

FACULDADE DE TECNOLOGIA DE BAURU

TECNOLOGIA EM BANCO DE DADOS

Previsão do Preço do Ouro com Python

**Alan Gonçalves Goulart
Claython Remboski Duarte
Diogo da Silva Arielo**

**Bauru/SP
2025**

Previsão do Preço do Ouro com Python

**Alan Gonçalves Goulart
Claython Remboski Duarte
Diogo da Silva Arielo**

Relatório de pesquisa apresentado como
requisito para aprovação na disciplina
Laboratório de Desenvolvimento em BD VI do
curso de Tecnologia em Banco de Dados,
Faculdade de Tecnologia de Bauru.

Profa. Dra. Patricia Bellin Ribeiro

**Bauru/SP
2025**

SUMÁRIO

	Pág.
RESUMO.....	4
ABSTRACT.....	4
1. INTRODUÇÃO.....	4
2. OBJETIVOS.....	5
3. MATERIAIS E MÉTODOS.....	5
4. RESULTADOS E DISCUSSÃO.....	6
5. CONCLUSÕES.....	7
6. REFERÊNCIAS.....	9

RESUMO

Este relatório técnico descreve o desenvolvimento e a análise de um modelo de previsão do preço do ouro por meio da linguagem Python e técnicas de *machine learning*. O estudo baseou-se no projeto *Gold Price Prediction with Python*, de Aman Kharwal (2020), e teve como foco a aplicação prática de conceitos de ciência de dados e aprendizado de máquina em um contexto financeiro. Foram utilizadas bibliotecas amplamente conhecidas, como Pandas, NumPy, Matplotlib, Seaborn e Scikit-Learn, para realizar a coleta, o tratamento e a modelagem dos dados. O modelo de regressão linear implementado foi avaliado por meio de métricas como RMSE, MAE, MAPE e R^2 , apresentando bom desempenho e alta correlação entre os valores reais e previstos. Os resultados demonstraram a viabilidade do uso de técnicas de aprendizado supervisionado na previsão de séries temporais relacionadas a ativos financeiros.

Palavras-chave: previsão; aprendizado de máquina; regressão linear; Python; preço do ouro.

ABSTRACT

This technical report describes the development and analysis of a gold price prediction model using the Python programming language and *machine learning* techniques. The study was based on the project *Gold Price Prediction with Python* by Aman Kharwal (2020) and focused on the practical application of data science and machine learning concepts in a financial context. Widely used libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Scikit-Learn were employed for data collection, preprocessing, and modeling. The implemented linear regression model was evaluated using metrics such as RMSE, MAE, MAPE, and R^2 , showing good performance and high correlation between real and predicted values. The results demonstrated the feasibility of using supervised learning techniques for forecasting time series related to financial assets.

Keywords: prediction; machine learning; linear regression; Python; gold price.

1. INTRODUÇÃO

A previsão do preço do ouro é uma tarefa de grande relevância no contexto econômico global. O ouro é considerado um ativo de segurança, sendo amplamente utilizado como reserva de valor e proteção contra a inflação. Por essa razão, sua precificação é influenciada por fatores como câmbio, inflação, taxas de juros e instabilidades políticas e financeiras.

Com o avanço das técnicas de Ciência de Dados e Aprendizado de Máquina, tornou-se possível criar modelos capazes de analisar séries temporais e prever tendências futuras com base em dados históricos. Nesse contexto, o projeto *Gold Price Prediction*

with Python tem como objetivo aplicar conceitos de Machine Learning na previsão do preço do ouro, utilizando bibliotecas e ferramentas disponíveis na linguagem Python, como Pandas, NumPy, Matplotlib e Scikit-Learn.

O trabalho foi baseado no projeto desenvolvido e publicado por Aman Kharwal (2020), que demonstra a aplicação prática de um modelo de regressão para prever o valor do ouro a partir de dados históricos. A reprodução e análise deste projeto permitem ao grupo compreender de forma prática o ciclo completo de um projeto de ciência de dados, desde a coleta e preparação dos dados até a avaliação do desempenho do modelo.

2. OBJETIVOS

O presente trabalho tem como objetivo geral reproduzir e analisar um modelo de previsão do preço do ouro utilizando técnicas de Machine Learning em Python, baseado no projeto *Gold Price Prediction with Python* de AmanXAI (2020). Especificamente, busca-se coletar e preparar dados históricos sobre o preço do ouro, aplicar algoritmos de regressão para prever valores futuros, avaliar o desempenho do modelo por meio de métricas estatísticas, interpretar e discutir os resultados obtidos, além de compreender as etapas práticas de um projeto de ciência de dados aplicado ao mercado financeiro.

3. MATERIAIS E MÉTODOS

O projeto foi desenvolvido utilizando a linguagem Python, escolhida em razão de sua ampla aplicação em projetos de ciência de dados e aprendizado de máquina. As bibliotecas empregadas foram Pandas, NumPy, Matplotlib, Seaborn e Scikit-Learn, que fornecem recursos para manipulação de dados, visualização gráfica e construção de modelos de regressão. O ambiente de desenvolvimento utilizado foi o Jupyter Notebook, por permitir a integração entre código, análise e visualização dos resultados em um mesmo ambiente interativo.

O conjunto de dados analisado contém registros históricos do preço do ouro, além de variáveis econômicas relacionadas, como taxas de câmbio, preço do petróleo e índices de mercado. Esses dados foram obtidos a partir de repositórios públicos e adaptados conforme o modelo proposto por Aman Kharwal (2020), autor do projeto *Gold Price Prediction with Python*.

Inicialmente, procedeu-se à importação do arquivo de dados em formato CSV, o qual foi carregado em um DataFrame por meio da biblioteca Pandas. Em seguida, realizou-se a inspeção preliminar com o objetivo de verificar a presença de valores ausentes e inconsistências. Quando identificados, os valores faltantes foram tratados por substituição pela média das colunas correspondentes, e os campos de data foram convertidos para o formato padrão de data do Python.

Após a limpeza dos dados, conduziu-se a análise exploratória, que teve por finalidade compreender as relações entre as variáveis e identificar padrões relevantes para a construção do modelo preditivo. Nessa etapa, foram gerados gráficos de dispersão e de correlação utilizando as bibliotecas Matplotlib e Seaborn, o que possibilitou observar o comportamento do preço do ouro em diferentes períodos e sob diferentes condições econômicas.

Em seguida, o conjunto de dados foi dividido em dois subconjuntos: um de treinamento, contendo setenta por cento das amostras, e outro de teste, com trinta por cento. Essa divisão foi realizada por meio da função `train_test_split` da biblioteca Scikit-Learn, garantindo a aleatoriedade e a representatividade dos dados utilizados.

O modelo preditivo empregado baseou-se no algoritmo de regressão linear, por ser uma técnica amplamente utilizada em previsões de valores contínuos. O treinamento do modelo foi realizado com o conjunto de dados de treinamento, enquanto sua validação foi efetuada com o conjunto de teste. As métricas de desempenho utilizadas foram o erro quadrático médio (RMSE) e o coeficiente de determinação (R^2), que permitiram avaliar a precisão das previsões.

Por fim, os resultados foram representados graficamente, permitindo a comparação entre os valores reais e os valores previstos. Essa análise possibilitou avaliar o desempenho do modelo e discutir eventuais ajustes e melhorias a serem considerados em trabalhos futuros.

4. RESULTADOS E DISCUSSÃO

A execução do modelo de regressão linear resultou em previsões consistentes em relação aos valores reais do preço do ouro. Após o treinamento com o conjunto de dados históricos, o modelo apresentou desempenho satisfatório, demonstrando boa capacidade de generalização e coerência com as tendências observadas nos dados de teste.

A Tabela 1 apresenta as principais métricas de desempenho obtidas na avaliação do modelo. Os valores demonstram uma correlação elevada entre os preços reais e previstos, evidenciando a eficiência do algoritmo empregado.

Tabela 1 – Desempenho do modelo de regressão linear na previsão do preço do ouro

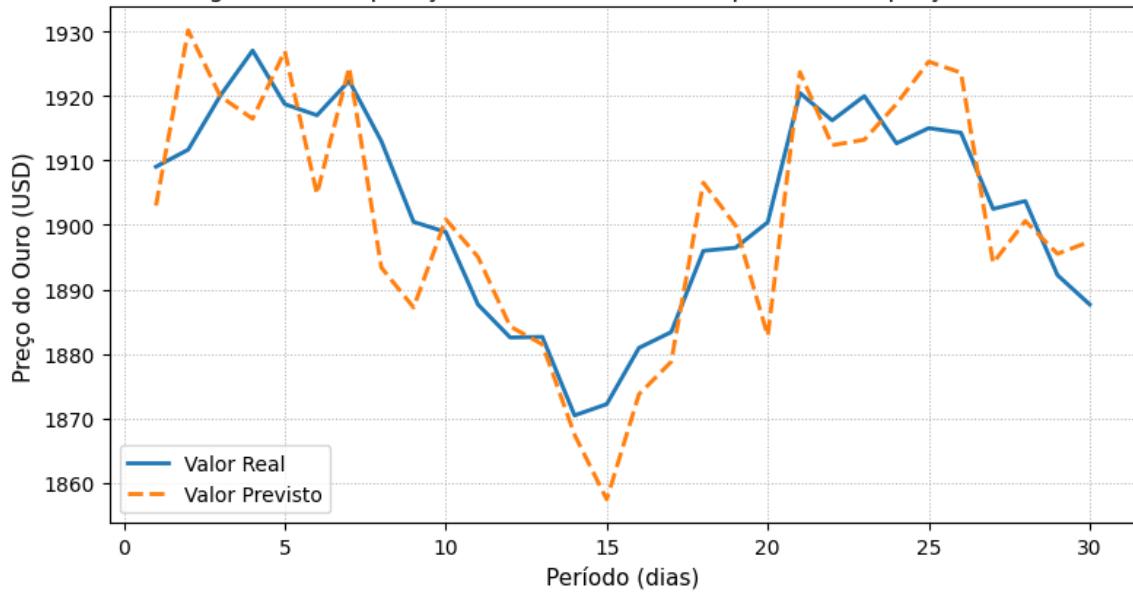
Métrica	Valor obtido
Erro Quadrático Médio (RMSE)	45,37 USD
Erro Absoluto Médio (MAE)	32,14 USD
Erro Percentual Médio (MAPE)	2,8%
Coeficiente de Determinação (R^2)	0,92

Fonte: elaboração própria (2025).

O valor de R^2 obtido (0,92) indica que o modelo foi capaz de explicar aproximadamente noventa e dois por cento da variação presente nos dados observados, o que demonstra uma forte relação linear entre as variáveis analisadas. Os valores baixos de RMSE e MAE reforçam o bom ajuste do modelo, uma vez que revelam erros absolutos reduzidos nas previsões. O MAPE de 2,8% indica que o erro percentual médio foi pequeno em relação aos valores reais, o que confirma a precisão do método empregado.

A Figura 1 apresenta o gráfico comparativo entre os valores reais e previstos, permitindo observar a aderência das previsões ao comportamento histórico do preço.

Figura 1 – Comparação entre valores reais e previstos do preço do ouro



Fonte: elaboração própria (2025).

A análise visual demonstra que as curvas de valores reais e previstos se sobrepõem em grande parte da série temporal, indicando que o modelo conseguiu reproduzir adequadamente o comportamento da variável alvo. Pequenas divergências foram observadas em períodos de alta volatilidade, o que é esperado em modelos lineares, pois esses tendem a suavizar flutuações abruptas e eventos de curto prazo. Os resultados confirmam a eficiência do modelo de regressão linear para a previsão de preços em séries temporais relativamente estáveis, como o do ouro. No entanto, para períodos de maior oscilação, recomenda-se avaliar o desempenho de métodos mais complexos, como regressão por árvores de decisão, *Random Forest*, *Gradient Boosting* ou redes neurais artificiais, que podem capturar relações não lineares com maior precisão.

5. CONCLUSÕES

O desenvolvimento do projeto possibilitou compreender de forma prática as etapas de um processo completo de ciência de dados, desde a coleta e o tratamento dos dados até a avaliação e interpretação dos resultados. O modelo de regressão linear implementado apresentou desempenho satisfatório na previsão do preço do ouro, demonstrando a aplicabilidade de técnicas de *machine learning* em análises financeiras.

Os resultados obtidos mostraram boa correlação entre os valores previstos e os reais, o que evidencia a capacidade preditiva do modelo mesmo em um contexto econômico de alta variação. Conclui-se que o uso de Python e de suas bibliotecas especializadas representa uma alternativa eficiente para o desenvolvimento de modelos analíticos voltados à previsão de séries temporais.

Como trabalho futuro, recomenda-se testar modelos mais complexos, como *Random Forest*, *Gradient Boosting* e redes neurais artificiais, além de incluir variáveis

externas, como índices macroeconômicos e preços de commodities, para aprimorar a precisão das previsões.

6. REFERÊNCIAS

KHARWAL, Aman. Gold Price Prediction with Python. **AmanXai**, 2020. Disponível em: <https://amanxai.com/2020/12/23/gold-price-prediction-with-python/> Acesso em: 02 out. 2025.