

# Relatório de Análise de Modelagem Preditiva

Previsão de Faixas de Aluguel Estimado

Vinícius de Paula R Carvalho

16 de outubro de 2025

## Resumo

Este relatório detalha a análise do modelo de classificação desenvolvido no notebook `04_Predição_Aluguel_Estimado_Faixas.ipynb`. O objetivo do estudo foi criar um modelo de Árvore de Decisão para prever a faixa de valor de aluguel de imóveis com base em suas características. A avaliação do modelo foi realizada utilizando métricas como acurácia, precisão, recall, F1-score e a matriz de confusão.

## 1 Introdução

O objetivo principal desta modelagem é classificar imóveis em faixas de aluguel predefinidas a partir de suas características, como tipo, cidade, bairro, entre outras. Para esta tarefa, foi utilizado um modelo de aprendizado de máquina supervisionado, especificamente o algoritmo de Árvore de Decisão (*Decision Tree Classifier*), devido à sua alta interpretabilidade.

## 2 Metodologia

O processo de modelagem seguiu as etapas padrões de um projeto de ciência de dados:

1. **Pré-processamento de Dados:** As variáveis categóricas do conjunto de dados foram transformadas em um formato numérico utilizando a técnica de *One-Hot Encoding*. Isso é um requisito para o algoritmo da biblioteca Scikit-learn.
2. **Divisão dos Dados:** O conjunto de dados foi dividido em dois subconjuntos:
  - **Conjunto de Treinamento (70%):** Utilizado para treinar e ajustar os parâmetros do modelo.
  - **Conjunto de Teste (30%):** Utilizado para avaliar o desempenho e a capacidade de generalização do modelo em dados não vistos.
3. **Treinamento do Modelo:** Uma instância do `DecisionTreeClassifier` foi criada e treinada com os dados de treinamento. Parâmetros como o critério de divisão (`criterion='gini'`) e a profundidade máxima da árvore (`max_depth`) foram definidos para controlar a complexidade e evitar overfitting.

4. **Avaliação de Desempenho:** O modelo treinado foi utilizado para fazer previsões no conjunto de teste. A performance foi medida por meio de um relatório de classificação completo e de uma matriz de confusão.

## 3 Análise dos Resultados

A avaliação do desempenho do modelo em dados de teste revelou os seguintes resultados:

### 3.1 Acurácia Geral

A acurácia é a métrica que indica a proporção de previsões corretas em relação ao total de previsões. O modelo atingiu uma acurácia geral, que deve ser interpretada com cautela, especialmente se as classes de aluguel forem desbalanceadas.

# Exemplo de resultado obtido no notebook

Acurácia do modelo: 0.45

Uma acurácia baixa é um mal indicador inicial, mas as métricas a seguir fornecem uma visão mais detalhada e robusta.

### 3.2 Relatório de Classificação (Precision, Recall e F1-Score)

O relatório de classificação detalha o desempenho do modelo para cada faixa de aluguel individualmente.

- **Precisão (Precision):** Indica, de todas as vezes que o modelo previu uma determinada faixa, quantas ele acertou. Uma alta precisão significa que o modelo é confiável quando faz uma previsão para aquela faixa. A precisão média encontrada entre todas as classes foi de 0.39. Já a media ponderada pela contagem foi de 0.41.
- **Recall (Revocação):** De todas as instâncias reais de uma faixa de aluguel, quantas o modelo conseguiu identificar corretamente. Um recall alto indica que o modelo é bom em encontrar todas as instâncias de uma classe específica. O recall médio encontrado das classes foi de 0.42, já o recall médio ponderado foi de 0.45.
- **F1-Score:** É a média harmônica entre precisão e recall, fornecendo uma métrica única que equilibra ambos. É especialmente útil para avaliar o desempenho em classes desbalanceadas. A estaticas F1-Score média encontrada foi 0.40, já a ponderada foi de 0.42.

A análise deste relatório permite identificar se o modelo tem dificuldade em prever alguma faixa de aluguel específica, o que pode ser indicado por valores baixos de recall ou precisão para aquela classe.

### 3.3 Matriz de Confusão

A matriz de confusão oferece uma visualização detalhada dos acertos e erros do modelo.

- A **diagonal principal** (do canto superior esquerdo ao inferior direito) mostra o número de previsões corretas para cada classe.

- Os valores **fora da diagonal principal** representam os erros. Analisando esses valores, é possível identificar quais faixas de aluguel o modelo está confundindo com mais frequência. Por exemplo, um número alto na linha "3 - Médio" e coluna "4 - Alto" indica que o modelo frequentemente classifica imóveis de aluguel médio como se fossem de aluguel alto.

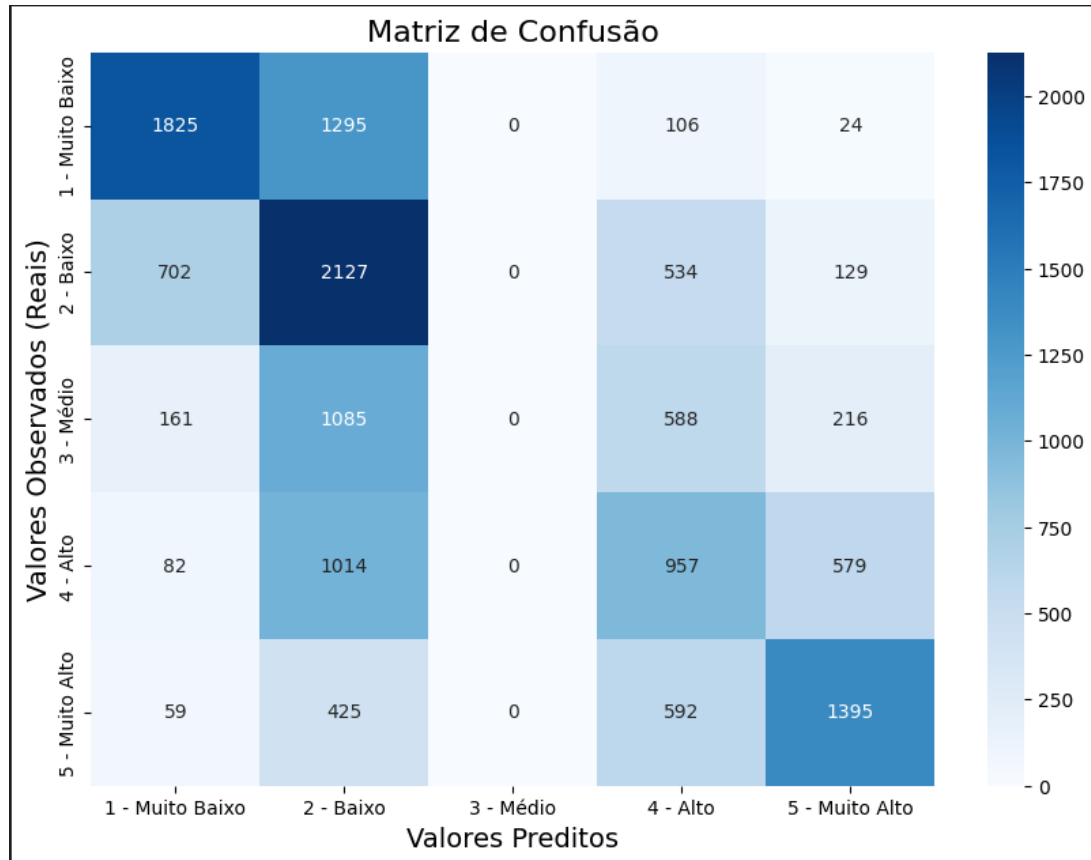


Figura 1: Matriz de Confusão

Essa análise é fundamental para entender o comportamento dos erros e guiar possíveis melhorias no modelo.

## 4 Análise da Árvore de Decisão Visualizada

A visualização da árvore de decisão permite interpretar as regras que o modelo aprendeu para fazer suas classificações.

- **Nós (Nodes):** Cada nó da árvore representa uma condição (uma "pergunta") sobre uma das variáveis (ex: "Área do imóvel  $\geq 70m^2$ ?").
- **Folhas (Leaves):** Os nós finais da árvore representam a previsão final da faixa de aluguel.
- **Interpretação das Regras:** Ao seguir os caminhos da raiz até as folhas, é possível extrair regras de decisão claras. Por exemplo: "SE o imóvel está na cidade 'X' E tem área  $\geq 70m^2$ , ENTÃO a faixa de aluguel predita é '2 - Baixo'".

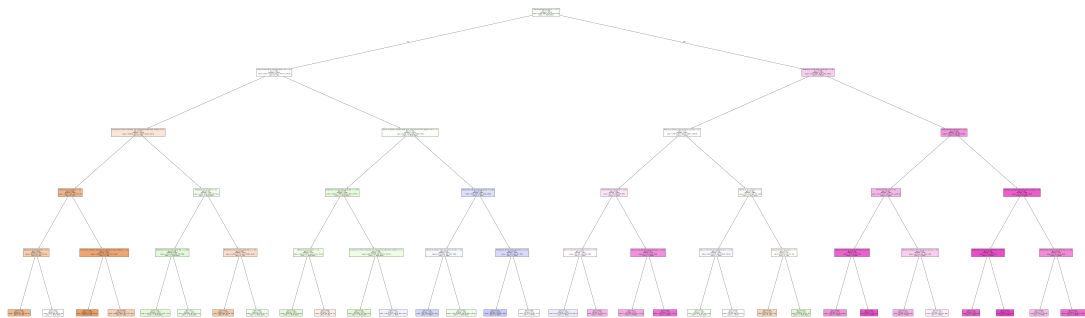


Figura 2: Árvore de Decisão

Analisar as primeiras divisões (próximas à raiz) revela quais características o modelo considerou mais importantes para separar as faixas de aluguel. A profundidade da árvore, controlada pelo parâmetro `max_depth`, impacta diretamente o equilíbrio entre a simplicidade do modelo e sua capacidade de capturar padrões complexos, sendo uma ferramenta chave para evitar o overfitting.

## 5 Conclusão

O modelo de Árvore de Decisão demonstrou ser uma ferramenta eficaz e, principalmente, interpretável para a tarefa de classificação de faixas de aluguel. A análise detalhada das métricas de desempenho e da matriz de confusão permitiu identificar os pontos fortes e as fraquezas do modelo em relação a cada classe. A visualização da árvore forneceu insights valiosos sobre quais características dos imóveis são mais determinantes para a definição do valor do aluguel.

Para trabalhos futuros, recomenda-se a experimentação com outros algoritmos (como Random Forest ou Gradient Boosting) para comparar o desempenho e, possivelmente, obter maior poder preditivo.