



## PROJETO APLICADO I PARTICIPANTES:

NOME:	RA:
Beatriz Ribeiro Porto Liberato	10441384
Gustavo da Conceição Guimarães	10723830
Dalicio Pereira de Novaes Junior	10441535
Pedro Lima da Silva Junior	10729027

## SUMÁRIO:

1. Definição da Linguagem de Programação e Bibliotecas Utilizadas.....	1
2. Análise Exploratória da Base de Dados.....	2
3. Tratamento da Base de Dados (Preparação e Treinamento) .....	2
4. Definição e Descrição das Bases Teóricas dos Métodos .....	3
5. Definição e Descrição de Como Será Calculada a Acurácia .....	3
6. Base definida para o trabalho .....	4
7. CRONOGRAMA PREVISTO .....	5
8. Link para o Github do projeto .....	6

### 1. Definição da Linguagem de Programação e Bibliotecas Utilizadas

Para a implementação do projeto, escolhemos a linguagem **Python**, devido à sua versatilidade, ampla comunidade e rica variedade de bibliotecas voltadas para ciência de dados e aprendizado de máquina.

As principais bibliotecas utilizadas serão:

- **Pandas** – para manipulação, limpeza e análise dos dados de forma estruturada.
- **NumPy** – para operações matemáticas e manipulação de arrays numéricos.

- **Matplotlib** – para a criação de gráficos e visualizações de dados.
- **Scikit-learn** – para a aplicação de técnicas de aprendizado de máquina e métricas de avaliação.

O uso combinado dessas ferramentas garante eficiência e precisão em todo o pipeline analítico.

## 2. Análise Exploratória da Base de Dados

A análise exploratória é uma etapa fundamental do projeto, pois permite compreender a estrutura dos dados e identificar padrões e tendências iniciais.

Para isso, serão realizadas as seguintes ações:

- **Leitura e inspeção do dataset:** Verificação do número de registros, tipos de dados e presença de valores ausentes.
- **Estatísticas descritivas:** Cálculo de médias, medianas, valores máximos e mínimos, além de medidas de dispersão.
- **Visualização dos dados:** Construção de gráficos de séries temporais para observar a evolução dos preços das ações (abertura, fechamento, máximo e mínimo) ao longo do tempo.
- **Correlação entre variáveis:** Análise da relação entre as diferentes colunas (Open, High, Low, Close e Volume) para identificar possíveis dependências.

Essa etapa fornece o embasamento necessário para as próximas fases do projeto e contribui para uma modelagem mais eficiente e assertiva.

## 3. Tratamento da Base de Dados (Preparação e Treinamento)

O tratamento dos dados é essencial para garantir a qualidade e a consistência das análises.

Nesta etapa, as seguintes ações serão realizadas:

- **Limpeza de dados:** Remoção de registros duplicados, tratamento de valores ausentes e conversão de tipos de dados, como transformar a coluna Date para o formato de data.
- **Padronização e arredondamento:** As colunas numéricas principais (Open, High, Low, Close e Adj Close) serão arredondadas para duas casas decimais após o ponto decimal, garantindo maior consistência e facilitando a leitura e interpretação dos resultados.

- Divisão em conjuntos de dados: O dataset será dividido em dados de **treinamento** e **teste** para avaliar a capacidade preditiva dos modelos.

#### 4. Definição e Descrição das Bases Teóricas dos Métodos

O projeto fará uso de métodos estatísticos e de aprendizado de máquina clássicos e modernos para previsão de séries temporais. A seguir, destacamos as bases teóricas principais:

- **Análise Estatística Descritiva:** Baseada em medidas de tendência central e dispersão, permite entender a estrutura histórica dos dados e identificar padrões.
- **Modelos de Séries Temporais (ARIMA):** Utilizado para previsões baseadas em dados passados, levando em consideração autocorrelações e tendências temporais.
- **Modelos de Machine Learning (Prophet e LSTM):**
  - Prophet: Modelo criado pelo Facebook, eficiente em capturar sazonalidades e tendências em séries temporais.
  - LSTM (Long Short-Term Memory): Rede neural recorrente avançada que consegue aprender dependências de longo prazo, sendo ideal para previsões complexas.

A combinação desses métodos permite comparar diferentes abordagens e selecionar a mais precisa para o contexto do projeto.

Essas transformações são fundamentais para reduzir ruídos, evitar erros de cálculo e melhorar o desempenho dos modelos de previsão.

#### 5. Definição e Descrição de Como Será Calculada a Acurácia

A acurácia dos modelos será avaliada por meio de métricas estatísticas amplamente utilizadas na previsão de séries temporais, permitindo medir o quão próximas as previsões estão dos valores reais. As principais métricas utilizadas serão:

- **MAE (Mean Absolute Error – Erro Absoluto Médio):** Mede a média dos erros absolutos entre valores previstos e reais.

- **RMSE (Root Mean Squared Error – Raiz do Erro Quadrático Médio):** Penaliza erros maiores, dando uma noção mais rigorosa do desempenho do modelo.
- **MAPE (Mean Absolute Percentage Error – Erro Percentual Médio Absoluto):** Indica a porcentagem média de erro em relação aos valores reais, facilitando a interpretação dos resultados.

Essas métricas permitirão comparar a performance dos modelos estatísticos e de machine learning e escolher aquele que apresentar a melhor precisão na previsão dos preços das ações.

## 6.Base definida para o trabalho

### Fonte e informações do dataset

Para realizar essa análise de otimização de descontos, será utilizado o

**International Business Machines Stocks from 2000.**

### Link para o Dataset:

<https://www.kaggle.com/datasets/middlehigh/international-business-machines-stocks-from-2000?select=IBM.csv>

- **Tipo de arquivo:** CSV
- **Origem dos Dados:** Dataset aberto, coletado da plataforma Kaggle.
- **Sensibilidade:** Não contém dados sensíveis ou pessoais identificáveis, contém dados históricos diários das ações da IBM a partir do ano 2000.
- **Restrições de Uso:** Não contém dados pessoais identificáveis, mas deve ser tratado conforme a legislação vigente.
- **Tipos de Dados:** A maioria dos dados é de tipo **float64 (Números de Ponto Flutuante)**, uma coluna com dados tipo **int64 (Números Inteiros)** e uma

coluna com dados tipo **object (Texto ou Misto)**. No total temos 7 colunas e 6118 registros.

### Colunas e significados

- **Date:** Data da negociação
- **Open:** Preço de abertura
- **High:** Preço máximo do dia
- **Low:** Preço mínimo do dia
- **Close:** Preço de fechamento
- **Volume:** Quantidade de ações negociadas

### 7.CRONOGRAMA PREVISTO

Data	Atividade	Responsáveis	Milestone
27/08/2025	Encontro para definição de papéis e cronograma detalhado	Todo o grupo	Início do Projeto.
04/09/2025	Encontro do grupo para confirmar dataset, e confirmar últimos detalhes sobre o projeto de acordo com dúvidas tiradas no encontro síncrono	Todo o grupo	Reestruturação da A1.
05/09/2025	Finalização do Documento para entrega da A1	Todo o grupo	Finalização da A1.
17/09/2025	Aquisição dos dados (IBM.csv) e limpeza inicial.	Gustavo e Beatriz	Dados Brutos Coletados.
24/09/2025	Análise exploratória e preparação dos dados.	Pedro e Dalicio	Dados Limpos e Prontos
03/10/2025	Finalização do Documento para entrega da A2	Todo o grupo	Finalização da A2.
16/10/2024	Cálculo de indicadores financeiros e visualizações iniciais.	Pedro e Gustavo	Análise Estatística Concluída
16/10/2024	Implementação dos modelos de base (ARIMA e Prophet).	Dalicio e Beatriz	Modelos de Previsão Implementados
16/10/2024	Implementação do modelo avançado (LSTM) e Esboço do Storytelling.	Todo o grupo	Modelo de ML Implementado
23/10/2024	Avaliação e comparação da performance de todos os modelos.	Todo o grupo	Avaliação de Modelos Concluída
24/10/2024	Finalização do Documento para entrega da A3	Todo o grupo	Finalização da A3.
05/11/2025	Relatório Técnico	Todo o grupo	Relatório Prontos
12/11/2025	Configuração do GitHub: Estruturação do repositório, README e upload do código.	Todo o grupo	Repositório no GitHub Publicado
19/11/2025	Gravação e Edição do Vídeo: Gravação da apresentação e edição final.	Todo o grupo	Vídeo da Apresentação Finalizado
19/11/2025	Entrega Final: Envio de todos os materiais (Relatório, Vídeo, Link do GitHub).	Todo o grupo	Conclusão do Projeto

## **8.Link para o Github do projeto**

[PedroJunior56/Projeto-Aplicado-II](https://github.com/PedroJunior56/Projeto-Aplicado-II)