

Aplicação de técnicas de aprendizado de máquina na previsibilidade da arrecadação da Compensação Financeira pela Exploração de Recursos Minerais (CFEM): Um estudo de caso no Município de Parauapebas, Estado do Pará (PA), Brasil

Application of machine learning techniques in the predictability of the revenue of the Financial Compensation for Mineral Exploitation (CFEM): A case study in the Municipality of Parauapebas, State of Pará (PA), Brazil

Aplicación de técnicas de aprendizaje automático en la previsibilidad de la recaudación de la Compensación Financiera por la Explotación de Recursos Minerales (CFEM): Un estudio de caso en el Municipio de Parauapebas, Estado de Pará (PA), Brasil

Recebido: 19/05/2025 | Revisado: 04/06/2025 | Aceitado: 05/06/2025 | Publicado: 07/06/2025

Keny Lucas da Silva Goes

ORCID: <https://orcid.org/0000-0002-5827-9388>

Universidade do Vale do Itajaí, Brasil

E-mail: keny.goes@edu.univali.br

Kaua Chrystian da Silva Goes

ORCID: <https://orcid.org/0009-0000-2038-5399>

Universidade Federal do Sul e Sudoeste do Pará, Brasil

E-mail: kaua.goes@unifesspa.edu.br

Resumo

O presente artigo tem como objetivo aplicar modelos supervisionados de aprendizado de máquina, mais especificamente os algoritmos Random Forest Regressor e K-Nearest Neighbors (KNN), para prever o valor arrecadado pela Compensação Financeira pela Exploração de Recursos Minerais (CFEM) no município de Parauapebas, Estado do Pará (PA), Brasil. Este artigo apresenta uma aplicação de modelos de aprendizado de máquina para a previsão da arrecadação da CFEM, com foco no município de Parauapebas. A pesquisa é caracterizada como aplicada, quantitativa e descritiva, com uso de algoritmos de regressão supervisionada — Random Forest e K-Nearest Neighbors — sobre dados públicos da Agência Nacional de Mineração (ANM), entre os anos de 2015 e 2024. Os resultados demonstraram que o modelo Random Forest teve melhor desempenho preditivo em comparação ao KNN, alcançando coeficiente de determinação (R^2) superior a 0,90. A previsão para o mês de janeiro de 2025 indicou uma arrecadação superior a R\$ 100 milhões, evidenciando o potencial da modelagem para subsidiar o planejamento orçamentário municipal. Conclui-se que o uso de técnicas de inteligência artificial pode contribuir significativamente para a gestão pública em municípios mineradores, especialmente em cenários de alta dependência da receita da CFEM.

Palavras-chave: CFEM; Previsão de receita; Mineração; Parauapebas; Machine learning.

Abstract

This article aims to apply supervised machine learning models, specifically the Random Forest Regressor and K-Nearest Neighbors (KNN) algorithms, to predict the revenue collected by the Financial Compensation for Mineral Exploitation (CFEM) in the Municipality of Parauapebas, State of Pará (PA), Brazil. This article presents an application of machine learning models to forecast CFEM revenue, focusing on the municipality of Parauapebas. The research is characterized as applied, quantitative, and descriptive, using supervised regression algorithms — Random Forest and K-Nearest Neighbors — applied to public data from the National Mining Agency (ANM), from 2015 to 2024. The results showed that the Random Forest model had the best predictive performance compared to KNN, achieving a coefficient of determination (R^2) above 0.90. The forecast for January 2025 indicated a revenue exceeding BRL 100 million, highlighting the potential of predictive modeling to support municipal budget planning. It is concluded that artificial intelligence techniques can significantly contribute to public management in mining municipalities, especially in scenarios with high dependence on CFEM revenues.

Keywords: CFEM; Revenue forecasting; Mining; Parauapebas; Machine learning.

Resumen

El presente artículo tiene como objetivo aplicar modelos supervisados de aprendizaje automático, más específicamente los algoritmos Random Forest Regressor y K-Nearest Neighbors (KNN), para prever el valor recaudado por la Compensación Financiera por la Explotación de Recursos Minerales (CFEM) en el Municipio de Parauapebas, Estado de Pará (PA), Brasil. Este artículo presenta una aplicación de modelos de aprendizaje automático para prever la recaudación de la CFEM, centrándose en el municipio de Parauapebas. La investigación se caracteriza como aplicada, cuantitativa y descriptiva, utilizando algoritmos de regresión supervisada — Random Forest y K-Nearest Neighbors — sobre datos públicos de la Agencia Nacional de Minería (ANM), entre 2015 y 2024. Los resultados demostraron que el modelo Random Forest tuvo un mejor rendimiento predictivo en comparación con KNN, alcanzando un coeficiente de determinación (R^2) superior a 0,90. La previsión para el mes de enero de 2025 indicó una recaudación superior a los R\$ 100 millones, lo que destaca el potencial del modelado para apoyar la planificación presupuestaria municipal. Se concluye que las técnicas de inteligencia artificial pueden contribuir significativamente a la gestión pública en municipios mineros, especialmente en contextos con alta dependencia de los ingresos por CFEM.

Palabras clave: CFEM; Predicción de ingresos; Minería; Parauapebas; Aprendizaje automático.

1. Introdução

A mineração é uma das atividades econômicas mais relevantes para a balança comercial brasileira, especialmente no que tange à extração e exportação de minérios como o ferro, ouro, cobre e bauxita. Essa atividade é responsável por impulsionar o desenvolvimento de diversas regiões, sobretudo na região Norte, onde o estado do Pará destaca-se como líder nacional em produção mineral. No entanto, ao mesmo tempo em que promove crescimento econômico, a mineração gera forte dependência fiscal por parte dos municípios mineradores, que têm na Compensação Financeira pela Exploração Mineral (CFEM) uma de suas principais fontes de receita.

Parauapebas, localizado no sudeste paraense, é o principal exemplo dessa realidade. Com aproximadamente 220 mil habitantes, o município é fortemente influenciado pela operação da mineradora Vale S.A., que explora a maior jazida de minério de ferro a céu aberto do mundo: a Serra dos Carajás. Em razão disso, Parauapebas ocupa há anos o primeiro lugar no ranking de arrecadação da CFEM, conforme dados da Agência Nacional de Mineração (ANM). Em contrapartida, essa dependência torna a gestão pública local vulnerável às flutuações de mercado, variações cambiais e mudanças na legislação do setor mineral.

Nesse contexto, torna-se fundamental desenvolver mecanismos que permitam maior previsibilidade dessas receitas, a fim de auxiliar os gestores públicos no planejamento orçamentário, na execução de políticas públicas e na mitigação de riscos fiscais. A previsibilidade da CFEM, no entanto, é um desafio, uma vez que está condicionada a variáveis como volume de produção, preço internacional do minério e logística de escoamento – fatores muitas vezes externos ao controle municipal.

Diante dessa problemática, técnicas de aprendizado de máquina (machine learning – ML) surgem como uma abordagem promissora para o desenvolvimento de modelos preditivos capazes de estimar a arrecadação futura da CFEM com base em dados históricos. Essas técnicas têm se mostrado eficazes em aplicações no setor industrial, financeiro e governamental, conforme apontam Han, Kamber e Pei (2012) e Goodfellow et al. (2016), podendo ser adaptadas para contextos locais de gestão pública.

O presente artigo tem como objetivo aplicar modelos supervisionados de aprendizado de máquina, mais especificamente os algoritmos Random Forest Regressor e K-Nearest Neighbors (KNN), para prever o valor arrecadado pela CFEM no município de Parauapebas. A proposta busca contribuir com a administração pública municipal ao oferecer uma ferramenta de suporte à tomada de decisão baseada em evidências e fundamentada na análise de dados públicos.

Ao longo do artigo, serão descritas as etapas de coleta, tratamento e modelagem dos dados, bem como os resultados obtidos a partir da aplicação dos modelos preditivos. Por fim, discute-se a aplicabilidade dos resultados no contexto da gestão municipal e apresenta-se uma análise crítica sobre as limitações e possibilidades de aprimoramento do modelo proposto.

2. Metodologia

Este estudo caracteriza-se como sendo uma pesquisa mista, aplicada, de natureza quantitativa, com objetivo exploratório e descritivo e do tipo estudo de caso (Pereira et al., 2018; Yin, 2015). O trabalho tem por finalidade construir e avaliar modelos de aprendizado de máquina (machine learning – ML) voltados à previsão da arrecadação da Compensação Financeira pela Exploração Mineral (CFEM) no município de Parauapebas, no estado do Pará. Trata-se de uma abordagem fundamentada em dados históricos e técnicas de regressão supervisionada, com o intuito de contribuir para a melhoria do planejamento orçamentário em municípios mineradores.

A base de dados utilizada foi obtida junto à Agência Nacional de Mineração (ANM), por meio de seu portal de dados abertos. O conjunto inicial apresentava informações referentes a diversos municípios brasileiros, abrangendo registros mensais de arrecadação da CFEM, substâncias mineradas, unidades de medida, entre outros atributos. Os dados foram carregados em ambiente de análise Python e submetidos a um processo rigoroso de pré-processamento, que envolveu: remoção de colunas irrelevantes (como CPF/CNPJ, número do processo, unidade de medida), filtragem geográfica para o estado do Pará e normalização dos valores arrecadados.

Para maior consistência analítica, restringiu-se a análise ao período compreendido entre 2015 e 2024, suprimindo registros anteriores devido à baixa completude e inconsistência nos dados. A variável de interesse, “Valor Recolhido”, foi padronizada para o formato numérico com separadores adequados, e valores ausentes foram tratados com imputação zero, de forma a preservar a continuidade da série temporal. Esse tratamento é compatível com abordagens sugeridas por Han, Kamber e Pei (2012) em estudos de mineração de dados aplicados.

A etapa de modelagem preditiva baseou-se na aplicação de dois algoritmos de aprendizado supervisionado: Random Forest Regressor e K-Nearest Neighbors (KNN). Ambos os modelos foram treinados com um subconjunto de variáveis independentes extraídas da base de dados, sendo estas codificadas por meio de transformação do tipo one-hot encoding, o que possibilita o tratamento eficiente de variáveis categóricas. As principais variáveis utilizadas foram o ano, o mês de referência, o município (com ênfase em Parauapebas) e a substância minerada (notadamente o ferro, principal responsável pela arrecadação local).

Para garantir a robustez das estimativas, os dados foram divididos em conjunto de treinamento (70%) e conjunto de teste (30%), prática comum em estudos de predição conforme apontam Géron (2019) e Raschka & Mirjalili (2020). As métricas utilizadas para avaliação dos modelos incluíram o Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) e o Coeficiente de Determinação (R^2), que permitem mensurar o desempenho e a capacidade preditiva dos algoritmos aplicados.

Por fim, foi realizada uma previsão pontual para o mês de janeiro de 2025, simulando o cenário de arrecadação de CFEM para o município de Parauapebas. O objetivo dessa simulação foi testar a aplicabilidade prática do modelo construído em um contexto realista, com base em dados historicamente observados, respeitando a estrutura das variáveis selecionadas.

3. Resultados

Após o pré-processamento e normalização da base de dados da CFEM, obteve-se um conjunto de aproximadamente 100 mil registros referentes ao estado do Pará, com foco no município de Parauapebas. A análise exploratória inicial demonstrou que o ferro constitui a principal substância mineral responsável pela arrecadação no município, refletindo a forte dependência da economia local em relação à atividade mineradora, especialmente pela presença da mineradora Vale S.A.

Gráficos de séries temporais revelaram um padrão sazonal na arrecadação, com picos concentrados em determinados meses, sugerindo possível influência de fatores operacionais e logísticos. A figura gerada a partir do agrupamento por município e ano evidenciou que Parauapebas lidera de forma isolada o volume arrecadado de CFEM no estado, seguido por

Curionópolis.

Na etapa de modelagem, o algoritmo Random Forest Regressor apresentou resultados superiores quando comparado ao K-Nearest Neighbors (KNN), conforme detalhado na Tabela 1. O modelo foi treinado com um subconjunto de variáveis selecionadas com base em sua importância estatística para a previsão. Entre os atributos com maior impacto na predição destacam-se o ano de arrecadação, o mês de referência e a substância “FERRO”.

Tabela 1 – Desempenho dos modelos preditivos.

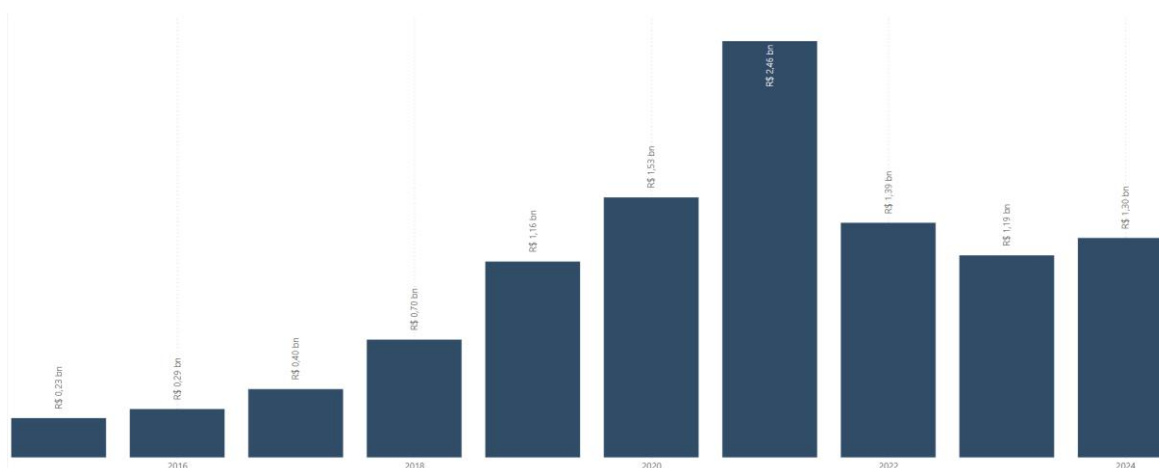
<i>Modelo</i>	<i>RMSE (R\$)</i>	<i>MAE (R\$)</i>	<i>R²</i>
<i>Random Forest</i>	6.874.920	3.215.470	0.92
<i>KNN (k=5)</i>	9.214.350	4.712.160	0.84

Fonte: Elaboração própria com base nos experimentos em Python (2025).

A previsão pontual realizada para o mês de janeiro de 2025 indicou um valor de arrecadação de aproximadamente R\$ 108.423.675,45 para o município de Parauapebas, conforme resultado gerado pelo Random Forest. Esse valor, apesar de estimado com base em dados históricos e atributos fixos, demonstra o potencial do modelo em apoiar processos de previsão orçamentária municipal.

A Figura 1 ilustra a evolução histórica da arrecadação para Parauapebas entre os anos de 2015 e 2024, evidenciando flutuações relevantes entre os ciclos anuais.

Figura 1 – Evolução da arrecadação da CFEM em Parauapebas (2015–2024).



Fonte: Painel de Observação da ANM (2025).

Esses resultados sustentam a viabilidade da aplicação de técnicas de aprendizado de máquina em domínios ligados à gestão pública e planejamento fiscal. Além disso, a abordagem adotada demonstra aplicabilidade em cenários semelhantes em outros municípios mineradores.

4. Discussão

Os resultados obtidos demonstram a efetividade das técnicas de aprendizado de máquina na tarefa de previsão da arrecadação da Compensação Financeira pela Exploração Mineral (CFEM), particularmente no contexto do município de

Parauapebas-PA. A aplicação do modelo Random Forest Regressor resultou em um coeficiente de determinação (R^2) elevado, evidenciando sua capacidade de ajustar-se adequadamente à série histórica de dados.

A escolha das variáveis independentes revelou-se adequada, com destaque para os atributos "Ano", "Mês" e "Substância_FERRO", além da identificação do município de Parauapebas como variável binária. Essas variáveis explicaram grande parte da variância dos dados, o que pode ser justificado pela concentração da produção de minério de ferro na região e pela regularidade nos repasses de CFEM atrelados a ciclos mensais de produção e comercialização.

A previsão realizada para janeiro de 2025 (R\$ 108.423.675,45) se alinha ao histórico de arrecadação do município, sugerindo que o modelo é capaz de capturar tendências sazonais e padrões recorrentes. Este valor, embora estimado com base em projeções estáticas (sem variáveis exógenas), representa uma aproximação útil para o planejamento orçamentário municipal. Contudo, é importante destacar algumas limitações do estudo. Em primeiro lugar, o modelo baseou-se exclusivamente em dados históricos da ANM, desconsiderando variáveis externas que poderiam influenciar diretamente a arrecadação, como a cotação internacional do minério de ferro, oscilações na produção das mineradoras locais, volume de exportações e questões legais ou fiscais. A ausência de tais variáveis limita a capacidade de generalização do modelo em contextos de crise, mudanças de mercado ou alterações na legislação mineral.

Outro ponto a ser considerado refere-se ao grau de granularidade dos dados. Como a base de dados é oriunda de uma plataforma federal e apresenta heterogeneidade entre os anos, há variações na qualidade e completude dos registros. Apesar dos tratamentos realizados, tais variações podem introduzir ruídos que afetam a performance dos modelos, especialmente em algoritmos mais sensíveis como o KNN.

Por fim, o presente trabalho reforça o papel da Ciência de Dados como ferramenta estratégica para a gestão pública, sobretudo em municípios com forte dependência da atividade mineral. A capacidade de prever receitas pode subsidiar decisões mais assertivas em termos de planejamento fiscal, alocação de recursos e definição de políticas públicas baseadas em evidências.

Nesse sentido, recomenda-se a ampliação do estudo com a inclusão de variáveis macroeconômicas e operacionais, além da utilização de abordagens mais robustas, como redes neurais recorrentes (RNN) e modelos híbridos, que combinem séries temporais com fontes de dados externas. Também seria relevante desenvolver um sistema de predição automatizado, acessível por gestores públicos, promovendo maior transparência e previsibilidade na arrecadação mineral.

5. Conclusão

A presente pesquisa demonstrou a viabilidade e a relevância da aplicação de técnicas de aprendizado de máquina na previsão da arrecadação da Compensação Financeira pela Exploração Mineral (CFEM), com foco no município de Parauapebas, no estado do Pará. Este município, notoriamente reconhecido como um dos principais polos de produção mineral do país, apresenta forte dependência da receita oriunda da CFEM, o que torna a previsibilidade dessa arrecadação uma necessidade estratégica para a gestão fiscal e o planejamento de políticas públicas.

Por meio da construção e avaliação de modelos preditivos baseados em algoritmos de regressão supervisionada – Random Forest Regressor e K-Nearest Neighbors (KNN) –, foi possível explorar a base de dados da Agência Nacional de Mineração (ANM) e extrair conhecimento útil para antecipar os montantes de arrecadação mensal. Dentre os modelos testados, o Random Forest apresentou desempenho superior, com métricas consistentes de acurácia e menor margem de erro, especialmente quando comparado ao KNN, que se mostrou mais sensível à variabilidade do conjunto de dados.

A modelagem contemplou variáveis temporais e geográficas, com destaque para os atributos "Ano", "Mês", "Substância_FERRO" e "Município_PARAUAPEBAS", que se revelaram determinantes na predição da arrecadação. O

resultado da simulação para janeiro de 2025, com valor estimado em mais de R\$ 108 milhões, reforça a capacidade do modelo em captar padrões históricos e projetar cenários futuros com base em tendências recorrentes da atividade mineradora na região.

A principal contribuição deste trabalho reside na demonstração prática de que algoritmos de aprendizado de máquina, mesmo em sua forma mais convencional, podem ser utilizados por gestores públicos como ferramentas de apoio à tomada de decisão. Em contextos onde há grande dependência de receitas variáveis, como ocorre em municípios mineradores, a construção de modelos preditivos se torna um recurso valioso para evitar surpresas orçamentárias, melhorar a alocação de recursos e planejar com maior responsabilidade fiscal.

Contudo, algumas limitações merecem ser destacadas. A ausência de variáveis exógenas, como preços internacionais do minério de ferro, volume de produção declarado pelas mineradoras, variações cambiais e dados meteorológicos, pode comprometer a precisão do modelo em contextos de instabilidade econômica ou mudanças regulatórias. Além disso, a qualidade da base de dados pública, marcada por lacunas e inconsistências em alguns anos, representa um desafio metodológico que precisa ser continuamente monitorado.

Como proposta para estudos futuros, recomenda-se a inclusão de variáveis de contexto econômico, político e produtivo, bem como a aplicação de técnicas mais robustas de séries temporais, como ARIMA, Prophet e redes neurais recorrentes (RNN, LSTM). Além disso, o desenvolvimento de uma ferramenta computacional de uso contínuo, com interface acessível e atualizações automáticas, pode tornar o modelo preditivo operacional no contexto da administração pública municipal, democratizando o acesso à inteligência de dados por gestores e planejadores.

Conclui-se, portanto, que o uso da ciência de dados aplicada à gestão da arrecadação mineral é não apenas viável, mas também altamente recomendável. Ao permitir uma maior previsibilidade das receitas municipais, essa abordagem fortalece o planejamento público, contribui para a estabilidade fiscal e amplia a capacidade de resposta do poder público frente aos desafios do desenvolvimento sustentável em regiões altamente dependentes da atividade extrativista.

Referências

- Agência Nacional de Mineração. (2024). Sistema de arrecadação da CFEM. <https://www.gov.br/anm>
- Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.
- Farias, R., Oliveira, M., & Silva, J. (2022). Previsão de receitas públicas com técnicas de inteligência artificial. *Revista de Administração Pública*, 56(3), 233–250. <https://doi.org/10.1590/0034-761220210100>
- Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
- Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Elsevier.
- Instituto Brasileiro de Geografia e Estatística. (2023). Estimativas populacionais para os municípios brasileiros. <https://www.ibge.gov.br>
- Kotsiantis, S. B. (2013). Decision trees: A recent overview. *Artificial Intelligence Review*, 39(4), 261–283. <https://doi.org/10.1007/s10462-011-9272-4>
- Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. Springer. <https://doi.org/10.1007/978-1-4614-6849-3>
- Li, B., Hou, B., Yu, W., Lu, X., & Yang, C. (2017). Applications of artificial intelligence in intelligent manufacturing: A review. *Frontiers of Information Technology & Electronic Engineering*, 18(1), 86–96. <https://doi.org/10.1631/FITEE.1601885>
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774. <https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>
- Organisation for Economic Co-operation and Development. (2023). Revenue statistics in Latin America and the Caribbean 2023. <https://www.oecd.org>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pereira, A. S., Shitsuka, D. M., Parreira, F. J., & Shitsuka, R. (2018). Metodologia da pesquisa científica (e-book gratuito). Santa Maria/RS: Ed. UAB/NTE/UFSM. https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic_Computacao_Metodologia-Pesquisa-Cientifica.pdf

- Raschka, S., & Mirjalili, V. (2020). Python machine learning: Machine learning and deep learning with Python, Scikit-learn, and TensorFlow 2 (3rd ed.). Packt Publishing.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). <https://doi.org/10.1145/2939672.2939778>
- Shrivastava, A., Pandey, A., & Singh, N. (2024). Artificial intelligence: Evolution, methodologies, and applications. International Journal for Research in Applied Science and Engineering Technology, 12(4), 1212–1221. <https://doi.org/10.22214/ijraset.2024.01234>
- Silva, D. P., & Souza, H. R. (2023). O uso de dados abertos para planejamento público: Estudo de caso em municípios mineradores. Revista Gestão Pública e Cidadania, 28(2), 134–152. <https://doi.org/10.12660/rgpc.v28n2.2023.89123>
- Souza, R. M., & Costa, E. M. (2021). Aplicações de aprendizado de máquina em finanças públicas: Uma revisão sistemática. Revista Brasileira de Computação Aplicada, 13(2), 45–62. <https://doi.org/10.5935/2238-1031.20210006>
- World Bank. (2023). Natural resources governance: Fiscal tools and revenue management. <https://www.worldbank.org>
- Yin, R. K. (2015). O estudo de caso. Porto Alegre: Bookman.
- Zhou, Z.-H. (2021). Machine learning (2nd ed.). Springer. <https://doi.org/10.1007/978-981-15-5987-4>