



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 METEOROLÓGICOS E ALGORITMOS DE APRENDIZADO DE MÁQUINA

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 Redes sociais para o monitoramento de alagamentos	5
4.3 Classificação	6
4.3.1 SVM	7
4.3.2 RF	8
4.3.3 MLP	8
5 Materiais e Métodos	8
5.1 Área de estudo e dados disponíveis	9
5.2 Ferramentas Computacionais	10
5.3 Proposta de algoritmo para definição do modelo	10
6 Plano de Atividades e Cronograma	12
6.1 Plano de atividades	12
6.2 Cronograma	13
REFERÊNCIAS BIBLIOGRÁFICAS	13

1 Introdução

Os alagamentos são fenômenos cada vez mais frequentes em regiões urbanas devido ao aumento da população e ao crescimento desordenado. Na cidade de São Paulo os alagamentos são recorrentes desde os primórdios de sua ocupação. A estrutura

urbana aliada 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, torna-se necessárias ações que visem a mitigação e antecipação deste fenômeno. Exemplos de ações são os sistemas de alertas, essenciais para que a comunidade seja informada com antecedência de fenômenos naturais intensos e, desta forma, minimizar e/ou 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 munidas de informações geográficas voluntárias podem ser empregadas como 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 através 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 àquelas “relacionadas” ou “não-relacionadas”, 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 das informações apontadas pelos Tweets em relação às regiões alagadas.

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. Em 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 na análise de múltiplos parâmetros físicos úteis na previsão de alagamentos, conforme mostrado por [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 está organizada na seguinte forma: 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 5 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

Este projeto de pesquisa tem como objetivo principal o desenvolvimento de um método para previsão de alagamentos com base em técnicas de Aprendizado de Máquina, dados temporais extraídos de Tweets e instrumentos meteorológicos como radares e pluviômetros.

Como objetivos específicos, são elencados:

- a) Analisar e definir o algoritmo de Aprendizado de Máquina de melhor de-

sempenho para previsão de alagamentos a partir de dados meteorológicos e extraídos de Tweets;

- b) Investigar sobre os atributo de maior relevância para predição de alagamentos;
- c) Estruturar e disponibilizar as bases de dados e códigos em repositório público;
- d) Publicação dos resultados obtidos em eventos e periódicos científicos.

3 Justificativa e Relevância

O projeto tem relevância nas pesquisas em desastres quanto no desenvolvimento de técnicas computacionais. De modo específico:

- No ponto de vista dos desastres, a deflagração de alagamentos causam diversas perdas diretas e indiretas em diferentes setores da economia. Em vista das consequências deste fenômeno, faz-se necessário o desenvolvimento de medidas mitigadoras capazes de antecipar a possível deflagração do fenômeno hidrológico em questão;
- Em virtude da ascensão de tecnologia e sua presença no cotidiano da população, este projeto contempla o desenvolvimento de métodos computacionais direcionados aos fenômenos de alagamento com suporte de dados advindos/extraídos da redes sociais.

4 Fundamentação Teórica

4.1 Alagamentos

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

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 assim o pico da vazão e consequentemente saturando a drenagem pluvial do local (HANSMANN, 2013). A topografia local também é um fator preponderante para a ocorrência de alagamentos, uma vez que verificado que locais com maior frequência de alagamentos tem características morfométricas planas, depressões ou fundos de vales, dificultando o processo de escoamento superficial local (BRAGA, 2016). Ainda, devido a ausência de educação ambiental da população, o descarte inadequado de resíduos sólidos surge como outro fator causador da obstrução dos sistema de drenagem do local, levando mais uma vez aos eventos de alagamento.

4.2 Redes sociais para o monitoramento de alagamentos

O desenvolvimento da sociedade na esfera tecnológica permitiu a ascensão meteórica das redes sociais e suas funcionalidades. A quantidade massiva de dados gerados por estas redes consolidam a interação do universo virtual com o mundo concreto, onde usuários expressam suas percepções e emoções acerca dos eventos circundantes (NAMAN, 2011). A atividade das redes sociais e sua heterogeneidade espacial demonstra a potencialidade para o monitoramento de eventos meteorológicos (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 fazem uso de dados georreferenciados, permitindo por sua vez a visualização não só da postagem como também da localização do usuário. Redes sociais como Twitter ou aplicativos como OpenStreetMap, os quais são munidos de georreferenciamento são ditos “de informações geográficas voluntárias”. Neste contexto, 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 quanto a sua incorporação em pesquisas para o monitoramento e análise de diferentes tipos de eventos. Segundo [Albuquerque et al. \(2015\)](#), a utilização de informações geográficas voluntárias, principalmente advindas da rede social Twitter, são componentes fundamentais para a maior consciência dos eventos de desastres, consolidando assim uma melhor percepção do ambiente além de possibilitar maior compreensão das possíveis consequências.

4.3 Classificação

Técnicas de Aprendizado de Máquina são cada vez mais empregadas nos estudos sobre desastres de naturais. Alguns estudos utilizam tais técnicas para analisar a semântica atrelada às postagens em redes sociais, permitindo assim aprimorar os resultados de classificação de determinada tipo de evento ([ALBUQUERQUE et al., 2015](#); [DEPARDAY et al., 2019](#)).

A classificação compreende os métodos utilizados para associação de cada item em uma série de dados em relação a uma determinada classe. Formalmente, 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 $n \in \mathbb{N}^*$ por intermédio de um indicador de classe $\mathcal{Y} \in \{1, 2, \dots, n\}$. Nestas condições, quando $x \in \mathcal{X}$ e $y \in \mathcal{Y}$, a função $y = F(x)$ indica que x pertence à Ω_y .

Em relação aos modelos de classificação com aprendizado supervisionado, a função F faz uso de informações extraídas de conjunto de treinamento $\mathcal{D} = \{(x_j, \omega_j) \in \mathcal{X} \times \Omega : i = 1, \dots, m; j = 1, \dots, c\}$. Dentre diversas propostas existentes na literatura, *Support Vector Machine* (SVMs), *Random Forest* (RF) (MOHRI et al., 2018) e *Multilayer Perceptron* (MLP) são empregados frequentemente nos mais diversos domínios de aplicação. ===== Atualmente os principais algoritmos empregados para classificação são o *Support Vector Machine* (SVMs), *Random Forest* (RF) e *Multilayer Perceptron* (MLP) (MOHRI et al., 2018).

4.3.1 SVM

O método SVM realiza a distinção entre exemplos de treinamento com base em um hiperplano com maior margem de separação, seja no espaço original dos dados ou remapeado de forma conveniente. Segundo Lian e Lu (2006), este método é utilizado por diversos autores devido a sua alta acurácia e capacidade de generalização.

O hiperplano corresponde ao lugar geométrico onde $f(x) = \langle w, x \rangle + b$ é nula. A variável w é o vetor ortogonal ao hiperplano e b a distância entre o hiperplano e a origem do espaço de atributos. A determinação do hiperplano com maior margem de separação é obtido com a otimização do problema (THEODORIDIS et al., 2010):

$$\begin{aligned} & \max_{\gamma} \left(\sum_{i=1}^m \gamma_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \gamma_i \gamma_j y_i y_j \langle x_i, x_j \rangle \right) \\ & \text{sujeito a } \begin{cases} 0 \leq \gamma_i \leq C, i = 1, \dots, m \\ \sum_{i=1}^m \gamma_i y_i = 0 \end{cases} \end{aligned} \quad (4.1)$$

onde C é o parâmetro utilizado para regularização do hiperplano e γ_i são multiplicadores de Lagrange.

A definição dos parâmetros que determinam o hiperplano são $w = \sum_{x_i \in SV} y_i \gamma_i x_i$ e $b = \frac{1}{\#SV} (\sum_{x_i \in SV} y_i + \sum_{x_i \in SV} \cdot \sum_{x_j \in SV} \gamma_i \gamma_j y_i y_j \langle x_i, x_j \rangle)$, onde SV é um subconjunto

das amostrar em \mathcal{D} denominados por 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

O método RF tem sido amplamente utilizada em aplicações relacionadas a identificação de eventos hidrológicos. Em Zhu e Zhang (2021) e Liu et al. (2020) são demonstradas as potencialidades deste método na avaliação da resiliência e identificação dos padrões espaciais de alagamentos. De modo superficial, segundo Breiman (2001), este método é representado por um conjunto de árvores de decisão que combinam suas respectivas saídas através de um esquema de votação por maioria a fim de tomar uma decisão final.

4.3.3 MLP

O método MLP tem sido empregado para emissão de alertas hidrológicos e mapeamentos de suscetibilidade (SILVA et al., 2016; QUEVEDO et al., 2020).

Este método compreende em um sistema de conexões ponderadas entre neurônios artificiais distribuídos em diferentes camadas. Matematicamente as camadas de neurônios de entrada e saída são vetores definidos como \mathbf{i} e \mathbf{o} respectivamente, e os pesos por uma matriz \mathbf{W} . Netas condições, a saída da rede é dada por $\mathbf{o} = f(\mathbf{iW})$, onde

$$f(x) \begin{cases} 1, & x > 0 \\ 0, & \text{caso contrário} \end{cases}, \text{ atua como uma função de ativação para a entrada } x \text{ apre-}$$

sentada. Assumindo que \mathbf{t} é consiste na saída esperada, é calculado o erro associado através de $E(\mathbf{o}) = \mathbf{t} - \mathbf{o} = \mathbf{t} - f(\mathbf{iW})$. O processo de treinamento do MLP é dados pela atualização dos pesos em \mathbf{W} a fim de minimizar os erros $E(\mathbf{o})$.

5 Materiais e Métodos

5.1 Área de estudo e dados disponíveis

Este projeto admite como área de estudo a região de São Paulo inserida na bacia hidrográfica do Rio Tamanduateí (Figura 5.1). Esta bacia possui uma área de 323 km² e se estende até as bacias hidrográficas do Rio Pinheiro, Rio Guaió, Rio Aricanduva e Córrego de Tapuapé. Esta área foi definida a partir da vizinhança de um pluviômetro, segundo um raio espacial de 2000 m, que por sua vez abrange diferentes regiões de alagamento, Tweets disponíveis e a célula de um radar meteorológico.

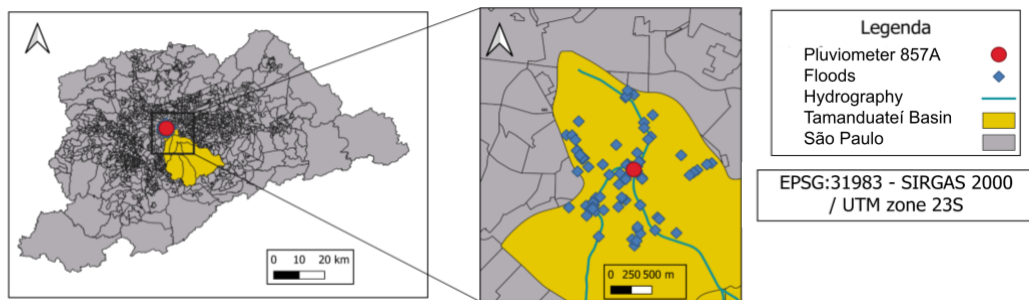


Figura 5.1 - Área de estudo

Os dados da rede social Twitter foram extraídas através de API (*Application Programming Interface*) disponibilizada pela própria rede social. Os dados pluviométricos são coletados do pluviômetro #833A, pertencente ao Centro Nacional e Alertas e Desastres Naturais (CEMADEN), os quais são disponibilizados publicamente pela 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). O equipamento monitora deslocamento, atuação de nuvens e núcleos de

instabilidade, mensurando o volume de precipitação em determinado local. 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

O desenvolvimento do projeto será pautado por programação via linguagem *Python*. A manipulação, filtragem e tratamento dos dados será suportada pelas bibliotecas *Pandas* (VANDERPLAS, 2016) e *Numpy* (MCKINNEY, 2012).

Para a aplicação de testes estatísticos, será feito uso da biblioteca *Scipy* (VIRTANEN et al., 2020). De forma semelhante, os métodos de classificação empregados na pesquisa serão obtidos da biblioteca Scikit-Learn (PEDREGOSA et al., 2011).

Por fim, operações necessárias sobre banco de dados serão realizada com apoio do Sistema de Informação Geográfica *QGIS* (SAMELA et al., 2018).

5.3 Proposta de algoritmo para definição do modelo

A concepção inicial desta proposta de pesquisa consiste no uso de séries temporais dos alagamentos, Tweets, e dados registrados por pluviômetro e radar meteorológico, a fim de construir um método para previsão de eventos de alagamento.

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 metereológico 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.

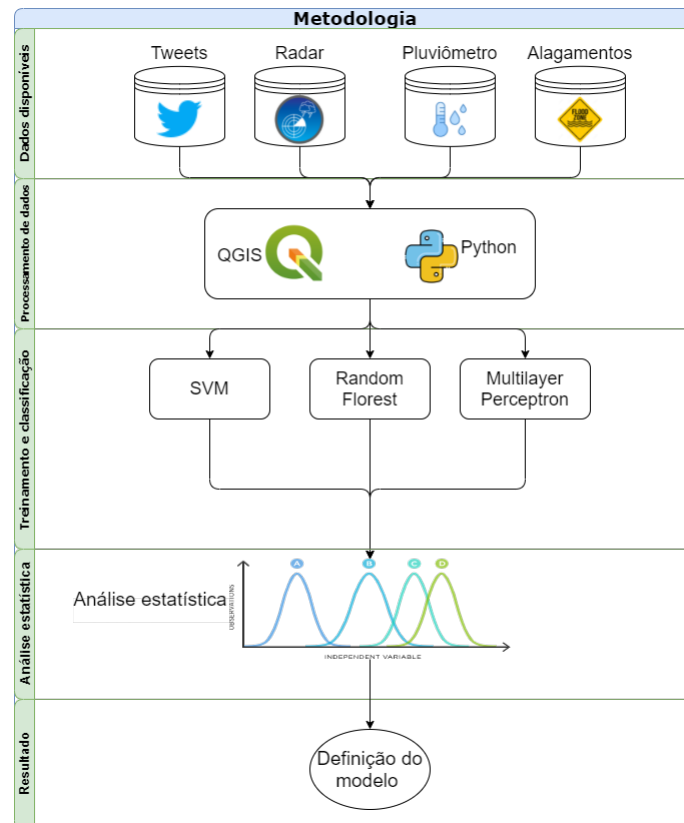


Figura 5.2 - Metodologia

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 realizado 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 realizados 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 5 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 listado 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. [5](#), [10](#)

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-df. 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.

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.; UEYAMA, 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. [5](#)

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](#)