

# RELATÓRIO – ANÁLISE DE MINERAÇÃO DE DADOS

**Disciplina:** Tópicos Especiais em Computação I

**Aluno:** Eduardo Thomaz Centenaro - 103526

**Data:** 06/07/2025

---

## Dataset Utilizado

- **Nome:** Titanic Dataset
  - **Fonte:** [Kaggle – Titanic: Machine Learning from Disaster](#)
  - **Registros:** 891 passageiros
  - **Colunas:** 12 variáveis (como idade, sexo, tarifa, classe, etc.)
  - **Objetivo:** Prever quem sobreviveu ao naufrágio com base nas características dos passageiros
- 

## Técnica Aplicada

Foi utilizada a técnica de **classificação**, por meio do algoritmo **Random Forest**, para prever a sobrevivência dos passageiros com base em variáveis como idade, sexo, tarifa, classe e tamanho da família.

---

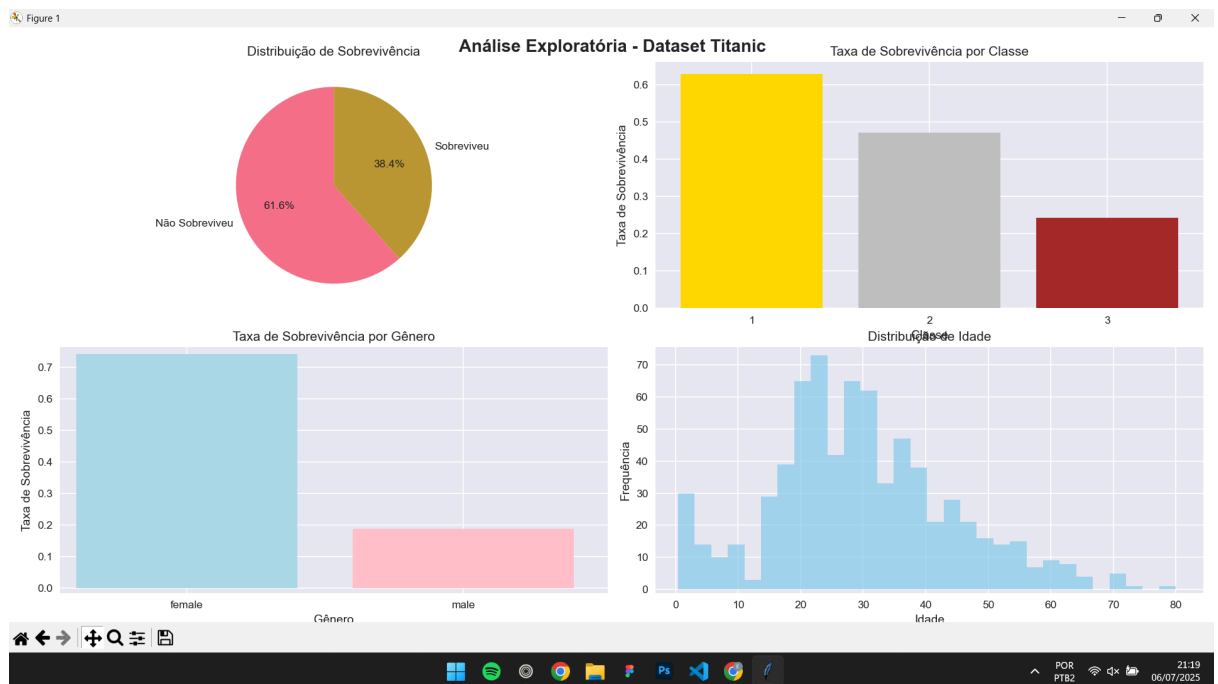
## Principais Resultados

- **Acurácia do modelo:** aproximadamente **82%**
- **Variáveis mais importantes:** sexo, tarifa paga, idade, classe e tamanho da família
- **Sobrevivência por gênero:** mulheres tiveram alta taxa de sobrevivência (~74%), enquanto a maioria dos homens não sobreviveu

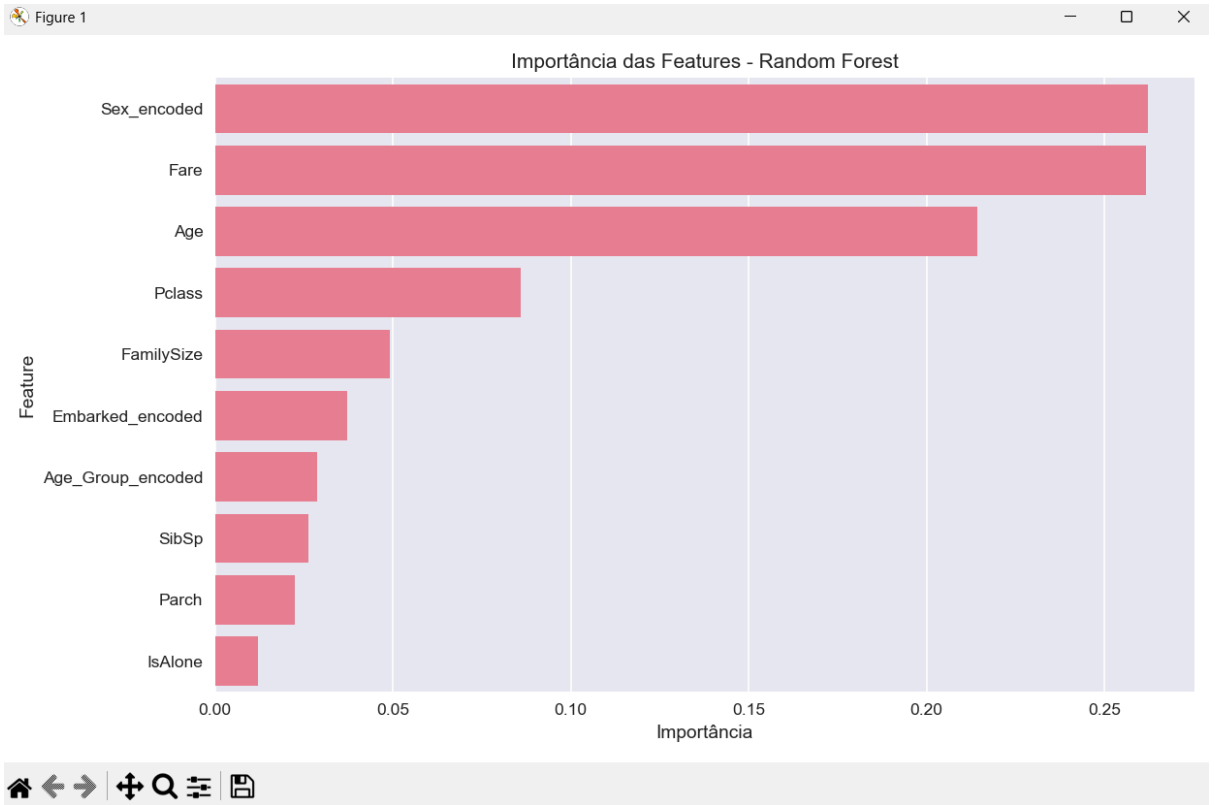
- **Classe social:** passageiros da 1ª classe tiveram maior taxa de sobrevivência
- **Idade:** passageiros mais jovens apresentaram maior chance de sobrevivência

## Gráficos Gerados

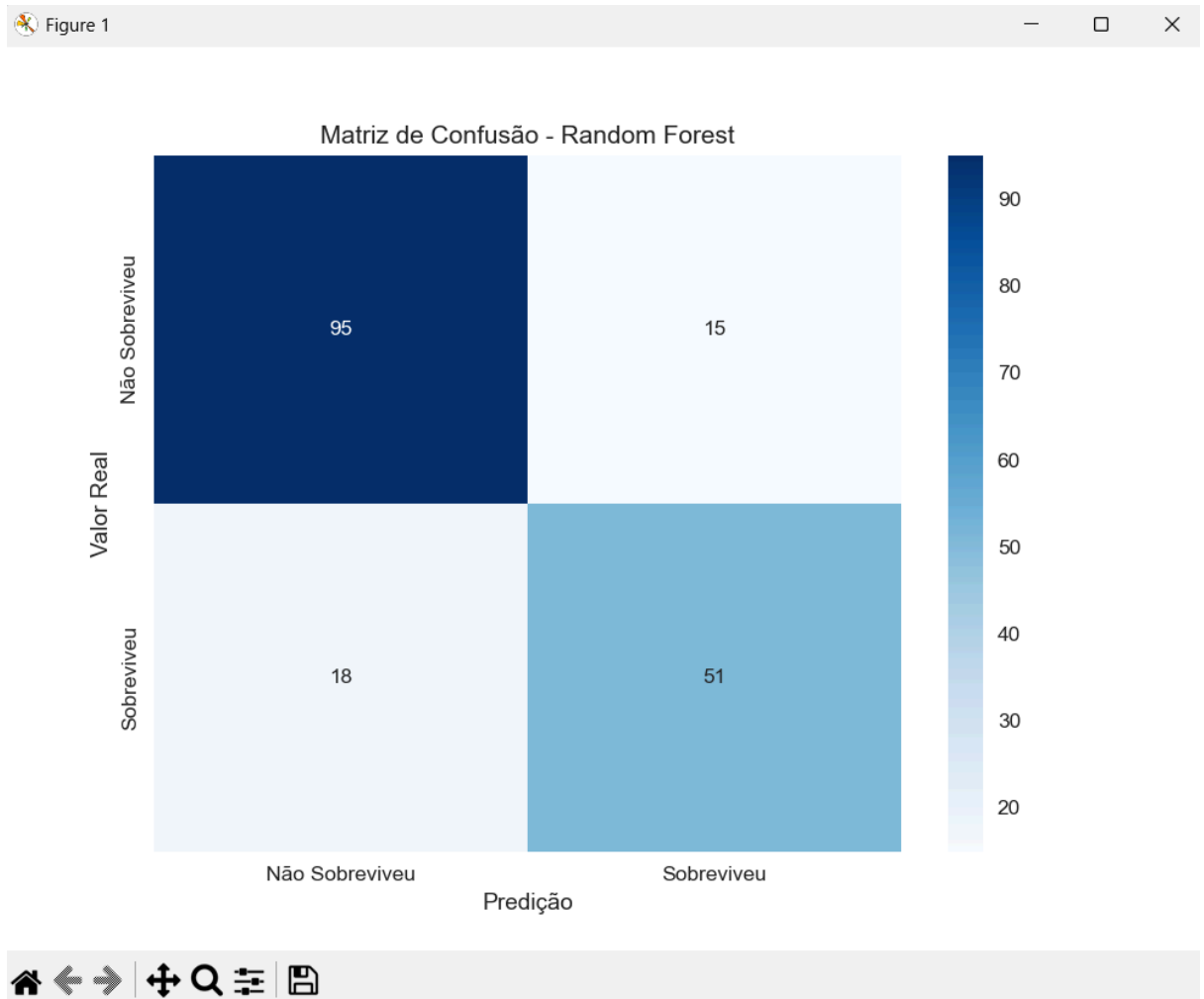
1. **Distribuição de Sobrevivência:** gráfico de pizza
2. **Taxa de sobrevivência por classe e gênero:** gráficos de barra
3. **Distribuição das idades:** histograma



4. **Importância das features:** gráfico de barras



## 5. Matriz de confusão: desempenho da predição



## Conclusão

A análise confirmou que **sexo**, **classe social** e **idade** foram os principais fatores de influência na sobrevivência dos passageiros. O modelo Random Forest obteve bons resultados e as visualizações facilitaram a interpretação dos padrões encontrados nos dados históricos do Titanic.