados – de 01/01/1997 até a última atualização - por períodos selecionáveis, considerando ou não determinados atributos, para ambos os tipos de base de dados: original e classificada. A Figura 6 mostra parte da tabela de resultados ao consultar toda a base de dados, considerando todos os atributos, em seu formato original. Essa é a configuração padrão do sistema.

A Figura 7 mostra a consulta desconsiderando os atributos área e fluxo de rádio no período de 01/02/2018 a 11/02/2018.

Figura 6 - Consulta à base de dados no formato original.

SARA Base de dados Regras de Associação Sobre

Mostrar

Download CSV

Year	Month	Day	Region	Area	Sunspot Number	Sunspot Magnetic Classification	X-Ray Flares	Radio Flux
1997	Jan	5	8009	10	3	beta		
1997	Jan	6	8009	20	5	beta		
1997	Jan	7	8009	10	5	beta		
1997	Jan	8	8009	0	2	alpha		
1997	Jan	10	8010	20	2	beta		
1997	Jan	11	8010	30	3	beta		
1997	Jan	12	8010	20	2	beta		
1997	Jan	16	8011	30	4	beta		
1997	Jan	17	8011	20	6	beta		
1997	Jan	18	8011	0	2	alpha		
1997	Jan	19	8011	0	1	alpha		
1997	Jan	19	8012	0	2	beta		

Fonte: Produção do autor.

Figura 7 – Consulta à base de dados original, alguns atributos e período.

SARA Base de dados Regras de Associação Sobre

Mostrar

Download CSV

Year	Month	Day	Region	Sunspot Number	Sunspot Magnetic Classification	X-Ray Flares
2018	Feb	01	2697	03	Alpha	
2018	Feb	02				
2018	Feb	03	2698	01	Alpha	
2018	Feb	04				
2018	Feb	05	2699	01	Alpha	
2018	Feb	06	2699	03	Beta	B8.1
2018	Feb	07	2699	07	Beta	
2018	Feb	08	2699	10	Beta-Gamma	
2018	Feb	09	2699	12	Beta-Gamma	
2018	Feb	10	2699	13	Beta	
2018	Feb	11	2699	25	Beta	B1.3

Fonte: Produção do autor.

Os mesmos mecanismos se aplicam ao se consultar a base de dados em seu formato classificado, apenas com uma ressalva: não foram encontradas na literatura, métricas para a elaboração de uma classificação do número de

manchas, portanto, esse atributo não pode ser consultado em seu formato classificado aparecendo desativado nessa opção. A Figura 8 mostra esse comportamento.

Figura 8 - Base de dados classificada.

The screenshot shows the SARA application interface. At the top, there is a navigation bar with the following items: SARA, Base de dados, Regras de Associação, and Sobre. Below the navigation bar, there is a main content area. At the top of this area, it says "Informações da base de dados: Último dia na base de dados 2018-11-04" with a green "Atualizada" button next to it. The main content area is divided into two columns. The left column is titled "Tipo de Base de dados" and has two radio buttons: "Original" (selected) and "Classificada" (selected). Below this, there is a section titled "Atributos" with three checked checkboxes: "Área", "Número de manchas solares", and "Classificação magnética da mancha". The right column is titled "Período" and has two date input fields: "Começo" with the value "01/01/1997" and "Fim" with the value "04/11/2018". Below these fields, there is a checked checkbox labeled "Use toda a base de dados". At the bottom right of the main content area, there is a blue "Pesquisar" button. At the bottom center of the main content area, there is a blue "Esconder" button.

Fonte: Produção do autor.

A Figura 9 mostra a tabela de resultados da consulta à base de dados classificada, sem o atributo número de manchas.

No Menu Regras de Associação o usuário pode criar regras considerando atributos, níveis de suporte e confiança e período. A configuração padrão para essa consulta considera todos os atributos por todo período disponível da base de dados com níveis de suporte e confiança a 0.01, que equivale a 1%. A Figura 10 mostra essa tela.

Figura 9 - Consulta à base de dados classificada, sem número de manchas.

Year	Month	Day	Region	Area	Sunspot Magnetic Classification	X-Ray Flares	Radio Flux
2000	Mar	3	8882	area_gigantesca	beta_gamma	raiox_X	radio_altissimo
2000	Mar	3	8882	area_gigantesca	beta_gamma		radio_altissimo
2000	Mar	3	8882	area_gigantesca	beta_gamma	raiox_M	radio_altissimo
2000	Mar	3	8882	area_gigantesca	beta_gamma	raiox_C	
2000	Mar	3	8883	area_pequena	alpha		
2000	Mar	3	8886	area_pequena	beta		
2000	Mar	3	8888	area_pequena	beta		
2000	Mar	3	8889	area_media	beta_gamma		
2000	Mar	3	8891	area_gigantesca	beta_gamma		
2000	Mar	3	8892	area_pequena	beta		
2000	Mar	3	8896	area_pequena	beta		
2000	Mar	3	8897	area_pequena	beta		

Fonte: Produção do autor.

Figura 10 - Menu Regras de Associação.

Informações da base de dados: Último dia na base de dados 2018-11-04 Atualizada

Atributos

- Área
- Classificação magnética da mancha
- Explosões por Raio-X
- Fluxo de rádio

Suporte

0.01

Confiança

0.01

Período

Começo: 01/01/1997

Fim: 04/11/2018

Use toda a base de dados

Gerar

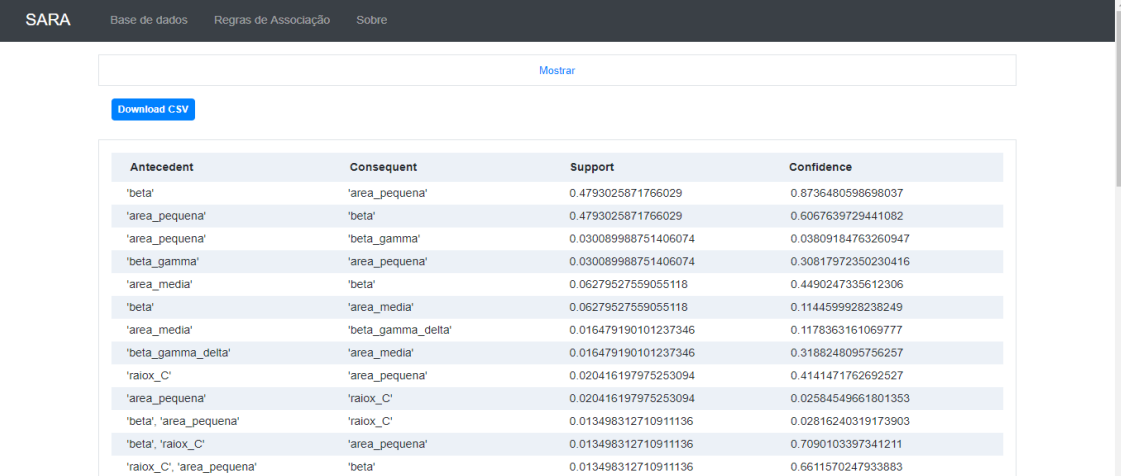
[Esconder](#)

Fonte: Produção do autor.

A Figura 11 mostra a tabela com as regras criadas de acordo com os parâmetros selecionados.

No menu Base de dados e no menu Regras de Associação, onde são efetuadas consultas, é possível exportar a consulta realizada num relatório em formato CSV através do botão “Download CSV”, como mostrado na Figura 7.

Figura 11 - Regras de associação geradas.



Antecedent	Consequent	Support	Confidence
'beta'	'area_pequena'	0.4793025871766029	0.8736480598698037
'area_pequena'	'beta'	0.4793025871766029	0.6067639729441082
'area_pequena'	'beta_gamma'	0.030089988751406074	0.03809184763260947
'beta_gamma'	'area_pequena'	0.030089988751406074	0.30817972350230416
'area_media'	'beta'	0.06279527559055118	0.4490247335612306
'beta'	'area_media'	0.06279527559055118	0.1144599928238249
'area_media'	'beta_gamma_delta'	0.016479190101237346	0.1178363161069777
'beta_gamma_delta'	'area_media'	0.016479190101237346	0.3188248095756257
'raiox_C'	'area_pequena'	0.020416197975253094	0.4141471762692527
'area_pequena'	'raiox_C'	0.020416197975253094	0.02584549661801353
'beta', 'area_pequena'	'raiox_C'	0.013498312710911136	0.02816240319173903
'beta', 'raiox_C'	'area_pequena'	0.013498312710911136	0.7090103397341211
'raiox_C', 'area_pequena'	'beta'	0.013498312710911136	0.6611570247933883

Fonte: Produção do autor.

No menu Sobre, mostrado pela Figura 12, há uma breve descrição sobre o que é o sistema, suas funcionalidades e os conceitos relacionados.

Figura 12 - Menu Sobre.



Fonte: Produção do autor.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALEXANDER, D. **Guide to the Universe: The Sun**. Westport, USA: Greenwood Press, 2009. 228 p. ISBN10 0313340773.
- ARGENTO, R. S. V. **Utilização de Ensembles de Redes Neurais MLP para Previsão de Explosões Solares**. 2016. 98 f. Dissertação (Mestrado) - Curso de Mestre em Tecnologia, área de Tecnologia e Inovação, Faculdade de Tecnologia - Universidade Estadual de Campinas, Campinas, 2016.
- COLAK, T.; QAHWAJI, R. Automated Prediction of Solar Flares Using Neural Networks and Sunspots Associations. **Advances in Soft Computing**. v.39, p. 316-324, 2007.
- DIERCKXSENS, M. et al. Relationship between Solar Energetic Particles and Properties of Flares and CMEs: Statistical Analysis of Solar Cycle 23 Events. **Solar Physics**, v.290, n.3, p.841-874, 2015.
- FLASK. Flask RESTful. Disponível em: <https://flask-restful.readthedocs.io/en/latest/>. Acesso em: 01 dez. 2018.
- GITHUB. GitHub. Disponível em: <https://github.com/>. Acesso em: 10 out. 2018.
- HAN, J. et al.. Data Mining: Concepts and Techniques. 3. ed. Morgan Kaufmann, 2012. (The Morgan Kaufmann Series in Data Management Systems).
- LEE, D. et al. Discovering Relationships between Factors of Round-trip Car Sharing by Using Association Rules Approach. In: WORLD MULTIDISCIPLINARY CIVIL ENGINEERING-ARCHITECTURE-URBAN PLANNING SYMPOSIUM– WMCAUS, 161., 2016, Praga. **Procedia Engineering**. Elsevier, 2016, p.1282 - 1288.
- LOSKUTOV, A. et al. Testing and forecasting the time series of the solar activity by singular spectrum analysis. **ArXiv**. 2001, Disponível em: ArXiv preprint nlin/0010027, Acesso em: 2001.
- MIGLANI, S. Konstant Info. Top 10 JavaScript Frameworks for 2016. Disponível em: <https://www.konstantinfo.com/blog/top-10-javascript-frameworks-for-2016/>. Acesso em: 20 out. 2018.
- MLXTEND. MLXtend Documentation. Disponível em: <https://rasbt.github.io/mlxtend>. Acesso em: 10 out. 2018.
- NGUYEN, T. T. et al. Learning Sunspot Classification. IOS Press. p. 1-15. 2016.

NPM. NPM Package Library. Disponível em: <https://www.npmjs.com/>. Acesso em: 20 out. 2018.

PYTHON. Python Documentation. Disponível em: [<https://docs.python.org/3/>](https://docs.python.org/3/). Acesso em: 10 out. 2018.

SCHWENN, R. Space Weather: The Solar Perspective. **Living Reviews in Solar Physics**, v.3, n.1, p.2-72, 2006.

SPACE Weather Prediction Center: NOAA / NWS Space Weather Prediction Center. 2015. Disponível em: <http://www.swpc.noaa.govpub/warehouse>>. Acesso em: 02 jul. 2018.

"Sunspots". NASA. 1 Abril 1998, Página Web consultada em 21/12/2018, Disponível em: <https://image.gsfc.nasa.gov/poetry/workbook/sunspot.html>>. Acesso em: 2018.

TAN, P. et al. Association Analysis: Basic Concepts and Algorithms. In: TAN, Pang-ning; STEINBACH, Michael; KUMAR, Vipin. Introduction to Data Mining. Pearson, 2005. Cap. 6. p. 328-414.

VASCONCELOS, L. M. R. de; CARVALHO, C. L. de. Aplicação de Regras de Associação para Mineração de Dados na Web. Goiás: Instituto de Informática Universidade Federal de Goiás, 2004.

YOUNG, C. Alex. Solar Activity. 2010. Disponível em: <http://www.thesuntoday.org/the-sun/solar-activity/>>. Acesso em: 10 set. 2018.