

# Análise da Base de Dados "CritérioProvas" Usando Árvores de Decisão

## 1. Introdução

Este relatório apresenta a análise da base de dados **CritérioProvas**, que contém informações sobre notas de provas e percentual de faltas de alunos, com o objetivo de classificar os estudantes como **Aprovado** ou **Reprovado**. O estudo inclui a construção da árvore de decisão, sua visualização e a análise dos resultados com a matriz de confusão.

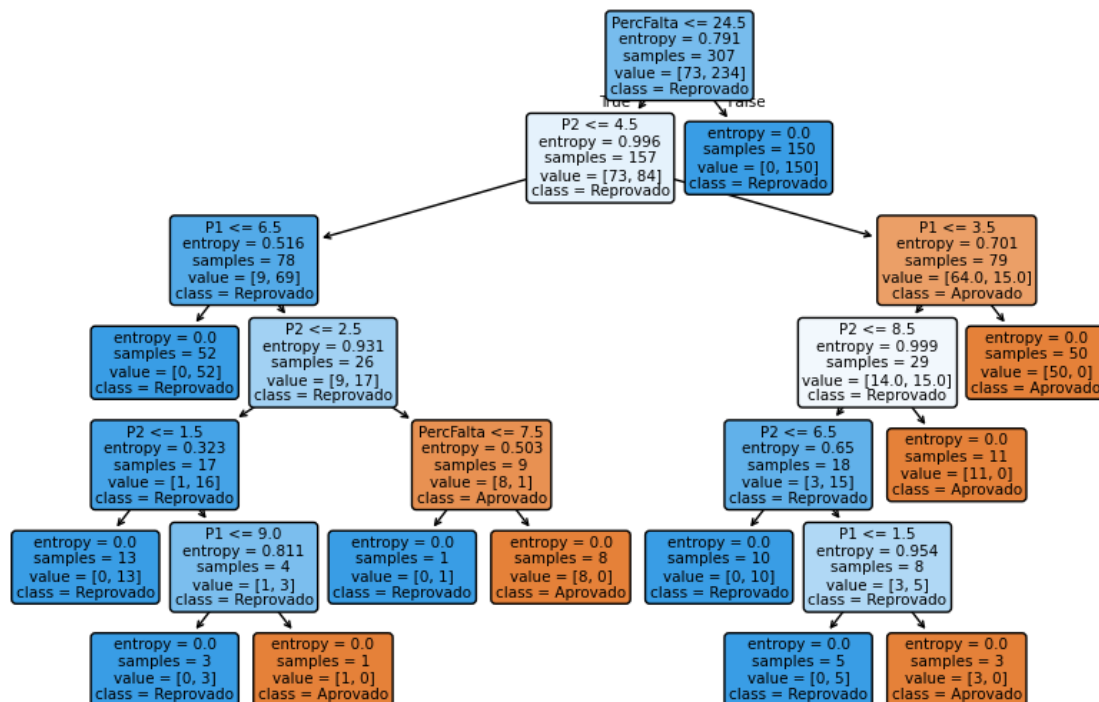
## 2. Descrição da Base de Dados

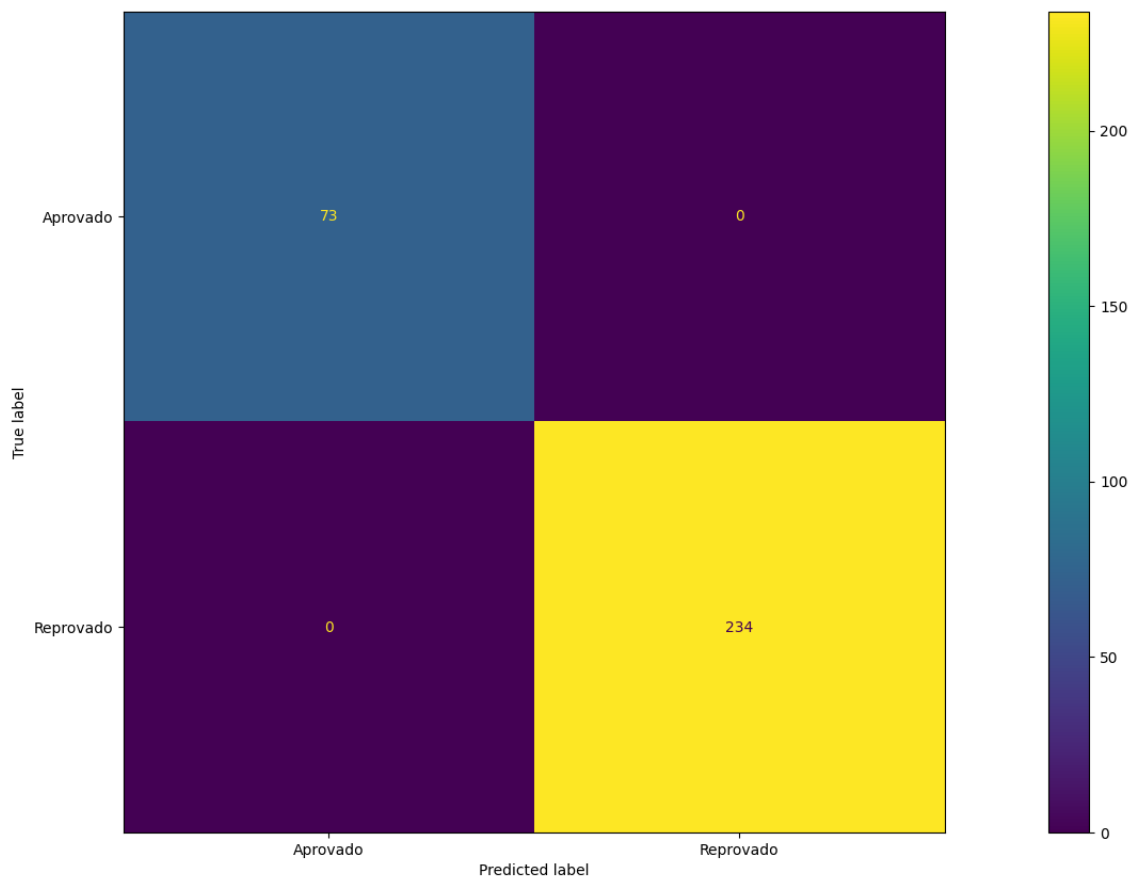
A base de dados **CritérioProvas** contém três atributos preditivos e uma variável alvo:

- **P1**: Nota da primeira prova (**inteiro de 0 a 10**).
- **P2**: Nota da segunda prova (**inteiro de 0 a 10**).
- **PercFalta**: Percentual de faltas (**inteiro representando a porcentagem de faltas do aluno**).
- **resultado**: Classificação final do aluno (**Aprovado** ou **Reprovado**).

## 3. Construção do Modelo de Árvore de Decisão

O modelo de árvore de decisão foi construído utilizando o algoritmo **ID3 (Information Gain - Entropy)** disponível no DecisionTreeClassifier da biblioteca scikit-learn. Esse algoritmo escolhe os atributos que proporcionam maior ganho de informação para a classificação.





#### 4. Conclusão

- A análise mostrou que **Notas das provas (P1, P2) e Percentual de Faltas (PercFalta)** são determinantes para a aprovação ou reprovação de um aluno.
- A árvore de decisão apresentou **alta interpretabilidade**, permitindo visualizar claramente as regras utilizadas na classificação.
- A matriz de confusão indicou que o modelo teve **poucos erros**, mostrando boa precisão na classificação.
- A comparação com outros algoritmos indicou que **Random Forest pode aumentar a acurácia**, mas com menor interpretabilidade.