



## 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. Método analítico aplicado.....	5
8. Medidas de acurácia.....	6
9. Descrição dos resultados preliminares.....	7
10. Storytelling.....	8
11. Cronograma Previsto.....	9
12. Github.....	10

### 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. Método analítico aplicado

O método analítico aplicado neste estudo foi a Regressão Linear, utilizada para prever o volume de ações negociadas da IBM a partir do preço de fechamento ajustado. Esse modelo estatístico foi escolhido por sua simplicidade interpretativa e pela capacidade de representar relações lineares entre variáveis quantitativas, permitindo analisar a influência direta do preço sobre o volume de negociações.

Os dados foram extraídos do arquivo IBM.csv, contendo informações históricas do mercado financeiro, como data, preços de abertura, alta, baixa, fechamento, fechamento ajustado e volume de transações. Após o tratamento da base, incluindo a conversão de formatos, arredondamento de valores e remoção de nulo foi realizada uma análise exploratória, destacando correlações e distribuições relevantes.

A etapa de modelagem envolveu a divisão dos dados em conjunto de treino (70%) e conjunto de teste (30%), assegurando uma avaliação independente do desempenho preditivo. O modelo de Regressão Linear foi então ajustado e empregado para gerar previsões sobre o conjunto de teste, relacionando os valores de fechamento ajustado com o volume previsto.

## 8. Medidas de acurácia

Para avaliar o desempenho do modelo, foram aplicadas métricas estatísticas amplamente utilizadas em tarefas de regressão e previsão, com o objetivo de mensurar a precisão das estimativas geradas. As métricas calculadas foram:

- **MAE** (Mean Absolute Error – Erro Absoluto Médio): 2.137.166,85
- **RMSE** (Root Mean Squared Error – Raiz do Erro Quadrático Médio): 3.255.843,90
- **MAPE** (Mean Absolute Percentage Error – Erro Percentual Médio Absoluto): 38,81%
- **MSE** (Mean Squared Error – Erro Quadrático Médio): 10.600.519.534.031,02
- **R<sup>2</sup>** (Coeficiente de Determinação): 0,165

Os resultados indicam que, embora o modelo consiga capturar parcialmente a relação entre o preço de fechamento ajustado e o volume negociado, o baixo valor de  $R^2$  (16,5%) demonstra que a relação linear explica apenas uma fração da variabilidade dos dados.

Além disso, o MAPE relativamente alto ( $\approx 39\%$ ) reforça que o comportamento do volume não é fortemente determinado apenas pelo preço de fechamento, sugerindo que variáveis adicionais — como indicadores de mercado, sentimento dos investidores ou eventos externos — poderiam aumentar o poder explicativo do modelo.

Mesmo com essas limitações, o método aplicado fornece uma base sólida para análises comparativas e para a futura inclusão de abordagens mais robustas, como modelos de séries temporais (ARIMA, Prophet) ou algoritmos de aprendizado de máquina multivariados.

## **9. Descrição dos resultados preliminares e produto gerado**

A análise exploratória e a modelagem estatística aplicadas à base de dados da IBM permitiram gerar resultados preliminares que servem como base para o desenvolvimento de um produto voltado à inteligência preditiva no mercado financeiro.

O modelo de Regressão Linear implementado teve como objetivo prever o volume de ações negociadas a partir do preço de fechamento ajustado. Embora tenha apresentado um baixo poder explicativo ( $R^2 = 0,165$ ) e erros médios relativamente altos (MAE = 2.137.166,85; RMSE = 3.255.843,90; MAPE = 38,81%), os resultados fornecem insights iniciais importantes sobre a viabilidade da abordagem preditiva e os limites de um modelo linear simples.

Esses achados indicam que o comportamento do volume de negociações não é exclusivamente determinado pelos preços, sendo influenciado também por variáveis externas, como volatilidade do mercado, notícias corporativas, índices setoriais e fatores macroeconômicos. Essa constatação orienta os próximos passos do projeto em direção à utilização de modelos mais complexos, como ARIMA, Prophet e técnicas de Machine Learning multivariadas, capazes de capturar padrões mais sofisticados e não lineares.

Com base nesses resultados, delineia-se o produto gerado: uma plataforma de inteligência preditiva voltada à previsão de preços e volumes de ações, utilizando modelos estatísticos e de aprendizado de máquina para gerar projeções trimestrais e semestrais. O sistema será treinado a partir de séries históricas de empresas listadas na B3 e em principais bolsas internacionais, permitindo a análise de tendências e o suporte à tomada de decisão.

### **Segmento de Clientes:**

O produto será direcionado a três principais perfis:

Investidores individuais, que buscam previsões confiáveis para orientar suas operações no mercado de ações;

Gestores de fundos e analistas, que necessitam de apoio técnico para decisões estratégicas;

Educadores financeiros, que poderão utilizar os dados da plataforma em simulações e atividades didáticas.

### **Proposta de Valor:**

A proposta central é oferecer previsões baseadas em séries temporais, indicadores macroeconômicos e comportamento histórico, integrando análises estatísticas e inteligência artificial para gerar insights de alta precisão e relevância para diferentes perfis de usuários.

Fontes de Receita

### **A estrutura de monetização será composta por múltiplos canais:**

Assinaturas mensais em diferentes níveis (básico, premium e institucional), com acesso gradativo às funcionalidades da plataforma;

Venda de relatórios preditivos sob demanda, personalizados por setor ou empresa;

Treinamentos e mentorias, voltados ao ensino de modelagem preditiva e utilização da ferramenta.

### **Estrutura de Custos:**

Os principais custos estão relacionados a:

Desenvolvimento e manutenção dos modelos preditivos;

Infraestrutura em nuvem para processamento e armazenamento de dados;

Equipe técnica composta por cientistas de dados, desenvolvedores, suporte e marketing;

Custos regulatórios e de compliance, especialmente no caso de integração com corretoras.

### **Estratégia de Lançamento e Crescimento:**

O plano de expansão da plataforma envolve parcerias com influenciadores do mercado financeiro, produção de conteúdo educativo gratuito para atrair usuários e um programa de afiliados com corretoras, que ampliará o alcance da base de clientes e consolidará a presença da marca no ecossistema financeiro digital.

## **10. Esboço do Storytelling**

Quando começamos o Projeto Aplicado II, a gente sabia que queria fazer algo que fosse além da sala de aula. A ideia de prever o comportamento do mercado de ações surgiu meio que num brainstorm, quando percebemos que muita gente investe sem entender os ciclos do mercado. E aí veio a pergunta: será que dá pra usar dados e inteligência artificial pra prever tendências trimestrais ou semestrais? A partir disso, começamos a montar o cronograma. Primeiro, fomos atrás das bases de dados históricos de ações da IBM, indicadores econômicos, e tudo que pudesse influenciar o mercado. A parte de tratamento foi intensa: limpar, organizar, arredondar valores, remover nulos... parecia que os dados nunca acabavam! Depois veio a modelagem. A gente começou com uma abordagem exploratória, calculando estatísticas como média, mediana, moda, desvio-padrão, variância, assimetria e curtose. Isso ajudou a entender a distribuição dos preços e volumes.

Descobrimos que os preços de abertura, alta, baixa e fechamento se movem juntos, com correlações altíssimas. Já o volume tinha uma correlação inversa quando o preço cai, o volume sobe.

A primeira tentativa de previsão foi com regressão linear. Queríamos saber se o fechamento ajustado influenciava o volume negociado. O modelo até rodou, mas o  $R^2$  foi baixo, e o erro médio quadrático alto. A gente entendeu que esse tipo de relação talvez não seja linear e que precisávamos de algo mais robusto. Foi aí que entraram as métricas de avaliação. Usamos Baseline Naive (Last value), MAE, RMSE, MAPE e SMAPE pra medir o desempenho. Também aplicamos bootstrap pra calcular intervalos de confiança e o teste de Diebold-Mariano pra comparar previsões. Criamos gráficos de dispersão, histogramas de erro e visualizações de tendência tudo pra entender onde o modelo acertava e onde falhava.

No meio do caminho, surgiu a ideia de transformar isso num produto. Imaginamos uma plataforma onde o usuário pudesse ver previsões trimestrais, simular carteiras, receber alertas. Algo que ajudasse tanto quem está começando quanto quem já investe há anos.

O MVP ficou simples, mas funcional: um dashboard com gráficos, projeções e relatórios. E aí começamos a pensar no modelo de negócio. Assinatura mensal, relatórios premium, integração com corretoras... tudo isso entrou no plano. Mais do que um projeto acadêmico, isso virou uma experiência real. A gente aprendeu a trabalhar em equipe, a lidar com dados complexos, a pensar como empreendedores. E no fim, ficou claro que o maior valor não está só na previsão está em transformar dados em decisões.

Esse projeto foi nossa forma de mostrar que, com dedicação e propósito, dá pra transformar uma ideia em algo que realmente faz sentido no mundo lá fora.



## 11.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

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

[PedroJunior56/Projeto-Aplicado-II](#)