

Predição da produção de minério bruto de cobre por meio de técnicas de *Machine Learning* com aprendizado supervisionado

Prediction of copper ore raw material production through Machine Learning techniques with supervised learning.

Eduarda Carvalho de Almeida

Universidade Federal Fluminense, Programa de Pós-Graduação Dinâmica dos Oceanos e da Terra, Av. Gen. Milton Tavares de Souza s/nº – Gragoatá, Campus da Praia Vermelha, CEP: 24210-346, Niterói – RJ
(eduardaallmeida@hotmail.com)

Resumo

O Brasil possui um dos maiores potenciais de reservas minerais do mundo, tornando o setor mineral fundamental para a economia nacional, contribuindo para a geração de empregos e a arrecadação de impostos e royalties. Nesse cenário, o cobre é um dos elementos protagonistas dentre as substâncias metálicas exploradas, se caracterizando por ser um minério primordial para diversas indústrias, como a de eletrônicos, de construção civil, de energia e transporte. Sendo assim, este artigo se propõe a aplicar técnicas de *Machine Learning (ML)* para modelar o comportamento da produção de cobre e efetuar predições com base nos dados históricos disponíveis, tendo em vista que a predição da produção de minério bruto de cobre nos próximos anos pode ser estratégica para o planejamento do setor. O dataset para treinamento dos modelos supervisionado exibe a produção anual dos anos de 2001 a 2021 com base nos Anuários Minerais Brasileiros. Foram utilizadas técnicas de aprendizagem supervisionada através dos algoritmos de regressão: Árvores de Decisão e Regressão Polinomial, com auxílio da biblioteca *Scikit-Learn*. Ambos os modelos apresentaram desempenho relativamente bom nos conjuntos de treino e teste, com base na métrica *sMAPE* o modelo treinado com árvores de regressão com o *criterion = squared_error* e *max_depth=5* apresentou com menor erro, possuindo o melhor desempenho. Com base nesse modelo espera-se que a produção de cobre no ano de 2022 no Brasil totalize 94.830.000,00 de toneladas

Palavras-chave: Machine Learning; Predição; Regressão; Produção.

Abstract

Brazil has one of the greatest potential for mineral reserves in the world, making the mineral sector fundamental to the national economy, contributing to the creation of jobs and the collection of taxes and royalties. In this scenario, copper is one of the main elements among the exploited metallic substances, characterized by being a fundamental ore for several industries, such as electronics, civil construction, energy and transport. Therefore, this article proposes to apply Machine Learning (ML) techniques to model the behavior of copper production and make predictions based on available historical data, considering that the prediction of crude copper ore production in the coming years can be strategic for the strategic planning of the sector. The dataset for supervised model training displays the annual production from the years 2001 to 2021 based on the Brazilian Mineral Yearbooks. Supervised learning techniques were used through regression algorithms Decision Trees and Polynomial Regression, with the aid of the Scikit-Learn library. Both models performed relatively well in the training and test sets, based on the *sMAPE* metric, the model trained with regression trees with *criterion = squared_error* and *max_depth=5* presented the lowest error, having the best performance. Based on this model, copper production in 2022 in Brazil is expected to total 94,830,000.00 tons.

Keywords: Machine Learning; Prediction; Regression; Production.

1. Introdução

Com uma vasta riqueza de recursos minerais, o Brasil possui um dos maiores potenciais de reservas minerais do mundo, o que torna a atividade de mineração de grande importância para o país (IBRAM, 2019). Desta forma, contribuindo para a geração de empregos e a arrecadação de impostos e royalties (ANM, 2022).

Segundo Santos e Motta (2018), a indústria mineral brasileira tem uma história rica em substâncias metálicas, cuja importância remonta aos tempos da Colônia. As incursões dos bandeirantes em busca de metais preciosos, como o cobre, definiram novas rotas para a ocupação do interior do Brasil e culminaram com a exploração de jazidas na região das Minas Gerais. No entanto, outros metais também tiveram relevância para a indústria mineral brasileira.

De acordo com dados estatísticos recentes, as substâncias metálicas continuam a ser extremamente relevantes para a produção mineral brasileira. Em 2021, elas responderam por cerca de 89% do valor total da produção no país. Dentre essas substâncias, onze se destacam por representarem 99,7% do valor da produção, incluindo alumínio, cobre, cromo, estanho, ferro, manganês, nióbio, níquel, ouro, vanádio e zinco. A produção dessas onze substâncias alcançou um valor total de 312,9 bilhões de reais, com protagonismo do ferro, ouro e cobre, que contribuíram com 80,1 %, 8,65% e 5,83% respectivamente (ANM, 2022) (figura 01).

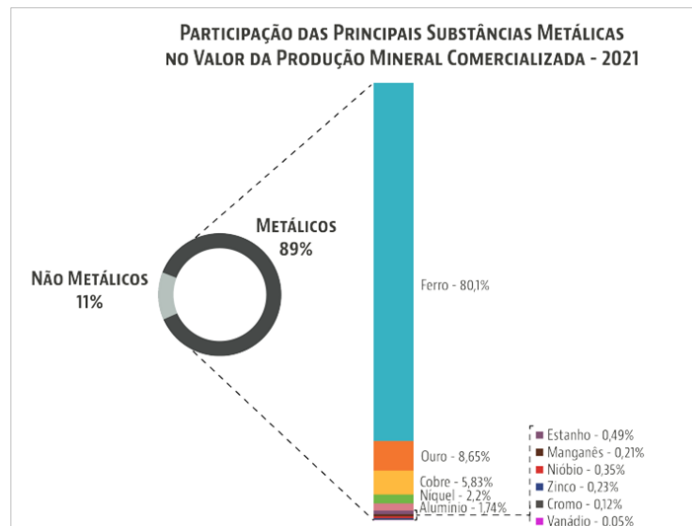


Figura 01 - Gráfico com a participação das principais substâncias metálicas no valor da produção comercializada com base no ano de 2021. Fonte: (ANM, 2022).

O cobre é um mineral fundamental para diversas indústrias, como a de eletrônicos, construção civil, energia e transporte, alcançando um valor de produção de R\$ 29,6 bilhões em 2021.

Dada a importância que esse metal representa para a economia mineral do país, a projeção da produção de minério bruto de cobre nos próximos anos pode ser estratégica para o planejamento estratégico do setor. Nesse sentido, o presente artigo visa aplicar técnicas de *Machine Learning* (ML), um ramo da Inteligência Artificial, para modelar o comportamento da produção de cobre e efetuar previsões com base nos dados históricos disponíveis.

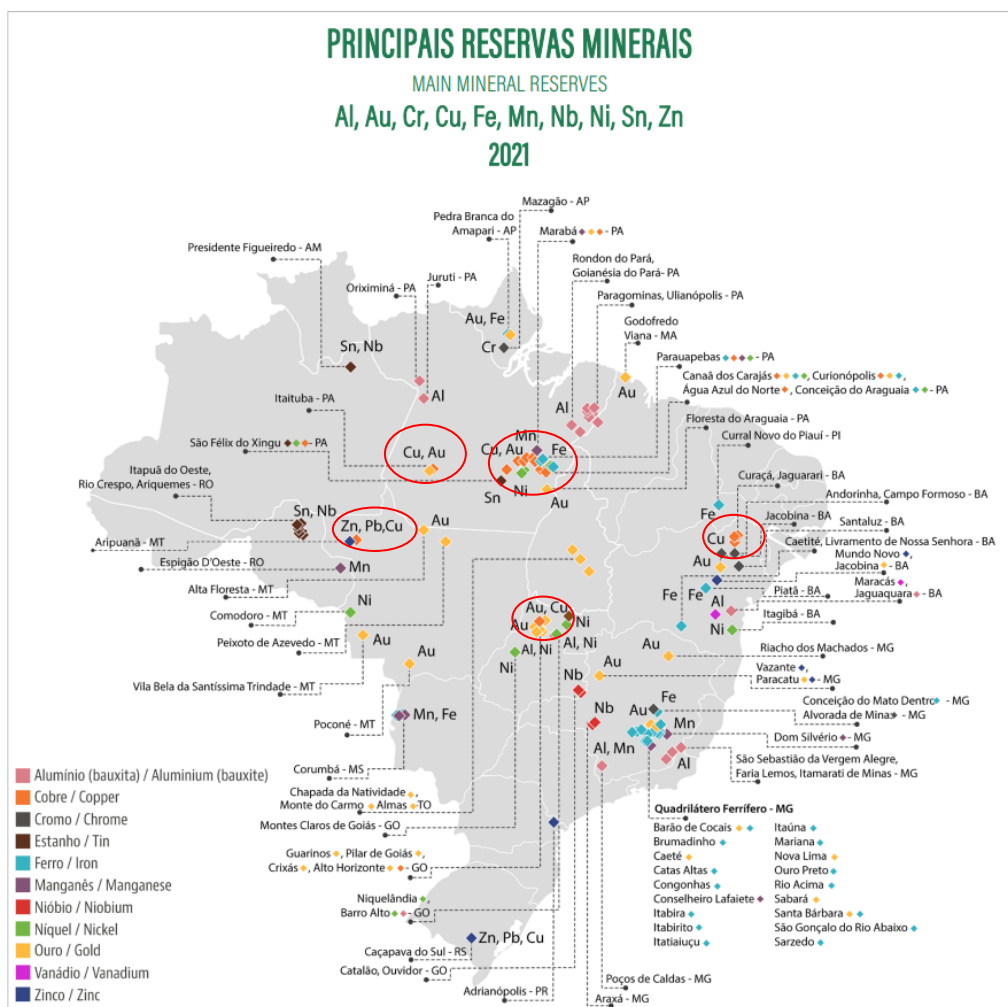


Figura 02: Localização das principais reservas minerais brasileiras de alumínio, cobre, cromo, estanho, ferro, manganês, nióbio, níquel, ouro, vanádio e zinco. Em destaque as reservas de cobre sinalizadas por círculos vermelhos. Fonte: (ANM, 2022).

2. O papel estratégico do Cobre na economia brasileira

O cobre é um elemento metálico com características excepcionais, sendo maleável, dútil, resistente à corrosão, antimicrobiano e excelente condutor de calor e eletricidade. Este elemento ocorre naturalmente na crosta terrestre em diversas formas, incluindo depósitos de sulfeto (como calcopirita, bornita, calcocita e covelita), depósitos de carbonato (como azurita e malaquita),

depósitos de silicato (como crisocola e diopside), além de ser encontrado em sua forma pura como "cobre nativo" (ICSG, 2023).

O potencial mineral do cobre no Brasil, se concentra principalmente em domínios do Pré-Cambrianos e o país consome cerca de 3% da produção mundial de cobre concentrado. A região de Carajás é a principal província mineral, com a maior quantidade de cobre metálico no país com um total estimado de 27,34 Mt de metal contido. (Costa *et al.*, 2023).

Enquanto a maioria dos países explora principalmente depósitos porfíricos, o Brasil possui cerca de 30 depósitos de cobre e perspectivas avançadas, distribuídos em diferentes tipos de ambiente geológicos, sendo os mais comuns os de segregação magmática IOCG (36%), IOCG (26%) e VMS (13%) (Silva *et al.*, 2023). Esses tipos de depósitos respondem por 74% do total de depósitos desse elemento no país, no ano de 2021 algumas empresas produtoras se destacaram pela sua participação anual nacional com a Salobo Metais e Vale (PA) (figura 03).

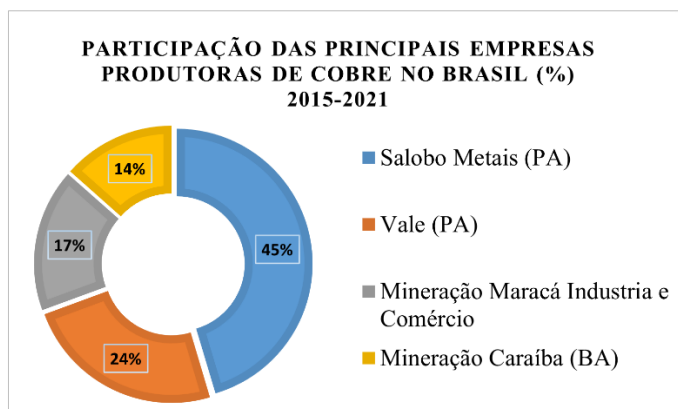


Figura 3 – Principais empresas produtoras de cobre no Brasil. Fonte:(ANM, 2022).

Embora o cobre seja um elemento abundante, nem sempre é encontrado em teores que permitam uma mineração lucrativa. Segundo o USGS, existem cerca de 2,1 bilhões de toneladas de recursos de cobre em todo o mundo. Entretanto, a produção global e as reservas são dominadas pelos países sul-americanos. O Chile é o maior produtor de cobre, responsável por cerca de 30% da produção global em 2020, com 5,7 milhões de toneladas. Em segundo lugar está o Peru, com 2,2 milhões de toneladas, seguido pela China, com 1,7 milhão de toneladas. Já os Estados Unidos produziram cerca de 1,2 milhão de toneladas, um pouco menos do que as 1,3 milhão de toneladas produzidas pela República Democrática do Congo (ICSG, 2023).

Em termos de Produção Bruta (ROM), isto é, a quantidade de minério bruto produzido obtido diretamente da mina sem sofrer qualquer tipo de beneficiamento, com base nos dados do

Anuário Mineral Brasileiro de 2022, em 2021 o valor em toneladas medida fechou em 99,57 milhões, com um crescimento de aproximadamente de 10 % em relação ao ano de 2020 (figura 04).

Percebe-se também, que a tendência do mercado do cobre a partir da produção comercializada é crescer por conta dos estudos de novas jazidas e ocorrências em todo território nacional.

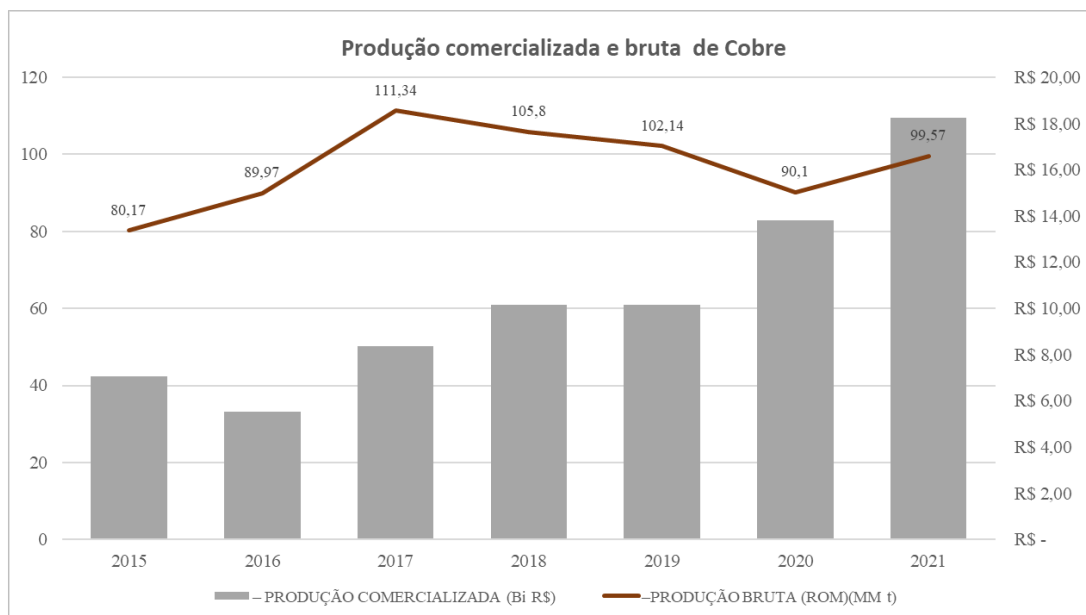


Figura 04 – Produção comercializada e bruta de cobre entre 2015 à 2021. Fonte:(ANM, 2022).

O cobre é uma commodity cujos preços são influenciados por diversas variáveis econômicas globais, devido à sua importância em uma vasta gama de setores industriais. Nos últimos cinco anos, o preço por tonelada variou consideravelmente, atingindo o patamar mais baixo em 2016, quando foi cotado a \$ 4.350 a tonelada. Após um período de crescimento constante, o preço caiu drasticamente em março de 2020, no início da pandemia de COVID-19. Desde então, o preço do cobre tem oscilado significativamente, tendo alcançado valores recordes no início de 2021, ultrapassando a marca dos US\$ 10.000 por tonelada (ICSG, 2023).

Na figura abaixo, é possível observar os valores das importações e exportações ao longo dos anos de 2015-2021, com um aumento significativo em cada ano. Esses valores são baseados no total geral, ou seja, no somatório por classificação do produto por fator agregado de países como Chile, Rússia, China, EUA, Coreia do Sul, Japão, Argentina, entre outros.

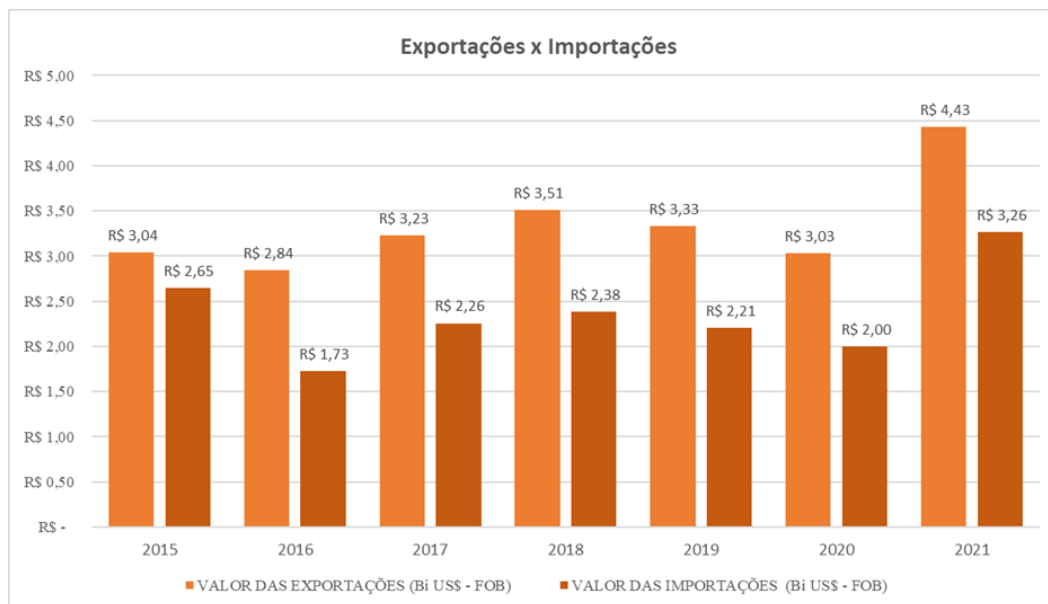


Figura 05 – Valor de exportações e importações de cobre entre 2015 a 2021. Fonte:(ANM, 2022)

Com a crescente demanda por tecnologias renováveis, o cobre tem se destacado como o metal mais transversal nesse setor. Segundo Hund *et al.* (2020), cerca de três quartos da demanda futura do metal estarão relacionados à energia solar fotovoltaica e à geração eólica. Além disso, o cobre também é fundamental em outras aplicações, como em baterias, na fiação de veículos elétricos, transmissão de rede e usinas geotérmicas. Na energia fotovoltaica, a tecnologia de seleneto de cobre-índio-gálio (CIGS) será responsável por cerca de um quinto dos painéis até 2050, enquanto as bobinas de cobre são componentes essenciais para geradores de indução usados em turbinas eólicas.

Em síntese, o cobre é uma commodity chave na economia mineral brasileira e possui uma importância crucial para a transição energética em curso e para a construção de um futuro sustentável, sendo metal estratégico para a implementação de tecnologias limpas, renováveis e de baixo carbono.

3. Materiais e métodos

Com o objetivo de treinar uma máquina sobre um conjunto de dados e, posteriormente, prever padrões de comportamento é necessário escolher o tipo de aprendizado que o será adotado (supervisionado, não supervisionado ou por esforço) e o algoritmo a ser utilizado, isto é, o conjunto de processos computacionais que geram e empregam instruções/funções a serem adotadas para a produção de um modelo que represente o fenômeno sob estudo. No âmbito do aprendizado

supervisionado fornecemos ao algoritmo um conjunto de exemplos rotulados (com identificações de classes em relação a um fenômeno ou característica), e esses exemplos são utilizados para tornar explícito (‘ensinar’) ao modelo a produzir a saída desejada. Esse conjunto de dados é utilizado para treinar o algoritmo para dois tipos principais de tarefas: a Classificação, que utiliza o algoritmo para rotular os dados em classes predefinidas e a Regressão, que busca compreender a relação funcional entre duas variáveis e possibilitar valores futuros com base nas informações de entrada utilizada para treinar o algoritmo.

Para modelar o comportamento da produção de cobre este artigo se propõe a aplicar técnicas de aprendizagem supervisionada aplicando os algoritmos de regressão Árvores de Decisão e Regressão Polinomial tendo em vista que a distribuição de produção ao longo dos anos não apresenta um comportamento linear (figura 06). O dataset para treinamento dos modelos supervisionado (quadro 1) exibe a produção anual dos anos de 2001 a 2021 com base nos Anuários Minerais Brasileiros publicado anualmente pela Agência Nacional de Mineração.

Produção Anual de Cobre (MM t)																				
2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
4,06	4,13	4,48	11,2	18,2	20,9	31,6	38,8	37,7	45,5	44,3	49,8	55,3	59,5	80,2	90	111	106	102	90,1	99,6

Quadro 1 - Produção de cobre ROM no Brasil entre 2001 à 2021.

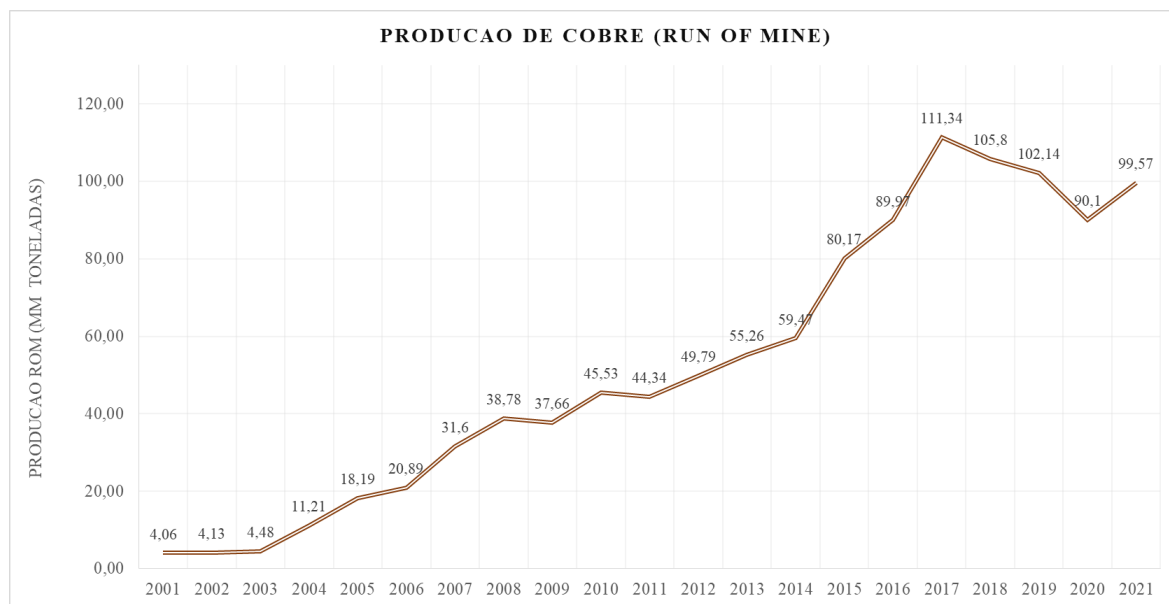


Figura 06 – Produção de cobre ROM no Brasil entre 2001 à 2021.

Para modelar o comportamento da produção de cobre ao longo do tempo que possibilite efetuar previsões foi aplicada técnicas de regressão, ou seja, utilizar análise estatísticas para entender a relação entre uma variável dependente contínua Y e uma ou mais variáveis independente

X_1, X_2, \dots, X_n . A Regressão Polinomial é um caso específico de regressão linear utilizada para efetuar um melhor ajuste na superfície curvilínea formada pelo comportamento dessas variáveis, na qual a função que descreve esse comportamento é uma função linear com a adição de N potências.

As Árvores de Decisão são algoritmos versáteis de aprendizado de máquina, especialmente por exigirem pouca preparação nos dados de treinamento do modelo, uma vez esse algoritmo não requer o escalonamento dos dados. Nesse algoritmo uma estrutura em forma de árvore é usada para modelar decisões e seus possíveis resultados, onde cada nó dessa árvore representa uma decisão baseada em uma determinada variável do conjunto de dados, e a partir de cada nó essa árvore se ramifica em mais direções, aumentando a sua profundidade (Hastie *et al.*, 2009). Em problema de regressão, o algoritmo utiliza critérios numéricos para dividir um ponto em dois nós com base em no *criterion* que é um parâmetro utilizado para determinar a medida utilizada para avaliar a qualidade da divisão dos dados em cada nó da árvore. Para treinamento de modelos de árvores de regressão a biblioteca *Scikit-Learn* é disponibiliza como *criterion* as métricas *squared_error*, *friedman_mse*, *absolute_error*, *poisson*, e o *mean absolute error*.

A aplicação desses algoritmos ao longo desse trabalho foi facilitada através da utilização da biblioteca *Scikit-Learn*, uma das bibliotecas mais populares para aprendizado de máquina em Python, que permitiu acessar os algoritmos *PolynomialFeatures* e *DecisionTreeRegressor*. Para medir a performance da regressão e avaliar o quão bem os modelos se ajustaram aos dados de treino e teste foi utilizada a métrica o *Symmetric Mean Absolute Percentage Error (sMAPE)*, que é capaz de calcular quão errado foram as previsões em relação ao resultado que era esperado levando em consideração a ordem de grandeza dos valores, dada pela equação:

$$SMAPE = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

onde F_t é o valor previsto e A_t é o valor real

4. Resultados e discussões

Para treinamento dos modelos a base de dados disposta no quadro 1 foi dividida em treino (70%) e teste (30%) através da função *train_test_split*. Com utilização da função

DecisionTreeRegressor foi possível efetuar o primeiro treinamento do modelo com as configurações padrões do algoritmo (figura 07), com o *criterion = squared_error* que apresentou um sobreajuste nos conjuntos de treinamento com o *sMAPE = 0.0*.

```
from sklearn.tree import DecisionTreeRegressor

# Treinando o modelo
regressor_dt = DecisionTreeRegressor()
regressor_dt.fit(X_train, y_train)
y_pred_dt = regressor_dt.predict(X_test)

# Avaliando o conjunto de treino
y_train_pred_dt = regressor_dt.predict(X_train)
smape_train_dt = smape(y_train, y_train_pred_dt) # Calculando o R² no conjunto de treinamento
print('sMAPE (Treinamento): %.3f' % smape_train_dt)

# Avaliando o conjunto de teste
smape_test_dt = smape(y_test, y_pred_dt) # Calculando o R² no conjunto de treinamento
print('sMAPE (Teste): %.3f' % smape_test_dt)

sMAPE (Treinamento): 0.000
sMAPE (Teste): 8.988
```

Figura 07 – Arvore de regressão sem ajuste nos hiperparâmetros.

Foi efetuado o treinamento dos modelos de árvores de regressão com os outros critérios a fim de selecionar o modelo com menores valores de erro através da métrica *sMAPE*, e para definição do melhor modelo, foi plotado o gráfico com a relação do erro *sMAPE* para profundidade das árvores nos conjuntos de treino e teste, a fim de podar o crescimento das árvores quando o aumento da profundidade não proporcionar melhora na performance do modelo. Os modelos *squared error*, *friedman mse*, *absolute error*, e *mean absolute error* apresentaram o mesmo desempenho, enquanto o modelo treinado com o critério *Poisson* exibe erros maiores através da métrica *sMAPE*, nesse sentido, com base (figura 8) é possível determinar a profundidade ideal de arvores 5 a fim de minimizar sobreajustes durante o treinamento com base no critério *absolute_error* e a profundidade 5

O modelo foi novamente treinando com a profundidade 5 (*criterion = squared_error*, *max_depth=5*), que apresenta um equilíbrio entre viés e variância exibindo os valores de *sMAPE* para o conjunto de treino foi de 0.189 e para o conjunto de teste foi de 8.798.

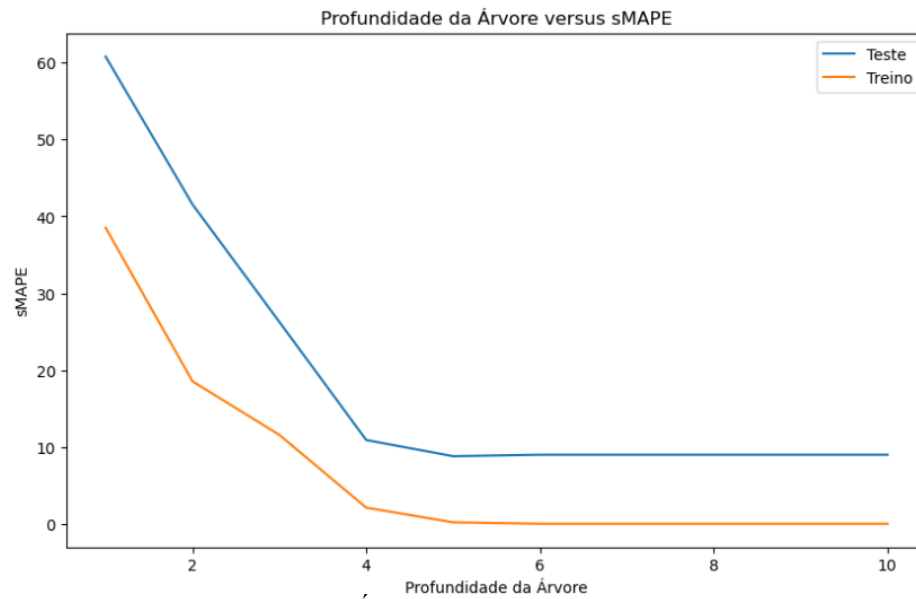


Figura 08 – Desempenho da Árvore de Regressão em Diferentes Profundidades.

Utilizando a classe *PolynomialFeatures* da biblioteca *Scikit-Learn*, foi possível realizar uma nova modelagem dos dados de treinamento, através de uma função com polinômio de grau 3, o que permitiu treinar um novo modelo (figura 09). Com a aplicação da regressão polinomial, foi calculado o desempenho desse modelo tanto no conjunto de treinamento quanto no conjunto de teste, obtendo *sMAPE* de 16.049 e 21.716, respectivamente.

```
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression

# Transformando os dados de entrada em recursos polinomiais
poly_features = PolynomialFeatures(degree=3)
X_train_poly = poly_features.fit_transform(X_train)
X_test_poly = poly_features.transform(X_test)

# Treinando o modelo de regressão linear com polinômio grau 3
regressor_poly = LinearRegression()
regressor_poly.fit(X_train_poly, y_train)
```

Figura 09 – Modelo com regressão polinomial de grau 3.

Com base nos dois modelos, árvore de regressão e regressão polinomial, utilizou-se a função *sMAPE* para determinar o modelo mais adequado para prever os valores de produção de cobre. O Quadro 02 apresenta um resumo das métricas de desempenho de ambos os modelos bem como o valor predito para o ano de 2022. É possível observar que o modelo de árvore de regressão possui um menor erro, de acordo com a função *sMAPE*.

Modelo	Árvore de regressão	Regressão polinomial
sMAPE treino	0.189	16.049
sMAPE teste	8.798	21.716
Valor predito 2022	94835	101.04

Quadro 02 – Síntese das métricas de desempenho dos modelos.

5. Conclusões

Este trabalho efetuou a aplicação prática do *Machine Learning* por meio do aprendizado supervisionado para solucionar desafios no setor mineral. Foi utilizada as técnicas de árvores de decisão e regressão polinomial, empregando os algoritmos *PolynomialFeatures* e *DecisionTreeRegressor* da biblioteca *Scikit-Learn*, a fim de realizar a predição da produção de cobre (ROM) para o próximo ano no Brasil.

Através da otimização dos hiperparâmetros foi possível encontrar a profundidade ideal da árvore de regressão e reduzir o nível de complexidade do modelo ajustando sua variância através da poda na profundidade 5. Ambos os modelos apresentaram desempenho relativamente positivos nos conjuntos de treino e teste, com base na métrica *sMAPE* o modelo treinando com árvores de regressão apresentou-se com menor erro, possuindo o melhor desempenho. Com base nesse modelo espera-se que a produção de cobre no ano de 2022 no Brasil totalize 94.830.000,00 de toneladas.

6. Referências

- Brasil. Agência Nacional de Mineração. Anuário Mineral Brasileiro: principais substâncias metálicas. Coordenação técnica de Karina Andrade Medeiros. – Brasília: ANM, 2023. p. 23.
- CORTI, C.W. e HOLLIDAY, R.J. Gold Bulletin, v. 37, p. 1-2, 2004.
- DA COSTA, Marco Aurélio; RIOS, Francisco Javier. The gold mining industry in Brazil: A historical overview. **Ore Geology Reviews**, v. 148, 2022. p. 105005,
- DNPM. Departamento Nacional de Produção Mineral. Produção Mineral Brasileira 2015. Brasília: DNPM, 2015.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The elements of statistical learning: data mining, inference, and prediction. Springer Series in Statistics. New York: **springer**. 2009. p. 767.
- HUND, K. et al. Minerals for climate action: the mineral intensity of the clean energy transition. Washington: World Bank, 2020.
- ICSG. International Copper Study Group. Copper Applications. Disponível em: <https://www.icsg.org/copper-applications>. Acesso em: 19 de abril de 2023.
- IBRAM. Instituto Brasileiro de Mineração. Mineração no Brasil. Disponível em: <http://www.ibram.org.br>. Acesso em: 20 de abril de 2023.
- SILVA, G.F.; CUNHA, I.A.; COSTA, I.S.L. (Orgs.) 2023. An overview of Critical Minerals Potential of Brazil. Brasília: **Serviço Geológico do Brasil**, 2023. p. 23

sklearn.tree.DecisionTreeRegressor. Disponível em: <<https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html>>. Acesso em: 04 de junho de 2023.

sklearn.metrics.r2_score. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html>. Acesso em: 04 de junho de 2023.

sklearn.preprocessing.PolynomialFeatures. Disponível em: <<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PolynomialFeatures.html>>. Acesso em: 04 de junho de 2023.

USGS. U.S. Geological Survey. Mineral Commodity Summaries 2022. Disponível em: <https://pubs.usgs.gov/>. Acesso em: 20 de abril de 2023.

World Gold Council. Uses of Gold. Disponível em: <https://www.gold.org/goldhub/uses-of-gold>. Acesso em: 20 de abril de 2023.