



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Campus de São José dos Campos
Instituto de Ciência e Tecnologia

PREVISÃO DE ALAGAMENTOS VIA INTEGRAÇÃO DE TWEETS, DADOS DE RADAR, PLUVIOMÉTRICOS E ALGORITMOS DE APRENDIZADO DE MÁQUINA (TEMPORÁRIO)

Aluno: Vitor Yuichi Hossaki

Orientador: Prof. Dr. Rogério Galante Negri

Projeto de Bolsa de Iniciação Científica apresentado à FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo.

Outubro de 2021

SUMÁRIO

| | <u>Pág.</u> |
|---|-------------|
| 1 Introdução | 1 |
| 2 Objetivos | 3 |
| 3 Justificativa e Relevância | 4 |
| 4 Fundamentação Teórica | 4 |
| 4.1 Alagamentos | 5 |
| 4.2 A utilização de redes sociais para o monitoramento de eventos | 5 |
| 4.3 Classificação | 6 |
| 4.3.1 SVM | 7 |
| 4.3.2 RF | 8 |
| 4.3.3 Redes Neurais | 8 |
| 5 Materiais e Métodos | 9 |
| 5.1 Área de estudo e dados disponíveis | 9 |
| 5.2 Ferramentas Computacionais | 10 |
| 5.3 Proposta de algoritmo para definição do modelo | 11 |
| 6 Plano de Atividades e Cronograma | 12 |
| 6.1 Plano de atividades | 12 |
| 6.2 Cronograma | 13 |
| REFERÊNCIAS BIBLIOGRÁFICAS | 14 |

1 Introdução

Os alagamentos são fenômenos cada vez mais frequentes em regiões urbanas devido ao aumento da população e o crescimento desordenado do processo de urbanização. Em São Paulo, os alagamentos são recorrentes desde os primórdios de sua ocupação,

a estrutura urbana aliado às características dos rios existentes auxiliam na deflagração destes fenômenos (HIRATA et al., 2013). Segundo Santos (2013), estima-se que os efeitos macroeconômicos dos alagamentos são de R\$172,3 milhões por ano, afetando setores logísticos e industriais.

Diante dos impactos materiais, econômicos e humanos causados pelos alagamentos, é necessário medidas que visem a mitigação e antecipação deste fenômeno. Uma destas medidas são os sistemas de alertas, essencial para que a comunidade seja alertada com antecedência de fenômenos naturais intensos e, desta forma, minimizar e prevenir possíveis danos materiais e humanos (KOBAYAMA et al., 2006).

Nesse ínterim, algumas pesquisas como as apresentadas por Horita et al. (2015) e Hirata et al. (2013), demonstram que a utilização de redes sociais que contém informações geográficas voluntárias podem ser um instrumento efetivo no desenvolvimento de sistemas de monitoramento e alerta de alagamento.

De modo similar, Albuquerque et al. (2015) demonstra a potencialidade de dados espaço-temporais obtidos da rede social Twitter para a gestão de desastres. Na pesquisa citada, as postagens são filtradas através de palavras chaves e então classificadas de forma binária dentre as classes “relacionado” ou “não-relacionada”, que por sua vez são alinhadas a uma série temporal de deflagrações dos alagamentos em determinada região. Por conseguinte, os dados são analisados através de modelos lineares generalizados. Os resultados obtidos indicam a relevância estatísticas dos Tweets relacionados às regiões de alagamento.

Em tempo, cabe destacar a recente ascensão quanto ao emprego de técnicas de Inteligência Artificial, principalmente dos métodos de Aprendizado de Máquina, na gestão e tomada de decisão perante eventos de desastres. Na pesquisa de Sit et al. (2019) são utilizados algoritmos de processamento de linguagem natural para identificação semântica de Tweets relacionados ao contexto de desastres. O Aprendizado

de Máquina é também incorporado ainda na análise de múltiplos parâmetros físicos úteis na previsão de alagamentos (MOSAVI et al., 2018).

Dessa forma, no contexto da problemática que envolve os eventos de alagamento e do potencial oferecido pelos métodos de Aprendizado de Máquina, este projeto de pesquisa tem o objetivo de construir um algoritmo capaz de prever a ocorrência de alagamentos através de informações extraídas automaticamente da rede social Twitter. Dados obtidos por radar meteorológico, pluviômetro e a base de dados de alagamentos são componentes que integram o algoritmo proposto.

A estrutura geral desta proposta contém: a Seção 2 apresenta os objetivos principal e específicos relacionados à pesquisa; a Seção 3 expõe a relevância e importância do tema proposto; discussões teóricas essenciais ao entendimento desta proposta são incluídas na Seção 4; a Seção 6.1 inclui questões relacionadas à delimitação da área de estudo, disponibilidade/origem dos dados, ferramentas computacionais, bem como apresenta uma concepção inicial do algoritmo proposto para previsão de alagamentos; por fim, a Seção 6 discute o plano de atividade e seu cronograma de execução.

2 Objetivos

A finalidade deste projeto é o desenvolvimento de um modelo de Aprendizado de Máquina para a predição de alagamentos, utilizando-se dos diferentes algoritmos propostos pela pesquisa. Neste projeto será utilizado séries temporais da rede social Twitter, a base de dados de alagamentos em São Paulo, e dados pluviométricos e meteorológicos.

Objetivos específicos:

- a) Definir qual é o algoritmo mais preciso para relacionar os dados meteoroló-

gicos, pluviométricos e frequência de tweets, para o modelo de classificação de alagamentos em dias de "ocorrência" e "não-ocorrência".

- b) Estabelecer o atributo com maior relevância estatística para predição de alagamentos dentre os propostos, combinando-os de forma conjunta com os algoritmos.
- c) Publicação dos resultados em congressos, eventos e periódicos científicos.
- d) Estruturar e disponibilizar as bases de dados e códigos em um repositório público.

3 Justificativa e Relevância

O desenvolvimento acelerado de São Paulo culminou na urbanização descontrolada causando diversas consequências na região. A impermeabilização do solo, a drenagem urbana deficitária e a topografia favorável ao acúmulo de água, são reflexos desta expansão desordenada. A deflagração de alagamentos causaram diversas perdas diretas e indiretas nos mais diversos setores da economia no estado. Em vista das consequências que os alagamentos vêm causando ao longo das décadas, é necessário medidas mitigadoras e o desenvolvimento de tecnologias que possam antecipar a possível deflagração do fenômeno hidrológico.

Além disso, em virtude da ascensão meteórica da tecnologia e da influência das redes sociais, o projeto visa o desenvolvimento da computação aplicada aos fenômenos hidrológicos de alagamentos, utilizando-se conceitos de ciência de dados e Aprendizado de Máquina.

4 Fundamentação Teórica

4.1 Alagamentos

Os alagamentos são fenômenos associados ao acúmulo de água em determinado local, favorecidos pela microdrenagem e macrodrenagem insuficientes. A ausência de planejamento urbano e a rápida modificação do espaço culmina na impermeabilização do solo, contribuindo para diminuição do tempo concentração e o aumento do volume de escoamento superficial, amplificando-se assim, o pico da vazão e consequentemente saturando a drenagem pluvial do local ([HANSMANN, 2013](#)).

A topografia e a elevação do local também são fatores preponderantes para a ocorrência de alagamentos, ou seja, verifica-se que os lugares com maior frequência de alagamentos tem características morfométricas planas, depressões ou fundos de vales, dificultando o processo de escoamento superficial do local ([BRAGA, 2016](#)).

Fatores como o descarte inadequado de resíduos sólidos, podem causar obstrução dos sistema de drenagem do local, isto ocorre em decorrência da ausência de educação ambiental da população.

Os alagamentos são fenômenos complexos, uma vez que a sua causa está interrelacionada a uma gama parâmetros como o clima que incluem precipitação forte ou persistente, falhas estruturais urbanas, sistemas de drenagem inadequados, bacias hidrográficas, proximidades de corpos aquáticos, uso e ocupação inapropriado do solo e entre outros ([DOOCY et al., 2013](#)).

4.2 A utilização de redes sociais para o monitoramento de eventos

O desenvolvimento da sociedade na esfera tecnológica permitiu a ascensão meteórica das redes sociais e suas funcionalidades. A quantidade massiva de dados gerados das redes sociais consolidam a interação do universo virtual com o mundo concreto, onde usuários expressam suas percepções e emoções acerca dos eventos circundan-

tes (NAAMAN, 2011). A atividade das redes sociais e sua heterogeneidade espacial demonstra a potencialidade para o monitoramento de eventos meteorológicos como a precipitação (ANDRADE et al., 2021).

Através das plataformas de mídia social, uma única postagem pode ser vista por milhares de usuários simultaneamente, além disso algumas plataformas utilizam-se de georreferenciamento que permite a visualização não só da postagem como também a localização do usuário com seu dispositivo móvel. Redes sociais (Twitter) ou aplicativos como Open Street Map que permitem a tecnologia de georreferenciamento são denominadas informações geográficas voluntárias, o trabalho de (HORITA et al., 2015), integra estas plataformas para o gerenciamento de risco dos alagamentos.

A utilização das redes sociais apresentam uma crescente tendência na sua incorporação em pesquisas para o monitoramento e análise de uma infinidade de eventos. Segundo (ALBUQUERQUE et al., 2015), a utilização de informações geográficas voluntárias, principalmente a rede social Twitter, são componentes fundamentais para a maior consciência dos eventos ocorrentes ou seja, consolida-se a percepção dos elementos no ambiente e possibilita maior compreensão das possíveis consequências.

4.3 Classificação

O Aprendizado de Máquina é cada vez mais empregado pelos pesquisadores na área de desastres de naturais, alguns autores utilizam esta ferramenta para analisar a semântica atrelada das postagens de rede social, e assim, aprimorar os resultados da classificação de determinada ocorrência (ALBUQUERQUE et al., 2015; DEPARDAY et al., 2019).

Esta tecnologia pode ser definida como um conjunto métodos computacionais para aprimorar performance ou realizar previsões acuradas. A classificação é um dos métodos computacionais amplamente utilizados para categorização de cada item em uma série de dados. Matematicamente, a classificação é descrita por uma função

$F : \mathcal{X} \rightarrow \mathcal{Y}$ que associa elementos no conjunto de atributos \mathcal{X} a uma classe de $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, com $n \in \mathbb{N}^*$, e partindo-se de um indicador de classe $\mathcal{Y} = \{1, 2, \dots, n\}$, portanto, quando $x \in \mathcal{X}$ e $y \in \mathcal{Y}$, a função $y = F(x)$ indica que x pertence à Ω_y .

Os modelos de aprendizagem supervisionada, a função F utiliza-se das informações do conjunto de treinamento representado pela equação $\mathcal{D} = \{(x_j, \omega_j \in \times \Omega : i = 1, \dots, m; j = 1, \dots, c)\}$, no qual m é a quantidade de dados no treinamento.

Atualmente os principais algoritmos empregados para classificação são o *Support Vector Machine* (SVMs) e *Random Forest* (RF) (MOHRI et al., 2018). e as redes neurais? (MLP?)

4.3.1 SVM

O método SVM realiza a distinção entre amostras de treinamento partindo-se de um hiperplano que possui maior abrangência de separação, mapeando o padrão de vetores para um espaço de alta dimensão, determinando-se o hiperplano mais adequado para separação de dados. Este algoritmo é utilizado por diversos autores devido à alta acurácia para problemas de classificação binária (LIAN; LU, 2006) .

O hiperplano corresponde ao lugar geométrico nos quais a função $f(x) = \langle w, x \rangle + b$ é nula. A variável w é o vetor ortogonal ao hiperplano e b a distância entre a função e a origem do espaço de atributos.

Para se encontrar o hiperplano mais adequado para separação entre as classes, é necessário a resolução do problema de otimização (THEODORIDIS et al., 2010) represen-

tado por: $max_{\gamma} (\sum_{i=1}^m \gamma_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \gamma_i \gamma_j y_i y_j \langle x_i, x_j \rangle)$,
$$\begin{cases} 0 \leq \gamma_i \leq C, i = 1, \dots, m \\ \sum_{i=1}^m \gamma_i y_i = 0 \end{cases},$$

a variável C é o parâmetro utilizado para regularização para ajustar o hiperplano e γ_i são os multiplicadores de Lagrange.

A definição dos parâmetros w e b que compõem o hiperplano são dadas por: $w = \sum_{x_i \in SV} y_i \gamma_i x_i$, $b = \frac{1}{\#SV} (\sum_{x_i \in SV} y_i + \sum_{x_i \in SV} \cdot \sum_{x_j \in SV} \gamma_i \gamma_j y_i y_j \langle x_i, x_j \rangle)$, SV é um subconjunto das amostras de treinamento \mathcal{D} , nos quais os elementos são os vetores suporte. Por fim indicação da classe pertencente do vetor analisado é dado pelo sinal da função discriminante $f(x)$ (MASELLI; NEGRI, 2019).

4.3.2 RF

A classificação através do algoritmo Floresta Aleatória vem sendo amplamente utilizada na literatura para avaliação e mapeamento dos padrões de eventos hidrológicos. Pesquisa como de (ZHU; ZHANG, 2021) e (LIU et al., 2020) demonstram a potencialidade do algoritmo para avaliar a resiliência e os padrões espaciais dos alagamentos.

Este modelo é um algoritmo de classificação que representa um conjunto de árvores de decisão, que combina a saída destas diversas árvores atribuindo-se uma classe ao conjunto de dados. Segundo (BREIMAN, 2001), a Floresta Aleatória consiste em uma coleção de classificadores em forma de árvore descritos por $\{h(x, \theta_k), k = 1, \dots\}$ onde θ_k são independentes e em cada árvore é lançado um voto unitário para a classe mais popular para o input x .

4.3.3 Redes Neurais

Este algoritmo vem sendo empregado para emissão de alertas hidrológicos e mapeamentos de suscetibilidade em alguns autores como (SILVA et al., 2016) e (QUEVEDO et al., 2020), demonstrando efetividade e acurácia elevada para os modelos de previsão associados aos fenômenos hidrológicos.

A técnica *Multilayer Perceptron* demonstra resultados relevantes em mais diversas áreas da ciência (GARDNER; DORLING, 1998). Este algoritmo consiste em um sistema interconectado de neurônios, estes nós são conectados entre si por um peso. Matematicamente as camadas de neurônios de entrada e saída são vetores definidos

como i e O respectivamente, e os pesos como uma matriz W . Portanto a saída da rede é dada por $O = f(IW_{io})$, ao final do processo uma função determina se aquele nó será ativado na condição $f(x) \begin{cases} 1 & x > 0 \\ 0 & otherwise \end{cases}$. Assumindo que T é o parâmetro de saída para o vetor de treinamento, o algoritmo o calcula o erro associado através de $E(O) = T - O = T - f(IW_{io})$. Algumas técnicas visam a redução do erro através da atualização dos pesos no processo representado matematicamente por $W_{io}(t+1) = W_{io}(t) + \alpha E_n$.

5 Materiais e Métodos

5.1 Área de estudo e dados disponíveis

O estudo foi realizado na região de São Paulo onde está localizado a bacia hidrográfica do Rio Tamanduateí (Figura 5.1). Esta bacia possui uma área de $323km^2$ e se estende até as bacia hidrográficas do Rio Pinheiro, Rio Guaió, Rio Aricanduva e Córrego de Tapuapé. Nesta região, foi analisado a partir de um pluviômetro um raio espacial de $2000m$ que abrange as regiões de alagamentos, tweets georreferenciados e a célula de radar.

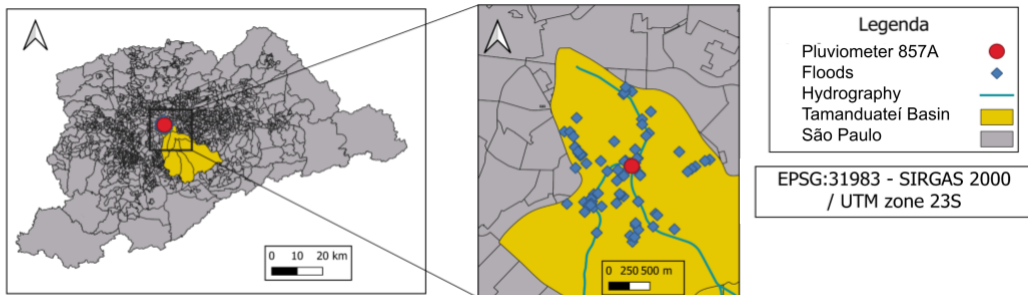


Figura 5.1 - Área de estudo

Os dados da rede social Twitter foram extraídas através da API (*Application Pro-*

gramming Interface). Os dados pluviométricos são coletados do pluviômetro 833A, pertencente ao Centro Nacional e Alertas e Desastres Naturais (CEMADEN), estes dados podem ser encontrados no próprio site da instituição.

A série histórica de alagamentos na área de estudo, foram concebidas por um dos integrantes da pesquisa. Os dados meteorológicos foram extraídos por estações pertencentes ao CEMADEN, o equipamento está localizado na cidade de São Roque - SP e atualmente está em operação pelo Departamento de Controle do Espaço Aéreo (DECEA). Esse radar tem alcance de 250 km, cobrindo toda a região metropolitana de São Paulo. O produto de radar usado para o CAPPI (Constant Altitude Plan Position Indicator) na altura de 3 km. Este produto possui uma resolução espacial de aproximadamente 1 km e uma resolução temporal de 10 minutos. Para a conversão da refletividade (dBZ) em taxa de separação (mm / h) foi utilizado em relação a Marshall-Palmer ([ANDW, 1948](#)) e a seguir os dados foram acumulados por dia.

5.2 Ferramentas Computacionais

A análise e aplicação do projeto será realizada de maneira geral com a ferramenta *Python*. Para a manipulação, filtragem e tratamento dos dados será utilizada a biblioteca *Pandas* ([VANDERPLAS, 2016](#)).

A aplicação de testes estatísticos na série de dados será usado *Scipy* ([VIRTANEN et al., 2020](#)) e *Numpy* ([MCKINNEY, 2012](#)), para os modelos de aprendizagem supracitados, a biblioteca específica para aprendizado de máquina denominado Scikit-Learn ([PEDREGOSA et al., 2011](#)). Por fim, algumas filtrações no banco de dados de alagamentos será realizada com ferramentas de geoprocessamento do software *QGIS* ([SAMELA et al., 2018](#)).

5.3 Proposta de algoritmo para definição do modelo

A concepção inicial deste trabalho é analisar as séries temporais dos alagamentos, tweets, pluviômetro e radar, para definir quais são o melhor conjunto de parâmetros em dias de alagamentos, associando-se ao número mínimo necessário de tweets para emissão de um alerta.

A série temporal analisada compreende os três primeiros meses do ano de 2019. Para a base de dados dos tweets, o processamento consiste no recorte temporal e filtração do tweets com base na lista de palavras associadas ao contexto meteorológico e hidrológico. Esta lista de palavra basea-se no trabalho de (ANDRADE et al., 2021).

Com base na estrutura (Figura 5.2), será registrado em único arquivo, na mesma série temporal, o número de tweets filtrados, os valores de precipitação do radar e o pluviômetro, e se houve alagamentos no dias analisados. Este dados processados em um único arquivo, possibilitarão a submissão nos modelos de aprendizados propostos, dividindo-se em base de dados para teste e treinamento.

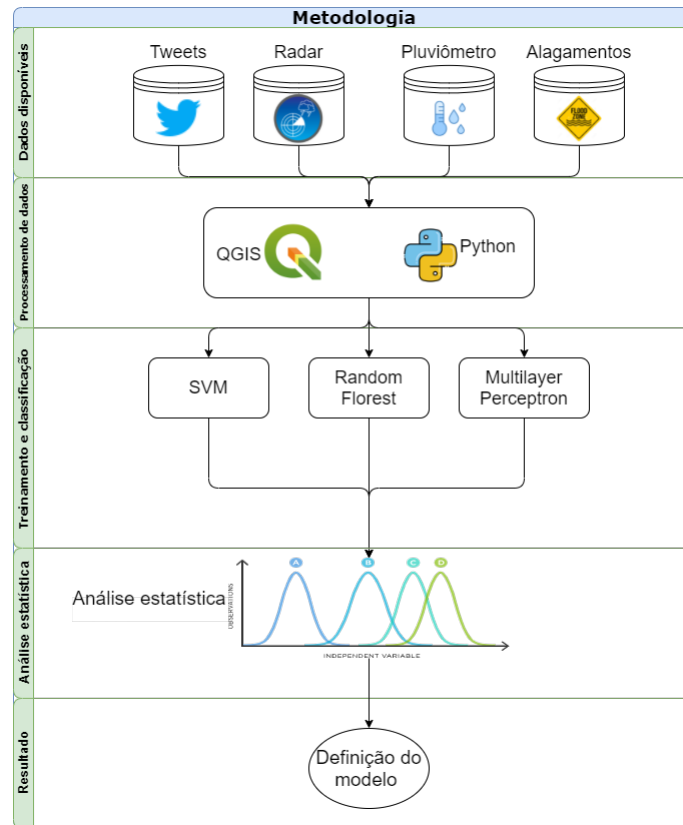


Figura 5.2 - Metodologia

Como a classificação binária consiste em dias de alagamento e não alagamento, a acurácia será medida a partir da base de dados de teste. Após o treinamento nos modelo SVM e Floresta Aleatória e Redes Neurais, serão analisadas a acurácia através da validação cruzada e subsequentemente testes estatísticos como ANOVA e coeficiente Kappa, determinando-se assim, o algoritmo que possui maior potencial para o desenvolvimento de um sistema de alerta com base nos dados disponíveis.

6 Plano de Atividades e Cronograma

6.1 Plano de atividades

O projeto tem início com a revisão sistemática com relação à literatura acerca dos fenômeno hidrológico de alagamento e os principais algoritmos de Aprendizado de

Máquina aplicados na área. Paralelamente, será realizado o estudo da linguagem de programação Python e do software de geoprocessamento QGIS.

Após o estudo dos conceitos necessários para o projeto, os dados de Twitter, pluviômetro, radar meteorológico e a base de dados de alagamentos serão tratados e processados, para estruturar uma série temporal com o conjunto de dados. Para analisar as relações estatísticas primárias entre os dados, será realizada uma análise exploratória através de gráficos e cálculos de parâmetros estatísticos.

Ainda na etapa supracitada, os resultados decorrentes da análise exploratória serão submetidos aos algoritmos de treinamento, nos quais serão realizadas as classificações binárias em dias de "alagamento" e "não-alagamento", com os respectivos algoritmos propostos. Após a classificação, será gerado um conjunto de dados através da validação cruzada, que incluem a acurácia do algoritmo utilizando bases de treinamento e testes distintos. Para a avaliação destes resultados, testes estatísticos citados na Seção serão utilizados para determinar o modelo com maior acurácia para classificação de alagamentos.

Dessa forma, com os resultados e inferências em mãos, serão realizados ajustes nos códigos e otimizações para possíveis melhorias na acurácia dos algoritmos e velocidade de execução. E assim, aprimorando de forma geral o modelo desenvolvido.

Por fim, através da obtenção do modelo e conclusões, os resultados serão escritos em trabalhos para divulgação em eventos científicos.

6.2 Cronograma

A pesquisa será realizada em 12 meses e será executada nos passos listados abaixo. A Tabela 6.1 exibe o respectivo cronograma de execução.

- A - Revisão sistemática em desastres associados à alagamentos e modelos de classificação;
- B - Estudo dos modelos de aprendizado de máquina e aplicação em Python;

- C - Processamento dos bancos de dados;
- D - Análise exploratória dos dados processados;
- E - Submissão dos dados processados para treinamento nos modelos propostos;
- F - Classificação;
- G - Cálculos estatísticos e inferências;
- H - Alterações, ajustes e otimizações no modelo de melhor desempenho;
- I - Análise e conclusão dos resultados;
- J - Relatório final

Tabela 6.1 - Cronograma

| Mês | | 1º | 2º | 3º | 4º | 5º | 6º | 7º | 8º | 9º | 10º | 11º | 12º |
|--------|---|----|----|----|----|----|----|----|----|----|-----|-----|-----|
| Etapas | A | • | • | | | | | | | | | | |
| | B | • | • | • | | | | | | | | | |
| | C | | | • | • | | | | | | | | |
| | D | | | | • | | | | | | | | |
| | E | | | | | • | • | | | | | | |
| | F | | | | | • | • | | | | | | |
| | G | | | | | | | • | • | | | | |
| | H | | | | | | | | • | • | | | |
| | I | | | | | | | | | | • | • | |
| | J | | | | | | | | | | | • | • |

REFERÊNCIAS BIBLIOGRÁFICAS

ALBUQUERQUE, J. P. D.; HERFORT, B.; BRENNING, A.; ZIPF, A. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. **International journal of geographical information science**, Taylor & Francis, v. 29, n. 4, p. 667–689, 2015. [2](#), [6](#)

ANDRADE, S. C. de; ALBUQUERQUE, J. Porto de; RESTREPO-ESTRADA, C.; WESTERHOLT, R.; RODRIGUEZ, C. A. M.; MENDIONDO, E. M.;

DELBEM, A. C. B. The effect of intra-urban mobility flows on the spatial heterogeneity of social media activity: investigating the response to rainfall events. **International Journal of Geographical Information Science**, Taylor & Francis, p. 1–26, 2021. [6](#), [11](#)

ANDW, J. M. Mc. k. palmer, the distribution of raindrops with size, J. **Meteor**, v. 5, p. 165–166, 1948. [10](#)

BRAGA, J. O. Alagamentos e inundações em áreas urbanas: estudo de caso na cidade de santa maria-rs. 2016. [5](#)

BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001. [8](#)

DEPARDAY, V.; GEVAERT, C. M.; MOLINARIO, G.; SODEN, R.; BALOG-WAY, S. Machine learning for disaster risk management. World Bank, 2019. [6](#)

DOOCY, S.; DANIELS, A.; MURRAY, S.; KIRSCH, T. D. The human impact of floods: a historical review of events 1980-2009 and systematic literature review. **PLoS currents**, Public Library of Science, v. 5, 2013. [5](#)

GARDNER, M. W.; DORLING, S. Artificial neural networks (the multilayer perceptron) a review of applications in the atmospheric sciences. **Atmospheric environment**, Elsevier, v. 32, n. 14-15, p. 2627–2636, 1998. [8](#)

HANSMANN, H. Z. Descrição e caracterização das principais enchentes e alagamentos de pelotas-rs. **Universidade Federal de Pelotas, Pelotas-RS**, 2013. [5](#)

HIRATA, E.; GIANNOTTI, M. A.; LAROCCA, A. P. C.; QUINTANILHA, J. A. Mapeamento dinâmico e colaborativo de alagamentos na cidade de são paulo. **Boletim de Ciências Geodésicas**, SciELO Brasil, v. 19, p. 602–623, 2013. [2](#)

HORITA, F. E.; ALBUQUERQUE, J. P. de; DEGROSSI, L. C.; MENDIONDO, E. M.; UHEYAMA, J. Development of a spatial decision support system for flood risk management in brazil that combines volunteered geographic information with wireless sensor networks. **Computers & Geosciences**, Elsevier, v. 80, p. 84–94, 2015. [2](#), [6](#)

KOBIYAMA, M.; MENDONÇA, M.; MORENO, D. A.; MARCELINO, I.; MARCELINO, E. V.; GONÇALVES, E. F.; BRAZETTI, L. L.; GOERL, R. F.; MOLLERI, G. S.; RUDORFF, F. d. M. **Prevenção de desastres naturais: conceitos básicos**. [S.l.]: Organic Trading Curitiba, 2006. [2](#)

LIAN, H.-C.; LU, B.-L. Multi-view gender classification using local binary patterns and support vector machines. In: SPRINGER. **International Symposium on Neural Networks**. [S.l.], 2006. p. 202–209. [7](#)

LIU, D.; FAN, Z.; FU, Q.; LI, M.; FAIZ, M. A.; ALI, S.; LI, T.; ZHANG, L.; KHAN, M. I. Random forest regression evaluation model of regional flood disaster resilience based on the whale optimization algorithm. **Journal of Cleaner Production**, Elsevier, v. 250, p. 119468, 2020. [8](#)

MASELLI, L. Z.; NEGRI, R. G. Integração entre estratégias multiclases e diferentes funções kernel em máquinas de vetores suporte para classificação de imagens de sensoriamento remoto. **Revista Brasileira de Cartografia**, v. 71, n. 1, p. 149–175, 2019. [8](#)

MCKINNEY, W. **Python for data analysis: Data wrangling with Pandas, NumPy, and IPython**. [S.l.]: "O'Reilly Media, Inc.", 2012. [10](#)

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. **Foundations of machine learning**. [S.l.]: MIT press, 2018. [7](#)

MOSAVI, A.; OZTURK, P.; CHAU, K.-w. Flood prediction using machine learning models: Literature review. **Water**, Multidisciplinary Digital Publishing Institute, v. 10, n. 11, p. 1536, 2018. [3](#)

NAAMAN, M. Geographic information from georeferenced social media data. **SIGSPATIAL Special**, ACM New York, NY, USA, v. 3, n. 2, p. 54–61, 2011. [6](#)

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. et al. Scikit-learn: Machine learning in python. **the Journal of machine Learning research**, JMLR. org, v. 12, p. 2825–2830, 2011. [10](#)

QUEVEDO, R. P.; OLIVEIRA, G. Garcia de; GUASSELLI, L. A. Mapeamento de suscetibilidade a movimentos de massa a partir de redes neurais artificiais. **Anuario do Instituto de Geociencias**, v. 43, n. 2, 2020. [8](#)

SAMELA, C.; ALBANO, R.; SOLE, A.; MANFREDA, S. A gis tool for cost-effective delineation of flood-prone areas. **Computers, Environment and Urban Systems**, Elsevier, v. 70, p. 43–52, 2018. [10](#)

SANTOS, E. T. d. **Impactos econômicos de desastres naturais em megacidades: o caso dos alagamentos em São Paulo**. Tese (Doutorado) — Universidade de São Paulo, 2013. [2](#)

SILVA, M. R. da; SANTOS, L. B. L.; SCOFIELD, G. B.; CORTIVO, F. D. Utilização de redes neurais artificiais em alertas hidrológicos: Estudo de caso na bacia do rio claro em caraguatatuba, estado de são paulo. **Anuário do Instituto de Geociências**, v. 39, n. 1, p. 23–31, 2016. [8](#)

SIT, M. A.; KOYLU, C.; DEMIR, I. Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of hurricane irma. **International Journal of Digital Earth**, Taylor & Francis, 2019. [2](#)

THEODORIDIS, S.; PIKRAKIS, A.; KOUTROUMBAS, K.; CAVOURAS, D.

Introduction to pattern recognition: a matlab approach. [S.l.]: Academic Press, 2010. [7](#)

VANDERPLAS, J. **Python data science handbook: Essential tools for working with data.** [S.l.]: "O'Reilly Media, Inc.", 2016. [10](#)

VIRTANEN, P.; GOMMERS, R.; OLIPHANT, T. E.; HABERLAND, M.; REDDY, T.; COURNAPEAU, D.; BUROVSKI, E.; PETERSON, P.; WECKESSER, W.; BRIGHT, J. et al. Scipy 1.0: fundamental algorithms for scientific computing in python. **Nature methods**, Nature Publishing Group, v. 17, n. 3, p. 261–272, 2020. [10](#)

ZHU, Z.; ZHANG, Y. Flood disaster risk assessment based on random forest algorithm. **Neural Computing and Applications**, Springer, p. 1–13, 2021. [8](#)