

## DESENVOLVIMENTO DE UM MODELO DE PREVISÃO DE CHUVA UTILIZANDO ÁRVORE DE DECISÃO

FRIEDRICH, Thalía Lopes<sup>1</sup>; CASSOL, Fabiano<sup>2\*</sup>

<sup>1</sup> UFFS, Acadêmica do curso de Engenharia Ambiental e Sanitária, Universidade Federal da Fronteira Sul, *Campus Cerro Largo*, RS, Brasil.

<sup>2</sup> UFFS, Docente do curso de Engenharia Ambiental e Sanitária, Universidade Federal da Fronteira Sul, *Campus Cerro Largo*, RS, Brasil.

\*Autor Correspondente: [fabiano.cassol@uffs.edu.br](mailto:fabiano.cassol@uffs.edu.br)

### RESUMO

O avanço na área de aprendizado de máquina possibilita o desenvolvimento de tecnologias para facilitar a previsão de chuva mesmo que em escala local. Diversas pesquisas apontam métodos de aprendizado de máquina como uma boa alternativa para a análise exploratória, classificação e previsão de dados utilizando grandes bancos de dados, como é o caso das informações meteorológicas. Nesse contexto, o objetivo deste artigo consiste no desenvolvimento e avaliação de um modelo para prever a ocorrência de precipitação pluviométrica local utilizando o método Árvore de Decisão. Foram selecionadas combinações de variáveis para o treinamento do modelo (temperatura do ar, umidade relativa, ponto de orvalho, pressão atmosférica, velocidade do vento, direção do vento, irradiação solar e precipitação) registradas pela estação meteorológica situada no *Campus Cerro Largo* da Universidade Federal da Fronteira Sul (UFFS) ao longo de um período de 3 anos (2016 a 2019). Foram realizados testes para determinação dos parâmetros que retornassem o melhor desempenho na previsão de ocorrência de chuva, por exemplo, a proporção na divisão de dados entre treinamento e validação do modelo. Após tratamento dos dados e treinamento do modelo, este foi avaliado a partir da matriz confusão, acurácia, precisão, especificidade, sensibilidade e *F1-score*. Os resultados demonstraram que o método utilizado pode ser considerado eficiente, retornando bons índices de acertos, por exemplo, 86,9 %, de acurácia e 89,9 % de *F1-score* nos dados de validação. Portanto, conclui-se que este estudo reforça a aplicabilidade de algoritmos de aprendizado de máquina nos estudos meteorológicos, apresentando resultados potencialmente satisfatórios.

**Palavras chave:** *Machine Learning*, Classificação de Dados, Precipitação Pluviométrica.

## DEVELOPMENT OF A RAIN FORECAST MODEL USING DECISION TREE

### ABSTRACT

The advance in machine learning allows the development of technologies to facilitate rain forecast even on a local scale. Several researches have pointed out machine learning methods as a good alternative for exploratory analysis, classification and data prediction using large databases, such as meteorological information. In this context, the objective of this article is to develop and evaluate a model to predict the occurrence of local rainfall using the Decision Tree method. Combinations of variables were selected for the model training (air temperature, relative humidity, dew point, atmospheric pressure, wind speed, wind direction, solar irradiation and precipitation) recorded by the weather station located at the Cerro Largo Campus of the Universidade Federal da Fronteira Sul (UFFS) over a period of 3 years (2016 to 2019). Tests were conducted to determine the parameters that would return the best performance in predicting the occurrence of rain, for example, the proportion in the data division between training and model validation. After processing the data and training the model was evaluated from the matrix confusion, accuracy, precision, specificity, sensitivity and F1 score. The results showed that the method used can be considered efficient, returning good hit rates, for example, 86.9%, accuracy and 89.9% F1-score in the validation data. Therefore, it is concluded that this study reinforces the applicability of machine learning algorithms in meteorological studies, presenting potentially satisfactory results.

**Keywords:** Machine Learning, Data Classification, Rainfall Precipitation.

### 1 INTRODUÇÃO

O aprendizado de máquina é uma das áreas de pesquisa mais ativas no campo da Inteligência Artificial. São diversos métodos de programação capazes de possibilitar conhecimento a uma máquina, tornando-a capaz de realizar previsões e classificações com base na experiência obtida por meio de um processo de aprendizagem (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012).

Um dos métodos de aprendizado de máquina é denominado Árvore de Decisão. Em geral, pode ser classificado em dois tipos: árvores de classificação, utilizada para prever variáveis discretas; e árvores de regressão, para quando as variáveis em questão são contínuas. O método objetiva desenvolver um modelo que preveja a classe ou o valor de um fator

dependente, aprendendo por meio de um processo de treinamento. A árvore de classificação é um dos métodos mais utilizados pela comunidade científica devido as suas vantagens (YEON; HAN; RYU, 2010).

Em estudo realizado por Freitas, França e Menezes (2019) na região metropolitana do Rio de Janeiro, Brasil, foi utilizado Árvore de Decisão para avaliar a ocorrência de eventos convectivos a partir de dados de descargas atmosféricas e de índices termodinâmicos. Os autores afirmam que esse método foi aplicado devido a seu histórico de gerar bons resultados em relação a classificação supervisionada, além de sua praticidade.

Um modelo de Árvore de Decisão foi elaborado por Marjanović et al. (2018) para detectar as condições de chuva que desencadearam deslizamentos de terra em uma região da Sérvia. Para o treinamento, foram utilizados dados de precipitação pluviométrica e registros de deslizamentos de terra. O método foi capaz de reconhecer os fatores que mais influenciaram na ocorrência de deslizamentos de terra, corroborando para demonstrar o bom desempenho em tarefas como, por exemplo, encontrar padrões de comportamento em dados históricos.

Moon et al. (2019) constataram que as técnicas de aprendizado de máquina podem ser implementadas para elaborar um sistema de previsão de chuvas fortes, utilizando como exemplos para o aprendizado dados obtidos de estações meteorológicas automáticas regionais, como registros de vento, temperatura, umidade do ar, pressão atmosférica e quantidade de precipitação. No estudo, foi realizado um comparativo entre várias técnicas de aprendizado, entre elas, a Árvore de Decisão.

Considerando os estudos meteorológicos, a estimativa de precipitação pluviométrica é um dos serviços que mais contribuem na sociedade. Conforme Nayak, Mahapatra e Mishra (2013), esse serviço é fundamental para o planejamento e gestão dos recursos hídricos, sistemas de drenagem pluvial, operação de reservatórios e prevenção de inundações, uma vez que exerce influência no escoamento superficial, infiltração e percolação. Além disso, possui direta influência em diversas atividades humanas, como a agricultura, o transporte aéreo e marítimo, influenciando, assim, na economia local.

Neste contexto, este trabalho tem como objetivo o desenvolvimento e a avaliação de um modelo utilizando o método de Árvore de Decisão para prever a ocorrência de precipitação pluviométrica na cidade de Cerro Largo/RS/Brasil, a partir de dados de uma Estação Meteorológica local.

## 2 DESENVOLVIMENTO

Os dados meteorológicos utilizados neste estudo são procedentes da Estação Meteorológica pertencente e situada na Universidade Federal da Fronteira Sul *Campus Cerro Largo*, uma estação automática composta por um *hardware* da marca *Agrosystem*, modelo *Davis Vantage Pro2* e pelo *software Wunderground*. No desenvolvimento do algoritmo utilizou-se o módulo *DecisionTreeClassifier*, da biblioteca *Scikit-Learn*, no ambiente de programação *Jupyter Notebook*, acessado através do *software Anaconda (Python 3.7)*.

Considerando os dados disponíveis pela Estação Meteorológica, os parâmetros selecionados para estudo foram: temperatura do ar (°C), umidade relativa (%), ponto de orvalho (°C), pressão atmosférica (hPa), velocidade do vento (m/s), direção do vento (°) e irradiação solar (KJ/m<sup>2</sup>), além da precipitação (mm), que é o parâmetro de interesse a ser previsto. Portanto, o modelo foi treinado utilizando sete variáveis em um conjunto de dados registrados por um período três anos (julho de 2016 a junho de 2019) em intervalos de 10 minutos. O conjunto passou por um processo de tratamento de dados para exclusão de valores ausentes ou com falhas de medição, além da padronização de unidades de medida.

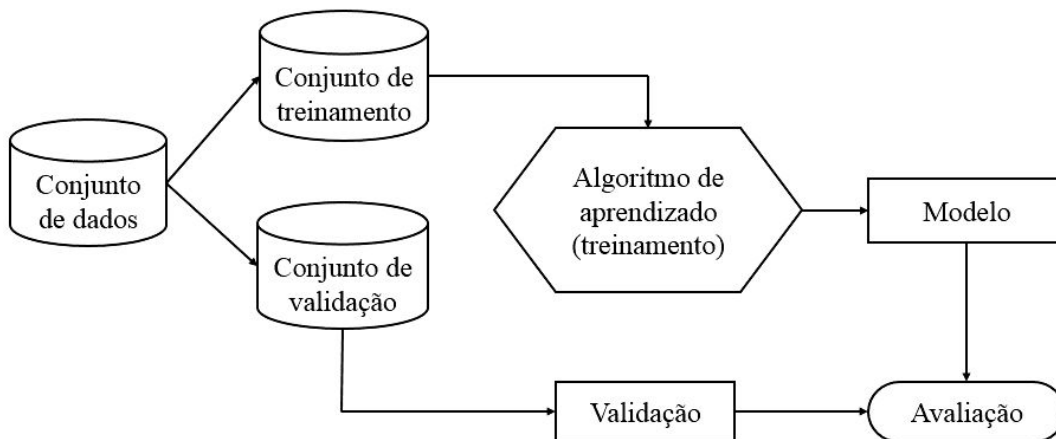
No desenvolvimento do modelo utiliza-se o conjunto de dados de treinamento. No entanto, na modelagem com aprendizagem de máquina, o mais importante é que a solução desenvolvida apresente bom desempenho diante de novos dados. Por essa razão, após o treinamento é pertinente avaliar o modelo com um conjunto de validação, dados até então desconhecidos pelo modelo. Além disso, realizou-se um segundo teste de avaliação, com um conjunto de dados chamados de teste, utilizando dados da Estação Meteorológica da UFFS de meses subsequentes ao período de treinamento (julho a dezembro de 2019), registrados em intervalos de 10 minutos, contendo as mesmas variáveis de entrada selecionadas para o treinamento e a validação.

### 2.1 APRENDIZADO DE MÁQUINA

O aprendizado de máquina pode ser classificado em aprendizado supervisionado ou não supervisionado (SATHYA; ABRAHAM, 2013). O método utilizado neste trabalho é classificado como supervisionado. A Figura 1 esquematiza o processo realizado desde a

divisão do conjunto de dados, separando-os entre treinamento e validação. Posteriormente, o conjunto de treinamento é implementado no algoritmo de aprendizado e utilizado para criação do modelo. Por fim, tanto o modelo quanto o conjunto de validação são submetidos à avaliação do desempenho.

Figura 1 - Processo de treinamento e validação de um aprendizado supervisionado.



Fonte: Adaptado de ROZA, 2016.

Ambos os conjuntos de dados são compostos por uma lista de exemplos, que são fatores independentes (as variáveis meteorológicas) e cada exemplo possui um fator dependente das demais variáveis. No caso de uma classificação, o fator dependente é dado por uma classe ou rótulo (a precipitação pluviométrica). Neste estudo, a precipitação originalmente era obtida em milímetros de chuva, portanto, foi convertida em duas classes: 1 e 0, sinalizando, respectivamente, a ocorrência ou não de chuva.

Dessa forma, após o treinamento, gera-se um modelo que é capaz de classificar novos dados. Assim, a finalidade do processo de treinamento é analisar as características e padrões presentes no conjunto de exemplos. A capacidade de classificação pode ser mensurada através de dados que se conhece previamente a qual classe estes pertencem, sendo possível confrontar com o resultado obtido pelo modelo (FERREIRA, 2016).

### 2.1.1 Árvore de Decisão

Árvore de decisão é um método que busca elaborar um modelo capaz de identificar a classe ou o estimar valor a partir do aprendizado de várias regras de decisão. Segundo

Hosokawa (2011), as árvores de decisão são formadas a partir de regras que são aprendidas por meio de um processo de divisão em um grande conjunto de dados, que ocorre sucessivamente. Em cada nova divisão, um grupo de dados são classificados e subdivididos conforme a presença de características em comum.

O algoritmo de uma árvore de decisão identifica um fator de particionamento e o utiliza para dividir o conjunto de dados (representado, na estrutura do modelo, por um nó), formando ramos (representado por linhas) para cada novo grupo de dados. Em cada ramo existe um nó interno que indica um novo fator de particionamento, subdividindo novamente o conjunto (denominado de nó terminal). Este processo ocorre inúmeras vezes de forma sucessiva e tem por finalidade separar os exemplos conforme suas características (GARCIA, 2003).

## 2.2 AVALIAÇÃO DO DESEMPENHO DO MODELO

A matriz confusão é uma ferramenta frequentemente utilizada em modelos de aprendizado de máquina que envolvam classificação de dados. É representada como uma tabela que mostra as frequências de classificações realizadas pelo modelo, dividindo os resultados em quatro categorias, segundo Castro e Ferrari (2016):

- Verdadeiros Positivos (VP): casos positivos que são previstos como positivos;
- Verdadeiros Negativos (VN): casos negativos que são previstos como negativos;
- Falsos Positivos (FP): casos negativos que são previstos como positivos;
- Falsos Negativos (FN): casos positivos que são previstos como negativos.

No modelo apresentado neste trabalho, os casos positivos representam os dados onde há ocorrência de chuva, enquanto que os casos negativos se referem aos dados onde não houve episódios de chuva. A partir da matriz confusão, são calculadas diversas métricas para avaliar o desempenho do modelo:

**Acurácia:** taxa de acertos do modelo, medindo a proporção de casos que foram previstos corretamente (CASTRO; FERRARI, 2016).

**Precisão, Sensibilidade e F1-score:** A precisão avalia os verdadeiros positivos em relação aos falsos positivos; enquanto que a sensibilidade é a proporção de verdadeiros positivos em relação aos falsos negativos, mensurando a habilidade do modelo para prever a

ocorrência de chuva. F1-score é uma métrica que equilibra a precisão e a sensibilidade, sendo calculada pela média harmônica entre elas (ZAINUDIN; JASIM; BAKAR, 2016).

Especificidade: refere-se à proporção de casos negativos que foram identificados corretamente (verdadeiros negativos) (DOU et al., 2019).

### **3 RESULTADOS E DISCUSSÃO**

Previamente às simulações são realizadas diversas operações relacionadas ao tratamento de dados (coleta, avaliação e limpeza de dados). Algumas informações necessitaram de tratamento, como por exemplo a direção do vento, que originalmente apresentava-se com caracteres (pontos cardeais) e foi necessário transformar para graus.

Outro procedimento efetuado previamente ao treinamento foi a normalização dos dados, sendo fundamental devido à grande variedade de escalas de unidades de medida e distribuições incomuns de valores que podem gerar instabilidade e dificultar o aprendizado computacional. A normalização consiste em transformar uma variável original de maneira que ela adquira outra escala, previamente determinada. Neste caso, os dados foram transformados para adquirirem valores entre 0 e 1.

Além disso, aplicou-se a técnica de redução de dados por seleção de atributos, (CASTRO; FERRARI, 2016), na qual efetuou-se a remoção parcial e aleatória de dados em que não havia precipitação pluviométrica, com a finalidade de manter quantidades semelhantes de dados com e sem ocorrência de chuva nos exemplos apresentados ao modelo. Após a redução, tem-se uma relação de aproximadamente 1:0,6 dados de precipitação para dados ausentes de chuva.

#### **3.1 DETERMINAÇÃO DOS PARÂMETROS DO MODELO**

A divisão do conjunto de dados foi determinada a partir de simulações variando a proporção de dados para treinamento e validação. Essa determinação é fundamental pois cada parâmetro de treinamento pode influenciar expressivamente nos resultados finais. O Quadro 1 apresenta os resultados observados em cada simulação, onde diversas métricas são avaliadas.

O ideal é buscar um equilíbrio entre as métricas, visto que para obter um bom desempenho é importante não apenas ter uma boa taxa de acerto (acurácia), mas também

apresentar bons índices na habilidade de previsão de ocorrência de chuva (como a precisão e a sensibilidade), além de prever corretamente a ausência de chuva (especificidade). Uma sugestão da literatura é considerar o *F1-score* como métrica principal, sendo a média harmônica entre duas métricas (precisão e sensibilidade), pode-se afirmar que é uma pontuação que representa bem o real desempenho do modelo.

Quadro 1 - Desempenho do modelo em função da proporção entre treinamento/validação.

Métricas	Proporção entre treinamento/validação								
	90/10	80/20	70/30	60/40	50/50	40/60	30/70	20/80	10/90
Precisão	0,876	0,867	0,858	0,863	0,860	0,857	0,852	0,837	0,839
Acurácia	0,866	0,869	0,860	0,858	0,857	0,857	0,850	0,845	0,835
Especificidade	0,778	0,761	0,742	0,756	0,752	0,743	0,737	0,697	0,710
Sensibilidade	0,918	0,935	0,931	0,919	0,920	0,926	0,919	0,934	0,911
<i>F1-score</i>	0,896	0,899	0,893	0,890	0,889	0,890	0,884	0,883	0,873

Fonte: Autores, 2020.

Quando 80 % dos dados foram utilizados para treinar o modelo e 20 % para validá-lo, observou-se o melhor *F1-score* (0,899), além de retornar o melhor índice de acurácia (0,869) e de sensibilidade (0,935). Assim, a divisão do conjunto de dados nas demais simulações realizadas neste trabalho foram fixadas com esta proporção.

No estudo realizado por Freitas, França e Menezes (2019), citado anteriormente, identificou-se que a melhor proporção de divisão de dados foi de 90 % para treinamento e 10 % para validação; assim, apresentou como resultados da avaliação do modelo de Árvore de Decisão, um total de 81,8 % de acertos em geral (casos classificados corretamente) e 18,2 % de alarmes falsos (casos classificados incorretamente).

Dessa forma, observa-se a importância de verificar qual proporção melhor se adapta aos exemplos de treinamento e validação, visto que parâmetros de treinamento, como este, podem influenciar significativamente no desempenho do modelo.

### 3.2 DESEMPENHO DO MODELO

Após o desenvolvimento e treinamento do modelo, o mesmo é avaliado com o conjunto de dados de validação (20 % do conjunto total de dados). A Tabela 1 representa a



matriz confusão gerada pelo modelo. Com ela, pode-se verificar a porcentagem de cada categoria em relação ao total de análises (2153 casos) previstas pelo modelo. No geral, o índice de acertos atingiu 86,9 % da amostra, contra 13,1 % de erros.

A maior porcentagem concentrou-se nos casos categorizados como verdadeiros positivos, ou seja, os casos em que o modelo previu corretamente que haveria ocorrência de chuva, com 1257 casos, representando 58,4 % da amostra. Os casos classificados como verdadeiros negativos são 615, compondo 28,6 % da amostra. As classificações efetuadas pelo modelo equivocadamente são 8,9 % de falsos positivos (193 casos) e 4,1 % de falsos negativos (88 casos).

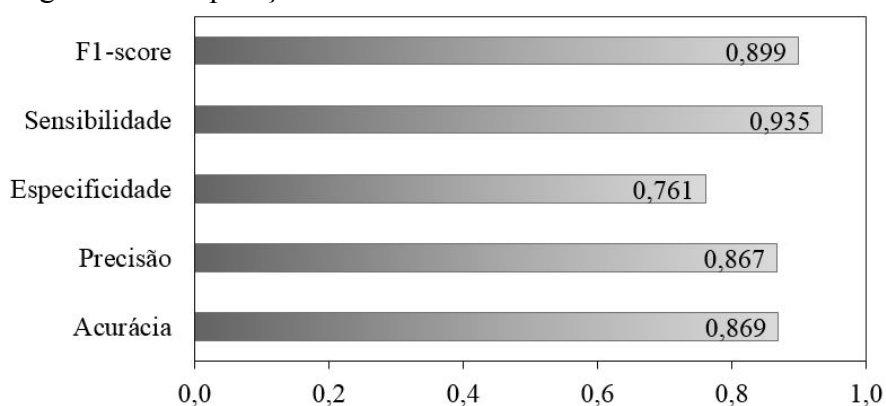
Tabela 1 - Matriz confusão (validação).

Classes	
1257 (VP)	88 (FN)
193 (FP)	615 (VN)

Fonte: Autores, 2020.

A Figura 2 exibe uma comparação entre as métricas analisadas na etapa de validação. Observa-se que 86,9 % do total de previsões foram corretas (acurácia). Dessas, a grande maioria foi a previsão de ocorrência de chuva, visto que a sensibilidade, que avalia esta condição, foi de 93,5 %; enquanto que a precisão foi de 86,7 %. A especificidade, que indica a proporção de casos sem ocorrência de precipitação previstos corretamente, retornou um valor de 76,1 %, indicando a presença de uma quantidade considerável de falsos positivos, ou seja, casos previstos como chuvosos quando não havia ocorrência de chuva. Por fim, a medida F1 com 89,9 %, evidencia o bom desempenho do método.

Figura 2 - Comparação das métricas avaliadas.



Fonte: Autores, 2020.

Para fins comparativos, observa-se o estudo realizado por Zainudin, Jasim e Bakar (2016), que analisaram diversos modelos classificadores entre eles, Árvore de Decisão para previsão de chuvas na Malásia, utilizando um conjunto de dados registrados por estações meteorológicas em Selangor, na Malásia no período de 2010 a 2014, contendo parâmetros como temperatura, umidade relativa, precipitação, entre outros. Entre as métricas de avaliação, os resultados obtidos com o modelo de Árvore de Decisão foram 74,9 % de precisão, 72,9 % de sensibilidade e 73,7 % de *F1-score*.

A relativa superioridade nos resultados do presente trabalho pode estar relacionada a diversos fatores, como as características do conjunto de dados em relação as variáveis utilizadas e a influência dos parâmetros de treinamento do modelo, por exemplo, a proporção utilizada na divisão do conjunto de dados, uma vez que no estudo referido tem-se 30 % de dados para treinamento e 70 % para validação. Além disso, considera-se que a técnica de redução de dados proporcionou uma melhora significativa nos resultados, sendo o fator principal para a obtenção de índices satisfatórios nas métricas avaliadas.

### 3.2.1 Desempenho do modelo em dados adicionais (teste)

Posteriormente ao treinamento e à validação, o modelo é submetido a dados de meses subsequentes (julho a dezembro de 2019), a fim de avaliar como seu comportamento sucede diante de novos dados. A Tabela 2 exhibe a matriz confusão gerada utilizando os dados de teste. Considerando o total de dados disponíveis para o teste (904), 42,1 % foram classificados como verdadeiros negativos (381 casos) e 25,5 % como verdadeiros positivos (230 casos); entre as classificações erradas, 25,2 % são falsos negativos (228 casos) e 7,2 % são falsos positivos (65 casos). No geral, 67,6 % das classificações foram realizadas corretamente, contra 32,4 % de previsões erradas.

Tabela 2 - Matriz confusão (teste).

		Classes	
	230 (VP)		228 (FN)
	65 (FP)		381 (VN)

Fonte: Autores, 2020.

A Tabela 3 apresenta os resultados em relação as métricas avaliadas. Verifica-se que ainda há uma boa taxa de acertos, com 67,6 % de acurácia.

Tabela 3 - Resultados obtidos com os dados de teste.

Métrica	Resultado
Precisão	0,780
Acurácia	0,676
Especificidade	0,854
Sensibilidade	0,502
F1-score	0,611

Fonte: Autores, 2020.

Porém, é perceptível que o desempenho decai na maioria das métricas, em comparação ao conjunto de validação. Apesar disso, nota-se um bom índice de especificidade (85,4 %) devido ao menor número de falsos positivos. Porém, a sensibilidade, com o menor índice (50,2 %), aponta que muitos casos onde havia precipitação não foram previstos como chuvosos, ou seja, há presença de muitos falsos negativos, que ocasiona redução neste índice.

## CONCLUSÃO

Neste trabalho avalia-se a potencialidade de um método de aprendizado de máquina (Árvore de Decisão) na tarefa de classificar e realizar previsão de eventos de precipitação na cidade de Cerro Largo/RS/Brasil.

Conclui-se que a metodologia utilizada neste trabalho pode ser utilizada como ferramenta auxiliar na classificação de dados meteorológicos, visto que foi possível registrar resultados satisfatórios, por exemplo, com 86,9 % de acurácia e 89,9 % de F1-score.

Portanto, de acordo com a análise da maioria das métricas nos conjuntos de dados de validação e de teste, os índices de acertos são bons, porém, deve ser considerada a presença de falsos positivos e negativos, que podem implicar na perda de confiabilidade do modelo.

## AGRADECIMENTOS

Os autores agradecem à Universidade Federal da Fronteira Sul pela concessão de bolsa de iniciação científica e apoio financeiro no projeto PES-2019-0644.

## REFERÊNCIAS

SATHYA, R; ABRAHAM, A. Comparison of supervised and unsupervised learning algorithms for pattern classification. **International Journal of Advanced Research in Artificial Intelligence**, v. 2, n. 2, 2013.

CASTRO, L. N.; FERRARI, D. G. **Introdução à Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações**. Editora Saraiva, 351 p., 2016.

DOU, J.; YUNUS, A. P; BUI, D; T.; MERGHADI, A.; SAHANA, M.; ZHU, Z.; CHEN, C.-W.; KHOSRAVI, K.; YANG, Y.; BINH THAI PHAM, B. T. Assessment of Advanced Random Forest and Decision Tree Algorithms for Modeling Rainfall-Induced Landslide Susceptibility in the Izu-Oshima Volcanic Island, Japan. **Science of the Total Environment** v. 662, p. 332-346, 2019.

FREITAS, J. H. V.; FRANÇA, G. B.; MENEZES, W. F. Previsão de Convecção Profunda Usando Árvore de Decisão na Região Metropolitana do Rio de Janeiro. **Anuário do Instituto de Geociências - UFRJ**, v. 42-1, p. 127-134, 2019.

FERREIRA, L. D. **Técnicas de aprendizado de máquina aplicadas à identificação de perfis de aprendizado em um ambiente real de ensino**. 2016. 73 p. Monografia (Mestrado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade Federal de São Paulo, São Carlos, 2016.

GARCIA, S. C. **O Uso de Árvore de Decisão na Descoberta de Conhecimento na Área da Saúde**. Monografia desenvolvida ao curso de Tecnologia em Processamento de Dados, da Faculdade de Tecnologia de São Paulo. 2003, 88 p. Dissertação (Programa de Pós-Graduação em Computação) – Universidade Federal do Rio Grande do Sul. Porto Alegre, 2003.

HOSOKAWA, E. O. **Técnica de Árvore de Decisão em Mineração de Dados**. 2011, 40 p. Monografia (Curso de Tecnologia em Processamento de Dados) - Faculdade de Tecnologia de São Paulo. São Paulo, 2011.

MARJANOVIĆ, M.; KRAUTBLATTER, M.; ABOLMASOV, B.; ĐURIĆ, U.; SANDIĆ, C.; NIKOLIĆ, V. The Rainfall-Induced Landsliding in Western Serbia: A Temporal Prediction Approach Using Decision Tree Technique. **Engineering Geology**, n. 232, p. 147-159, 2018.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. **Foundations of machine learning**. Cambridge. Massachusetts Institute of Technology - MIT Press, 427 p., 2012.

MOON, S-H; KIM, Y.-H.; LEE, Y. H.; MOON, B.-R. Application of Machine Learning to an Early Warning System for Very Shortterm Heavy Rainfall. **Journal of Hydrology**, v. 568, p. 1042-1054, 2019.

NAYAK, D. R.; MAHAPATRA, A.; MISHRA, P. A Survey on Rainfall Prediction using Artificial Neural Network. **International Journal of Computer Applications**, v. 72, n. 16, 2013.

ROZA, F. S. **Aprendizagem de máquina para apoio à tomada de decisão em vendas do varejo utilizando registros de vendas**. 2016. 62 p. Monografia. (Curso de Engenharia de Controle e Automação) - Universidade Federal de Santa Catarina, Florianópolis, 2016.

YEON, Y.-K.; HAN, J.-G.; RYU, K. H. Landslide Susceptibility Mapping in Injae, Korea, Using a Decision Tree. **Engineering Geology**, v. 116, p. 274-283, 2010.

ZAINUDIN, S.; JASIM, D. S.; BAKAR, A. A. Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction. **International Journal on Advanced Science Engineering Information Technology**, v. 6, n. 6. 2016.