

Leonardo Pereira Gonçalves, Nome do Colega 2, Nome do Colega 3, Nome do Colega 4, Nome do Colega 5

## **Análise Preditiva de Preços de Smartphones para o Mercado de 2025**

UEMG - Sistemas de informação  
Mineração de Dados - Rení Aparecido Norberto Pinto

Passos - MG  
14 de novembro de 2025

# Resumo

Este trabalho detalha o desenvolvimento de um projeto de mineração de dados focado na previsão de preços de smartphones, com base em um conjunto de dados simulado para o mercado de 2025. O processo abrange desde a estruturação do ambiente de trabalho colaborativo e a carga inicial dos dados, até a análise preliminar de sua consistência. Utilizando a biblioteca Pandas em um ambiente Google Colab, foi realizada a extração e a primeira inspeção dos dados, identificando as tarefas de limpeza necessárias para as etapas subsequentes. O objetivo final é construir e avaliar um modelo de regressão capaz de estimar o preço de um aparelho a partir de suas especificações técnicas. **Palavras-chave:** Mineração de Dados. Aprendizado de Máquina. Previsão de Preços. Regressão. Análise de Dados.

# Sumário

<b>Sumário</b> . . . . .	2
<b>0.1</b> <b>Introdução</b> . . . . .	3
0.1.1     Objetivos . . . . .	3
0.1.1.1    Objetivo Principal . . . . .	3
0.1.1.2    Objetivos Secundários . . . . .	3
<b>0.2</b> <b>Metodologia</b> . . . . .	3
0.2.1     Ferramentas e Ambiente . . . . .	3
0.2.2     Processo de Extração, Transformação e Carga (ETL) . . . . .	4
0.2.2.1    Extração e Carga Inicial . . . . .	4
0.2.2.2    Análise Preliminar e Direcionamento da Limpeza . . . . .	4
<b>0.3</b> <b>Análise Exploratória de Dados (Em Desenvolvimento)</b> . . . . .	5
<b>0.4</b> <b>Modelagem e Resultados (Em Desenvolvimento)</b> . . . . .	5
<b>0.5</b> <b>Conclusão</b> . . . . .	5
<b>REFERÊNCIAS</b> . . . . .	6

## 0.1 Introdução

O mercado de smartphones é segmentado por diversas faixas de preço, e o valor final de um aparelho é influenciado por uma complexa combinação de fatores como memória RAM, qualidade da câmera, capacidade da bateria, entre outros. Para consumidores e analistas de mercado, entender quais características mais contribuem para o custo de um aparelho é um desafio analítico. O problema central que este projeto busca resolver é: **quais especificações técnicas são os principais fatores que determinam o preço de um smartphone no mercado previsto para 2025?**

Para investigar esta questão, foi utilizado o conjunto de dados *Global Mobile Prices 2025 Extended*, obtido na plataforma Kaggle (SHAHZADI, 2024), que simula as características de 1000 modelos de smartphones.

### 0.1.1 Objetivos

Com base no problema definido, os seguintes objetivos foram traçados para guiar o projeto.

#### 0.1.1.1 Objetivo Principal

Desenvolver um modelo de aprendizado de máquina capaz de prever o preço (`price_usd`) de um smartphone com base em suas especificações técnicas.

#### 0.1.1.2 Objetivos Secundários

- Realizar uma análise exploratória para entender a correlação entre as especificações técnicas e o preço.
- Identificar o posicionamento de preço das principais marcas.
- Treinar e comparar diferentes modelos de regressão para determinar o de melhor performance.
- Gerar insights sobre quais características mais agregam valor a um smartphone.

## 0.2 Metodologia

Esta seção descreve as ferramentas, o conjunto de dados e os procedimentos realizados na fase inicial do projeto, que serviram de base para as etapas de análise e modelagem.

### 0.2.1 Ferramentas e Ambiente

O projeto foi desenvolvido utilizando a linguagem **Python 3**. O ambiente de desenvolvimento escolhido foi o **Google Colab**, por sua facilidade de uso e colaboração. A

principal biblioteca utilizada nesta fase inicial foi a **Pandas**, para manipulação e análise de dados. O controle de versão e o trabalho colaborativo foram gerenciados através de um repositório no **GitHub**.

### 0.2.2 Processo de Extração, Transformação e Carga (ETL)

Esta etapa foi focada na extração dos dados e em uma análise preliminar para guiar as tarefas de transformação e limpeza.

#### 0.2.2.1 Extração e Carga Inicial

O primeiro passo consistiu em carregar o arquivo `Global_Mobile_Prices_2025_Extended.csv` em um DataFrame do Pandas. O código utilizado para a carga e a primeira inspeção é apresentado abaixo.

```
1 # Importando a biblioteca Pandas
2 import pandas as pd
3
4 # Caminho do arquivo no ambiente Colab
5 caminho = '/content/Global_Mobile_Prices_2025_Extended.csv'
6
7 # Carga do CSV para um DataFrame
8 df = pd.read_csv(caminho)
9
10 # Exibindo informações gerais (tipos de dados e contagem de nulos)
11 df.info()
12
13 # Verificando a soma de valores nulos por coluna
14 print(df.isnull().sum())
```

Listing 1 – Código para carga e inspeção inicial dos dados.

#### 0.2.2.2 Análise Preliminar e Direcionamento da Limpeza

A execução do código (Listagem 1) revelou que o conjunto de dados possui 1000 registros e 15 colunas, sem nenhum valor ausente (nulo), o que simplifica a etapa de tratamento.

Contudo, a inspeção visual dos dados apontou a necessidade das seguintes tarefas de limpeza e transformação, que serão executadas pelo integrante responsável pela etapa de ETL:

- **Limpeza da Coluna ‘model’:** A coluna que identifica o modelo do aparelho contém ruídos numéricos (ex: “A98 111”). Estes devem ser removidos para padronizar os nomes.
- **Transformação de Variáveis Categóricas:** A coluna `5g_support`, com valores “Yes” e “No”, precisará ser convertida para um formato numérico (0 ou 1) para

ser utilizada em modelos matemáticos. Outras colunas, como `processor`, também deverão ser tratadas (ex: via One-Hot Encoding).

- **Remoção de Colunas Irrelevantes:** A coluna `year` possui um valor constante (2025) para todos os registros e, portanto, não agrega valor preditivo ao modelo, podendo ser removida.

### 0.3 Análise Exploratória de Dados (Em Desenvolvimento)

*[Esta seção será desenvolvida pelo integrante responsável pela análise exploratória. Aqui serão inseridos gráficos e visualizações para investigar a distribuição dos preços, a correlação entre as variáveis e outros insights extraídos dos dados já limpos.]*

### 0.4 Modelagem e Resultados (Em Desenvolvimento)

*[Esta seção será desenvolvida pelo integrante responsável pela modelagem. Serão detalhados os algoritmos de regressão escolhidos, o processo de treinamento e teste, e a apresentação dos resultados de performance dos modelos, como o Erro Quadrático Médio (RMSE) e o  $R^2$ .]*

### 0.5 Conclusão

*[Esta seção será preenchida ao final do projeto, resumindo os resultados, discutindo as limitações do modelo e do dataset, e sugerindo possíveis melhorias ou trabalhos futuros.]*

# Referências

SHAHZADI, A. *World Smartphone Market 2025*. Kaggle, 2024. Acesso em: 14 nov. 2025. Disponível em: <<https://www.kaggle.com/datasets/shahzadi786/world-smartphone-market-2025>>.