

**FINTECH  
BOOST**

**Inteligência Artificial e  
Análise de Dados.**

**Liderando as Finanças do Futuro**

## **Disclaimer**

A Responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do especialista.

**FINTECH  
BOOST**

# Machine Learning e Finanças



# Zaika dos Santos

Zaika dos Santos é cientista de dados, multi-artista, curadora, designer, engenheira de prompt, ficcionista, pesquisadora, especialista em Inteligência Artificial, divulgadora científica de Movimentos Especulativos como Afrofuturismo, Africanfuturism e Afropresentismo, criptoartista e pesquisadora NFT e Web 3.0. CEO da empresa Afrofuturismo Arte e STEM, uma empresa inovadora que combina Design Generativo, CGI, Ciência de Dados, Ciências Humanas, Ciências Exatas, STEAM e Educação 5.0 para criar soluções inovadoras e transformadoras. É realizadora, produtora e editora de ficções especulativas. Ela é coordenadora geral da Black Speculative Arts Movement Brasil e coordenadora da Black Speculative Arts Movement no Sul Global.

Tecnóloga em Audiovisual, em Rádio e TV, em Web Design. Licenciada em Artes Plásticas na Guignard – UEMG, especialista em Ciência de Dados pela Digital House, pós-graduada lato-sensu em Inteligência Artificial pela UNA.

Foi capa da Glamour Brasil em 2023, liderança negra brasileira pelo Fundo Alas – Fundação Tide Setubal, Porticus e Ibirapitanga. É cientista filiada ao 500 Women Scientists, membra da comunidade ForbesBLK, curadora da II Bienal Black Brasil Art, curadora da exposição 'Mulheres na Ciência: Olhares Sobre o Futuro' – Aliança Francesa, da 'Exposição Imersiva Afro-Remix' – BSAM Brasil e de outros artistas em exposições nacionais e internacionais. Também atua com Design Gráfico, Design Estratégico, Storytelling, possui treinamento em cenários do futuro através do 'Afrofuturism 2.0 and Global Security Foresight' utilizando a metodologia de Horizon Scanning, junto à Instituição N Square. Em 2022 apresentou o percurso em Arte, Ciência e Tecnologia do Afrofuturismo Brasileiro em Harvard, na Segunda Conferência em Estudos Afro-Latinoamericanos.



# AGENDA

01

Introdução

02

O que é Machine Learning?

03

Mergulhe nos conceitos

04

Como aplicar

05

Conclusão

06

Bibliografias de apoio

# Introdução

No curso **Machine Learning e Finanças** você terá uma introdução às técnicas de aprendizado de máquina aplicadas ao setor financeiro. Vamos explorar como modelos preditivos podem ser usados para analisar mercados, para gestão de riscos e otimizar investimentos.



## O que é Machine Learning?

Será apresentado o conceito de Machine Learning, trazendo um breve contexto histórico até os dias atuais, já fazendo uma ligação de como ele é utilizado nas finanças com alguns exemplos mais gerais e abrangentes.

## Mergulhe nos conceitos

Aqui faremos um “mergulho” nos conceitos principais, aprofundando em Aprendizado Supervisionado, Aprendizado Não Supervisionado, Aprendizado Semi-supervisionado e Aprendizado por Reforço.

## Como aplicar

Exploraremos o uso de Modelos Preditivos em técnicas para previsão de preços, análise de riscos e identificação de padrões. Também veremos a aplicação de Algoritmos, trading automatizado e otimização de estratégias financeiras e ética.

## Conclusão

Na conclusão veremos a integração de conhecimentos explorados no curso na aplicação prática de modelos desenvolvidos para resolver problemas financeiros reais. Além disso, veremos a preparação para o mercado na discussão de aplicações futuras, desafios e como utilizar as habilidades adquiridas em cenários profissionais.

## Bibliografias de apoio

Bibliografias para leitura, material de apoio para uso posterior com referencial teórico fundamental.



# O que é Machine Learning?

---

FINTECH  
BOOST

# Machine Learning (ML),

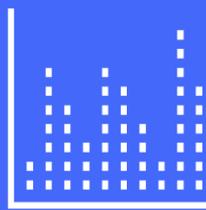
ou Aprendizado de Máquina, é um campo da Inteligência Artificial (IA) que permite que computadores aprendam e tomem decisões a partir da análise de uma base de dados, sem serem explicitamente programados, para a realização de tarefas específicas. Em outras palavras, é uma forma de ensinar a máquina a identificar padrões e fazer previsões ou tomar decisões com base nos padrões identificados em exemplos, ou experiências anteriores.

**"Aprendizado de máquina é sobre aprender com os dados para fazer previsões ou tomar decisões automaticamente."**

Pedro Domingos, autor de The Master Algorithm.

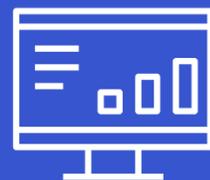
# Principais conceitos de Machine Learning:

**Curso:** Inteligência Artificial e Análise de Dados



## DADOS

A fonte do aprendizado. Incluem exemplos históricos ou simulados que alimentam o modelo para a execução do algoritmo.



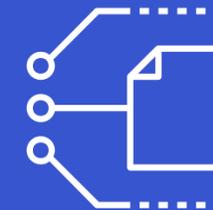
## MODELO

É a representação matemática em algoritmos aplicados que aprende com os dados.



## TREINAMENTO

O processo pelo qual o modelo ajusta seus parâmetros com base nos dados fornecidos.



## PREDIÇÃO

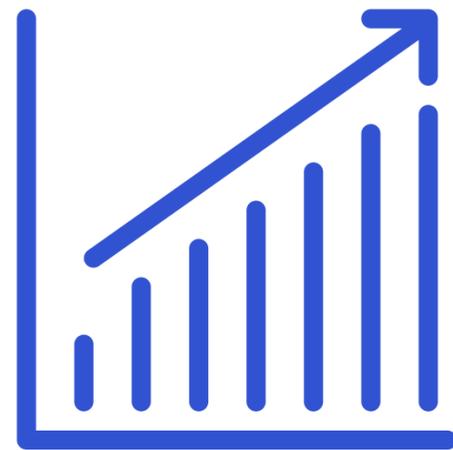
A habilidade do modelo de fazer inferências sobre novos dados após o treinamento.

"Em aprendizado de máquina, um modelo aprende padrões matemáticos a partir de dados; esses dados, por sua vez, são o combustível que alimenta o aprendizado. O processo de treinamento ajusta os parâmetros do modelo, permitindo que ele faça previsões precisas sobre novos exemplos."

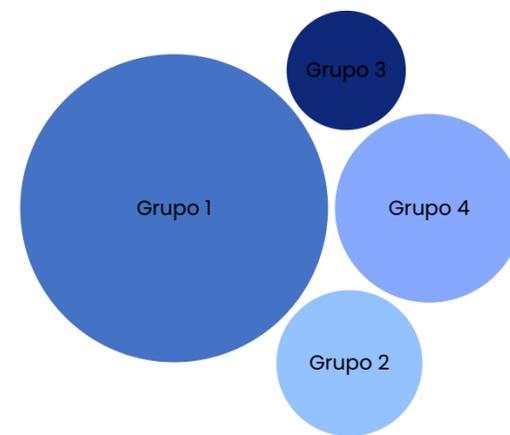
**Tom M. Mitchell, Machine Learning (1997).**

# Categorias do Machine Learning

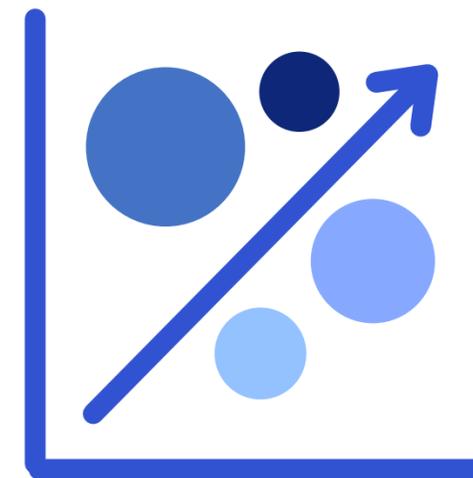
**Curso:** Inteligência Artificial e Análise de Dados



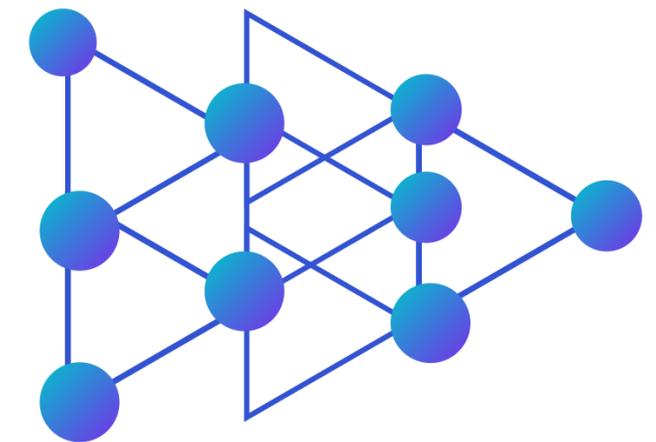
**Aprendizado Supervisionado**



**Aprendizado Não Supervisionado**



**Aprendizado Semi-supervisionado**



**Aprendizado por Reforço**



**Mergulhe nos  
conceitos**

---

# Regulamentação

SOX, ou Sarbanes-Oxley Act, é uma lei dos Estados Unidos aprovada em 2002 e tem como objetivo aumentar a transparência, a responsabilidade e a precisão das informações financeiras fornecidas por empresas públicas.

A LGPD (Lei Geral de Proteção de Dados) é uma lei brasileira que regula o uso, coleta, armazenamento e compartilhamento de dados pessoais e sensíveis. O objetivo da lei é proteger a privacidade e os direitos dos cidadãos em relação ao uso de suas informações pessoais.

A GDPR (General Data Protection Regulation) é a lei europeia de proteção de dados, em vigor desde 2018, que regula como dados pessoais devem ser coletados, armazenados e processados no continente. Garante direitos como acesso, exclusão e portabilidade de dados, exigindo transparência e consentimento explícito para seu uso.

**"Esta Lei estabelece normas gerais de caráter nacional para o desenvolvimento, implementação e uso responsável de sistemas de inteligência artificial (IA) no Brasil, com o objetivo de proteger os direitos fundamentais e garantir a implementação de sistemas seguros e confiáveis, em benefício da pessoa humana, do regime democrático e de desenvolvimento científico e tecnológico."**

Projeto de Lei 2338/2023

# CATEGORIAS DE MACHINE LEARNING

**Curso:** Inteligência Artificial e Análise de Dados

## SUPERVISIONADO

Modelo que aprende com exemplos já resolvidos (dados com respostas certas) para fazer previsões ou tomar decisões. É como ensinar alguém mostrando perguntas e respostas.

Exemplo: Prever se o preço de uma ação vai subir ou descer no próximo dia com base em dados históricos, como volume de negociações, preço de fechamento, médias móveis e volatilidade.

## NÃO SUPERVISIONADO

Quando o modelo aprende sozinho com dados que não tem resposta certas (rótulos). Ele procura padrões escondidos ou grupos nos dados, como separar clientes em grupos com comportamentos parecidos.

Exemplo: Usar dados históricos de preços, volume de negociações e volatilidade para agrupar ações com comportamentos similares, sem saber previamente quais grupos existem. Isso pode ajudar a identificar setores ou padrões de mercado, como ações que tendem a reagir de forma semelhante a eventos econômicos.

## SEMI – SUPERVISIONADO

Combina dados rotulados (com respostas certas) e não rotulados (sem respostas para treinar um modelo). É útil quando rotular todos os dados é caro ou demorado, mas ainda se pode usar os dados não rotulados para melhorar o aprendizado.

Exemplo: Usar um pequeno conjunto de dados rotulados indicando se uma ação subiu ou desceu após um padrão específico (ex.: cruzamento de médias móveis) e um grande conjunto de dados não rotulados de preços históricos para treinar um modelo que identifica automaticamente padrões similares e prevê movimentos futuros do mercado.

## APRENDIZADO POR REFORÇO

O aprendizado onde o modelo aprende a tomar decisões em um ambiente, experimentando diferentes ações e recebendo recompensas ou penalidades, com feedback constante. O objetivo é descobrir a melhor estratégia para maximizar o ganho ao longo do tempo.

Exemplo: Um agente treinado com aprendizado por reforço decide quando comprar, vender ou manter uma ação com base em dados de mercado, como preço, volume e volatilidade. O objetivo do agente é maximizar os lucros ao longo do tempo, ajustando sua estratégia conforme as condições do mercado mudam.

**MACHINE LEARNING**  
Aprendizagem de máquina

**CATEGORIAS DO APRENDIZADO DE MÁQUINA**

Supervisionado

**MODELOS DE TREINAMENTO DA APRENDIZAGEM DE MÁQUINA - ALGORITMOS**

Classificação

Regressão

Não supervisionado

Agrupamento

Associação

Sumarização

Semi-supervisionado

Propagação de Rótulos

Autoencoders

Co-training

Aprendizado por Reforço

Q-Learning

Deep Q-Learning

Policy Gradient



# Como aplicar

FINTECH  
BOOST



# DEFINA O PROBLEMA

Comece com perguntas simples, como:  
O que quero resolver?  
Quais dados podem me ajudar?

Entenda o contexto, defina objetivos e transforme em um problema técnico.



# BUSQUE AS BASES DE DADOS

Comece com perguntas simples, como:  
Qual o problema financeiro?  
Que tipo de dados são necessários?

Entenda o objetivo da pesquisa, busque bases de dados para finança referente ao problema definido, escolha a base correta e use ferramentas para tratar os dados antes de treinar os dados.

Supervisionado

Não supervisionado

Semi supervisionado

Aprendizado por reforço

**FINTECH BOOST**

**Agrupamento (Clustering)**  
O que é: Identificar grupos ou clusters em dados não rotulados.

**Propagação de Rótulos**  
O que é: Espalhar rótulos conhecidos para dados não rotulados com base em conexões.

**Q-Learning**  
O que é: Algoritmo de aprendizado por reforço baseado em tabela de valores para decisões.

**Classificação**  
O que é: Atribuir um dado a uma categoria ou classe pré-definida.

**Associação**  
O que é: Descobrir relações entre itens ou variáveis em um dataset.

**Autoencoders**  
O que é: Redes neurais usadas para compressão e reconstrução de dados.

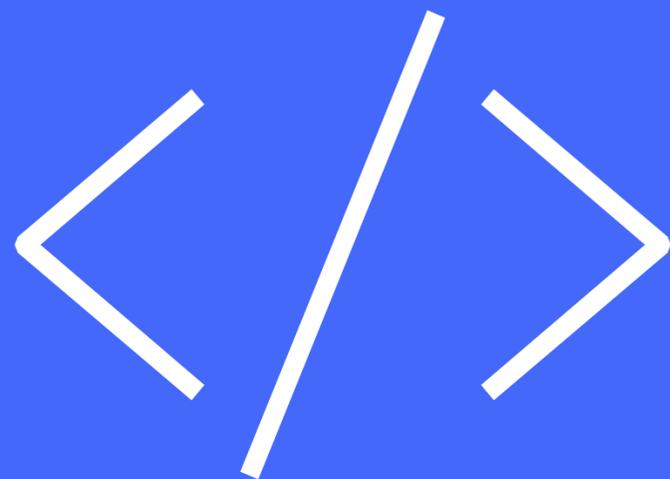
**Deep Q-Learning**  
O que é: Método de aprendizado por reforço que usa redes neurais para representar valores de ações.

**Regressão**  
O que é: Prever valores contínuos baseados em variáveis de entrada.

**Sumarização**  
O que é: Criar representações compactas ou resumos de dados.

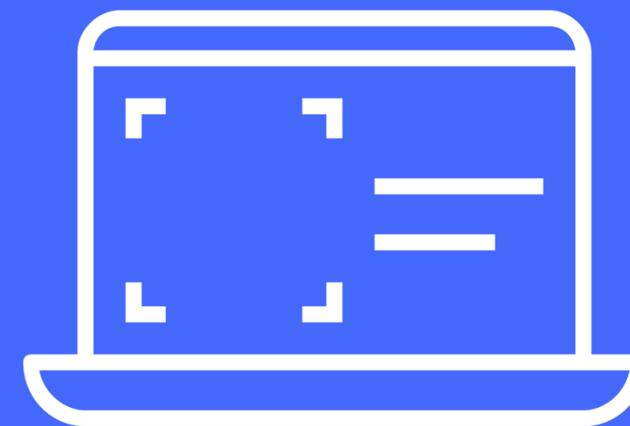
**Co-training**  
O que é: Treinamento colaborativo de modelos com conjuntos de características diferentes.

**Policy Gradient**  
O que é: Algoritmo de aprendizado por reforço que ajusta diretamente a política.



## CODE - CÓDIGO

Treinamento de modelos para finanças é geralmente feito em Python, porque essa linguagem possui um ecossistema robusto de bibliotecas que facilitam o trabalho com dados, modelagem de Machine Learning e visualizações. Além disso, Python é amplamente adotado na área de finanças por sua simplicidade e flexibilidade.



## NO CODE - SEM CÓDIGO

O treinamento no-code para finanças envolve um processo simples e visual que permite criar modelos de Machine Learning para resolver problemas financeiros sem necessidade de programar. Ele foca em ações práticas, aproveitando interfaces gráficas e automatização.



# Conclusão

FINTECH  
BOOST

# CAMINHO PARA TREINAR UM MODELO DE MACHINE LEARNING:

## 1. Definição do Problema

Identifique o problema a ser resolvido e os objetivos do modelo.

## 3. Preparação e Pré-processamento dos Dados

Nessa etapa é realizada a Limpeza (remoção de valores nulos, inconsistências ou duplicatas), a Transformação: (normalização/padronização dos dados) e a Divisão (separação dos dados em conjuntos de treinamento, validação e teste).

## 2. Coleta de Dados e Análise Exploratória

Reunião de todos os dados necessários para o modelo e análise exploratória

## 4. Escolha do Modelo

Selecione o tipo de algoritmo de aprendizado de máquina adequado ao problema (Regressão, Classificação ou Clustering).

## 5. Treinamento do Modelo

Use o conjunto de dados de treinamento para ajustar os parâmetros do modelo.



# CAMINHO PARA TREINAR UM MODELO DE MACHINE LEARNING:

## 6. Avaliação do Modelo

Teste do modelo com o conjunto de dados de validação, usando métricas como Acurácia, RMSE e Precisão.

## 7. Ajuste e Otimização

Etapa de ajustes no modelo como alteração de hiperparâmetros e escolhas de algoritmos alternativos.

## 8. Teste do Modelo

Utilização de conjunto de dados de teste nunca antes expostos ao modelo. Avaliação do desempenho em cenários reais.

## 9. Deploy (Implantação)

Selecione o tipo de algoritmo de aprendizado de máquina adequado ao problema (Regressão, Classificação ou Clustering).

## 10. Monitoramento e Manutenção

Monitoramento do modelo em uso no cenário real para o qual foi projetado e realização de atualizações do mesmo com novos dados, conforme a necessidade.

# Exemplos de projetos para finanças

**Curso:** Inteligência Artificial e Análise de Dados

## PREVISÃO DE PREÇO DE AÇÕES

Descrição: Use dados históricos de ações para prever o preço de fechamento do próximo dia.

Tipo de Problema: Regressão.

Passos:

Baixe os dados históricos de uma ação (ex.: AAPL).

Use variáveis como abertura, alta e volume.

Treine um modelo de Regressão Linear para prever o preço de fechamento.

## DETECÇÃO DE FRAUDES EM TRANSAÇÕES

Descrição: Identificar transações fraudulentas com base em padrões atípicos.

Tipo de Problema: Classificação ou Detecção de Anomalias.

Use variáveis como valor da transação e localização.

Treine um modelo com dados históricos de transações.

Avalie o modelo com precisão e recall.

## RECOMENDAÇÃO DE PRODUTOS FINANCEIROS

Descrição: Sugerir produtos financeiros (ex.: empréstimos, cartões de crédito) para clientes.

Tipo de Problema: Recomendação.  
Ferramentas: Scikit-learn, sistema baseado em conteúdo.

Passos:

Use dados sobre produtos já contratados por clientes.

Crie um sistema que recomenda produtos semelhantes com base em perfil.

## ANÁLISE DE SENTIMENTO EM NOTÍCIAS FINANCEIRAS

Descrição: Analisar notícias financeiras para determinar se o sentimento é positivo ou negativo.

Tipo de Problema: Classificação de Texto.

Passos:

Colete manchetes de notícias financeiras.

Classifique-as manualmente como positivas ou negativas.

Treine um modelo de classificação para automatizar o processo.

# Exemplo de projeto para finanças

Da Coleta de Dados ao  
Monitoramento Contínuo

Vá até o ChatGPT ou Google Gemini e pergunte:

## **Como instalar o Python pelo Anaconda?**

“Anaconda é uma plataforma de distribuição de ciência de dados e inteligência artificial de código aberto para linguagens de programação Python e R. Desenvolvida pela Anaconda, Inc. , uma empresa americana fundada em 2012, a plataforma é usada para desenvolver e gerenciar projetos de ciência de dados e IA.”

# Objetivo do Modelo

Desenvolver um modelo de aprendizado de máquina para prever o preço de fechamento de ações na bolsa de valores, utilizando dados históricos como:



Essa previsão ajuda investidores a tomar decisões mais informadas.

# Tecnologias Utilizadas

- Linguagem: **Python**.
- Bibliotecas:
  - **yfinance** para **coleta de dados financeiros**.
  - **scikit-learn** para **aprendizado de máquina**.
  - **pandas e numpy** para **manipulação de dados**.
  - **joblib** para **salvar o modelo**.

## Passo 1: Definição do Problema

Utilizamos dados históricos de ações para prever o preço de fechamento do próximo dia.

- Tipo de problema: **Regressão**.
- Tecnologias utilizadas: **Python, 'yfinance', 'scikit-learn'**.

**Objetivo:** Criar um modelo preditivo eficiente.

## Passo 2: Coleta de Dados e Análise Exploratória

Dados coletados de ações (exemplo: AAPL) usando **yfinance**.

### Código:

```
import yfinance as yf

symbol = "AAPL"
start_date = "2020-01-01"
end_date = "2023-12-31"

data = yf.download(symbol, start=start_date, end=end_date)
...
```

## Análise Exploratória de Dados (EDA):

– Entender os dados coletados antes do pré-processamento.

### Análises realizadas:

1. Resumo estatístico:

#### Código:

```
print(data.describe())
```

```
...
```

2. Evolução do preço de fechamento ao longo do tempo:

#### Código:

```
import matplotlib.pyplot as plt
```

```
plt.figure(figsize=(12, 6))  
plt.plot(data['Close'], label='Preço de Fechamento')  
plt.title('Evolução do Preço de Fechamento')  
plt.xlabel('Data')  
plt.ylabel('Preço ($)')  
plt.legend()  
plt.show()
```

### 3. Mapa de Correlação

#### Código:

```
import seaborn as sns  
  
plt.figure(figsize=(8, 6))  
sns.heatmap(data.corr(), annot=True, cmap='coolwarm')  
plt.title('Matriz de Correlação')  
plt.show()
```

#### Observações:

Identificar padrões, outliers e relações entre variáveis.

## Passo 3: Preparação e Pré-processamento dos Dados

1. Seleção de colunas: `Open`, `High`, `Low`, `Volume` e `Close`.
2. Remoção de dados ausentes para evitar problemas no treinamento.
3. Definição de variáveis:
  - **X (independentes)**: 'Open', 'High', 'Low', 'Volume'.
  - **y (dependente)**: 'Close'.

### Código:

```
data = data[['Open', 'High', 'Low', 'Volume', 'Close']]
data.dropna(inplace=True)

X = data[['Open', 'High', 'Low', 'Volume']]
y = data['Close']
```

## Passo 4: Escolha do Modelo

O modelo escolhido foi a **Regressão Linear**, devido à simplicidade e interpretabilidade.

### **Biblioteca usada:**

`'sklearn.linear_model.LinearRegression'`

## Passo 5: Treinamento do Modelo

- Dividimos os dados em treino e teste:
  - 80% para treino, 20% para teste.
- Treinamos um modelo de **Regressão Linear**.

### Código:

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
```

## Passo 6: Avaliação do Modelo

- Métricas de desempenho:
  - **Erro Quadrático Médio (MSE):** Mede o erro médio das previsões.
  - **R<sup>2</sup>:** Mede o quão bem o modelo explica os dados.

### Código:

```
from sklearn.metrics import mean_squared_error, r2_score

predictions = model.predict(X_test)
mse = mean_squared_error(y_test, predictions)
r2 = r2_score(y_test, predictions)

print(f"MSE: {mse}")
print(f"R2: {r2}")
```

## Passo 7: Ajuste e Otimização

Avaliar as métricas e ajustar hiperparâmetros do modelo (se necessário).

### Próximos passos:

- Adicionar variáveis derivadas (ex.: médias móveis, volatilidade).
- Testar outros modelos como árvores de decisão ou redes neurais.

## Passo 8: Teste do Modelo

Simular previsões com novos dados para avaliar o desempenho do modelo em cenários reais.

### Código:

```
import numpy as np

latest_data = np.array(data.iloc[-1][['Open', 'High', 'Low', 'Volume']]).reshape(1, -1)
predicted_close = model.predict(latest_data)[0]
print(f"Preço previsto: ${predicted_close:.2f}")
```

## Passo 9: Deploy (Implantação)

**Salvar** o modelo para uso em sistemas externos ou aplicações.

### Código:

```
import joblib  
  
joblib.dump(model, "modelo_regressao.joblib")
```

**Próximos passos:** Integrar o modelo a APIs ou sistemas web.

## 10: Monitoramento e Manutenção

### **Nesta etapa devemos:**

- Atualizar dados periodicamente para manter a precisão do modelo.
- Avaliar a performance do modelo ao longo do tempo.

**Próximos passos:** Implementar pipelines automáticos de monitoramento.

# Conclusão



## Resumo do Projeto:

- Utilizamos dados históricos para prever preços futuros com base em variáveis simples.
- Realizamos análise exploratória, pré-processamento, treinamento e avaliação.

## Resultados obtidos:

- Métrica  $R^2$ : Avaliou o quão bem o modelo se ajusta aos dados.
- Modelo salvo para reutilização.

## Próximos Passos:

- Explorar novas variáveis e modelos mais complexos para melhorar a performance.

**NÃO TENHA MEDO DA MATEMÁTICA!  
NÃO TENHA MEDO DE MINEIRAR DADOS!  
NÃO TENHA MEDO DOS CÓDIGOS!  
USE SEU PENSAMENTO CRÍTICO!**

---

# Bibliografias de apoio

---

1. **Google AutoML** – GOOGLE. AutoML. Disponível em: <https://cloud.google.com/automl>.
2. **Teachable Machine** – GOOGLE. Teachable Machine. Disponível em: <https://teachablemachine.withgoogle.com>.
3. **DataRobot** – DATAROBOT. DataRobot AI Platform. Disponível em: <https://www.datarobot.com>.
4. **BigML** – BIGML. BigML – Machine Learning Made Easy. Disponível em: <https://bigml.com>.
5. **Microsoft Power Platform** – MICROSOFT. Power Platform. Disponível em: <https://powerplatform.microsoft.com>.
6. **Azure Machine Learning Studio** – MICROSOFT. Azure Machine Learning Studio. Disponível em: <https://azure.microsoft.com/services/machine-learning>.
7. **KNIME** – KNIME. KNIME Analytics Platform. Disponível em: <https://www.knime.com>.
8. **RapidMiner** – RAPIDMINER. RapidMiner – Data Science Platform. Disponível em: <https://rapidminer.com>.
9. **Alteryx** – ALTERYX. Alteryx – Analytics Automation Platform. Disponível em: <https://www.alteryx.com>.
10. **PyCaret** – PYCARET. PyCaret – An open-source, low-code machine learning library in Python. Disponível em: <https://pycaret.org>.



## Contato



Zaika dos Santos



afrofuturismo.tech



@zaikadosantos



in/zaika-dos-santos



**FINTECH  
BOOST**

Muito  
Obrigada!