

Previsão de Preços de Imóveis em Perdizes Utilizando Modelos de Regressão Supervisionada

Patrick Gonçalves, Nicholas dos Santos Leal, Gustavo Ibara

Instituto de Informática – Universidade Presbiteriana Mackenzie (UPM)

Faculdade de Tecnologia e Informação (FCI) – São Paulo, SP – Brazil

10409363@mackenzista.com.br, 10389067@mackenzista.com.br,
10409210@mackenzista.com.br

Abstract. This project applies Artificial Intelligence techniques to predict real estate prices in the Perdizes district of São Paulo. The analysis uses structural variables such as property size, number of bedrooms, number of bathrooms, parking spaces, and price per square meter, obtained from a private and anonymized dataset. The methodology includes exploratory data analysis, data preprocessing, train-test splitting, and the application of supervised regression models, specifically Linear Regression and Decision Tree Regression. Models were evaluated using metrics such as MSE, RMSE, MAE, MAPE, and R^2 . Results show that the Decision Tree model achieved the best performance, reaching an R^2 of 0.9236 on the test set, outperforming Linear Regression. The study demonstrates the potential of AI to support real estate decision-making by providing accurate, transparent, and reliable price estimates.

Resumo. O presente projeto aplica técnicas de Inteligência Artificial para prever o preço de imóveis no bairro de Perdizes, em São Paulo. A análise utiliza variáveis estruturais como metragem, número de quartos, banheiros, vagas de garagem e valor do metro quadrado, provenientes de um dataset privado previamente anonimizado. A metodologia inclui análise exploratória dos dados, preparação da base, divisão treino-teste e aplicação de modelos de regressão supervisionada, especificamente Regressão Linear e Árvore de Decisão. Os modelos foram avaliados por métricas como MSE, RMSE, MAE, MAPE e R^2 . Os resultados demonstram que a Árvore de Decisão apresentou melhor desempenho, alcançando R^2 de 0,9236 no conjunto de teste, superando a Regressão Linear. O estudo evidencia o potencial da IA para apoiar decisões no mercado imobiliário e fornecer estimativas transparentes e confiáveis.

1. Introdução

a. Contextualização

O mercado imobiliário brasileiro apresenta forte variação de preços devido a fatores como localização, metragem, infraestrutura e características estruturais dos imóveis. Estimar corretamente o valor de uma propriedade é um desafio relevante tanto para proprietários quanto para compradores. A Inteligência Artificial surge como alternativa

capaz de identificar padrões complexos e produzir estimativas mais rápidas e consistentes.

b. Justificativa

Modelos de IA permitem automatizar a avaliação de imóveis, reduzindo subjetividade na precificação e aumentando a transparência do processo. Além disso, possibilitam análises em larga escala, oferecendo suporte para imobiliárias, investidores e empresas do setor. A escolha do bairro de Perdizes se dá pela disponibilidade de um dataset privado confiável.

c. Objetivo

O objetivo deste trabalho é desenvolver e comparar modelos de regressão supervisionada para prever o valor aproximado de imóveis no bairro de Perdizes, utilizando variáveis estruturais presentes no dataset. O estudo avalia o desempenho da Regressão Linear e da Árvore de Decisão por meio das métricas MSE, RMSE, MAE, MAPE e R².

d. Opção do projeto

Machine Learning (Regressão).

2. Fundamentação Teórica

A previsão de preços de imóveis é tradicionalmente tratada como um problema de regressão, no qual algoritmos aprendem relações entre variáveis estruturais e o valor final da propriedade. A Regressão Linear modela relações lineares entre atributos e preço, sendo amplamente utilizada pela sua interpretabilidade. Já as Árvores de Decisão permitem capturar relações não lineares e interações entre variáveis, particionando o espaço de dados em regras de decisão. As métricas MSE, RMSE, MAE, MAPE e R² são comumente utilizadas para avaliar a acurácia de modelos de regressão. Estudos recentes (Truong, 2020; Hernes, 2024) indicam que modelos supervisionados, mesmo os mais simples, podem alcançar excelente desempenho quando o dataset está limpo e bem estruturado.

3. Descrição do Problema

O problema abordado consiste na dificuldade de estimar preços de imóveis de maneira padronizada e confiável, considerando as variações próprias de um bairro urbano consolidado como Perdizes. Avaliações manuais podem ser subjetivas e inconsistentes. Assim, busca-se desenvolver um modelo de IA capaz de oferecer previsões rápidas, replicáveis e transparentes, contribuindo para decisões mais assertivas no setor imobiliário.

4. Aspectos Éticos e Responsabilidade no Desenvolvimento da Solução

O uso de Inteligência Artificial na precificação de imóveis traz benefícios, mas também exige responsabilidade. Como os modelos dependem da qualidade dos dados, é essencial garantir que não haja vieses que favoreçam ou prejudiquem determinados perfis de imóveis ou regiões. Todo o dataset utilizado neste trabalho foi devidamente

anonimizado para preservar a privacidade dos proprietários. Outro aspecto ético importante é garantir transparência: modelos preditivos não devem ser usados como verdade absoluta, mas como apoio à decisão. A responsabilidade final permanece com o avaliador humano, que deve interpretar os resultados e considerar fatores externos não presentes no dataset. Assim, o uso da IA deve ser feito com cautela, garantindo equidade, privacidade e integridade dos dados.

5. Dataset

O dataset utilizado neste trabalho consiste em registros reais de imóveis do bairro de Perdizes, São Paulo, provenientes de uma base privada previamente tratada e anonimizada. A base contém variáveis estruturais relevantes para o processo de especificação: metragem total do imóvel, número de quartos, número de banheiros, vagas de garagem, valor do metro quadrado e valor aproximado do imóvel. A preparação dos dados envolveu análise exploratória (histogramas, boxplots, matriz de correlação), identificação e remoção de outliers e verificação de valores ausentes. Em seguida, os dados foram divididos em conjuntos de treino e teste. A variável alvo escolhida foi o valor aproximado do imóvel, utilizando as demais variáveis como atributos preditores.

6. Metodologia e Resultados Esperados

A metodologia aplicada seguiu as etapas:

1. **Carregamento e limpeza da base**, incluindo verificação de outliers e normalização quando necessário.
2. **Análise exploratória**, com visualização de distribuições, correlações e relações entre variáveis e o preço.
3. **Separação da base em treino (80%) e teste (20%)**.
4. **Treinamento de dois modelos supervisionados**:
 - o Regressão Linear com padronização (StandardScaler)
 - o Árvore de Decisão (DecisionTreeRegressor)
5. **Avaliação dos modelos** utilizando MSE, RMSE, MAE, MAPE e R².
6. **Comparação direta** dos resultados dos modelos e interpretação dos erros.
7. **Implementação de função interativa** para permitir que o usuário informe características de um imóvel e receba previsões dos dois modelos.

7. Resultados

Os resultados obtidos foram:

Regressão Linear – Teste

- MSE: 7.526.565.579
- RMSE: 86.755
- MAE: 68.964
- MAPE: 11.58%

- R^2 : 0.8835

Árvore de Decisão – Teste

- MSE: 4.934.057.564
- RMSE: 70.242
- MAE: 46.727
- MAPE: 6.56%
- R^2 : **0.9236**

A Árvore de Decisão apresentou desempenho superior em todas as métricas, especialmente na redução do erro percentual médio e no aumento de R^2 . O resultado indica que relações não lineares entre as variáveis têm papel relevante na especificação dos imóveis da região. O modelo linear teve desempenho consistente, porém inferior ao modelo baseado em árvore.

8. Conclusão

O projeto atingiu o objetivo proposto ao desenvolver e comparar modelos preditivos para estimar o preço de imóveis em Perdizes. As análises demonstraram que a Árvore de Decisão foi o modelo mais eficaz, alcançando R^2 de 0,9236 no conjunto de teste e apresentando menor erro percentual. Os resultados confirmam que técnicas de IA podem apoiar com precisão o processo de avaliação de imóveis, oferecendo estimativas confiáveis e replicáveis. Como trabalhos futuros, sugere-se testar modelos mais avançados, como Random Forest, Gradient Boosting e Redes Neurais, além de incorporar variáveis externas como localização exata, idade do imóvel e dados socioeconômicos.

9. Endereço GitHub e Endereço do vídeo no YouTube

GitHub: <https://github.com/patrickkalvesz/previsao-precos-imoveis-IA>

YouTube: <https://www.youtube.com/watch?v=1WRPYA1jGdA>

10. Referências

- Q. Truong, “Housing Price Prediction via Improved Machine Learning,” *Procedia Computer Science*, vol. 174, pp. 433–442, 2020.
 M. Hernes, “Real estate valuation using machine learning,” *Procedia Computer Science*, vol. 227, pp. 321–329, 2024.