

UNIVERSIDADE ESTADUAL PAULISTA – UNESP  
FACULDADE DE CIÊNCIAS DE BAURU  
DEPARTAMENTO DE COMPUTAÇÃO

## Previsão de Preços de Casas Utilizando Redes Neurais

Autor  
Paulo Henrique Dionysio

2024

# 1 Introdução

A precificação de imóveis é um desafio complexo, pois envolve diversas variáveis que influenciam os preços, como localização, área construída, idade do imóvel e condições estruturais. Com a crescente disponibilidade de dados e avanços em inteligência artificial, é possível criar modelos que aprendam padrões a partir de dados históricos e gerem previsões precisas de preços. Este trabalho apresenta uma aplicação prática de técnicas de aprendizado de máquina para prever preços de casas com base em um conjunto de dados previamente definido.

## 2 Objetivos

O principal objetivo deste trabalho é desenvolver um modelo de aprendizado de máquina para prever preços de casas, utilizando como base dados históricos de propriedades. Os objetivos específicos incluem:

- Explorar e pré-processar os dados fornecidos para identificar e tratar valores ausentes e variáveis categóricas.
- Construir um modelo de rede neural utilizando a biblioteca TensorFlow com a função de ativação ReLU para aprendizado profundo.
- Avaliar o desempenho do modelo em dados de validação e gerar previsões precisas em um conjunto de dados de teste.
- Visualizar os resultados por meio de gráficos interativos que facilitem a análise das previsões e dos resíduos do modelo.

## 3 Metodologia

### 3.1 Materiais

- Linguagem de Programação: Python.
- Bibliotecas Utilizadas:
  - Streamlit: Criação de uma interface interativa para o projeto.
  - TensorFlow: Construção e treinamento do modelo de rede neural.
  - Pandas e NumPy: Manipulação e análise de dados.

- Scikit-learn: Pré-processamento de dados e divisão de conjuntos de treino e validação.
- Plotly: Geração de gráficos interativos
- Conjunto de Dados: Arquivos train.csv e test.csv contendo variáveis numéricas e categóricas sobre características das propriedades e seus respectivos preços de venda.

## 3.2 Método proposto

### 1. Carregamento e Pré-processamento dos Dados:

- Dados categóricos foram transformados em variáveis dummy.
- Dados ausentes foram preenchidos utilizando a mediana das variáveis numéricas.
- Escalonamento foi aplicado utilizando o MaxAbsScaler para normalizar as variáveis numéricas.

### 2. Divisão dos Dados:

- O conjunto de dados foi dividido em conjuntos de treino e validação em uma proporção de 50%.

### 3. Modelo de Rede Neural:

- Arquitetura com camadas densas (dense layers) utilizando a função de ativação ReLU.
- Regularização adicionada com dropout para evitar overfitting.
- Otimização realizada com o algoritmo Adam e a métrica de erro utilizada foi o Erro Logarítmico Médio Quadrático (MSLE).

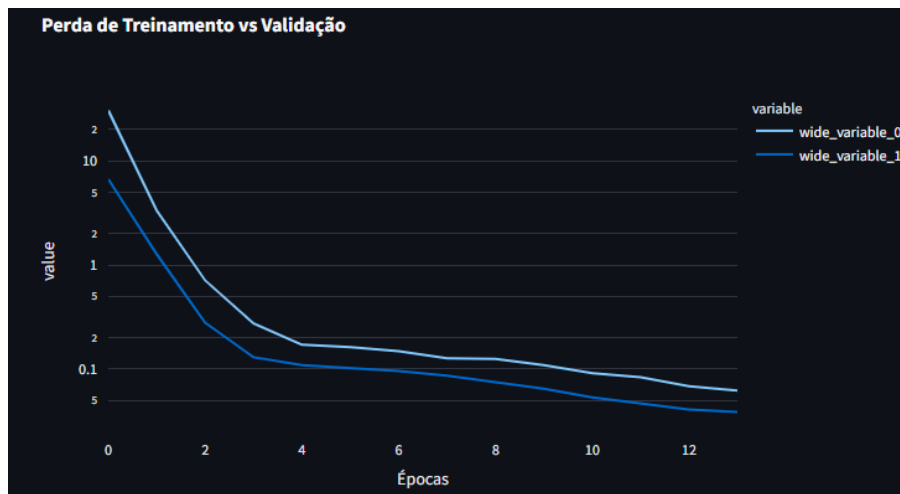
### 4. Treinamento e Avaliação:

- O modelo foi treinado com callbacks para interromper o treinamento se não houvesse melhoria significativa.
- Gráficos interativos foram gerados para comparar os valores reais e previstos, além de visualizar a distribuição dos resíduos.

## 4 Respostas Obtidas

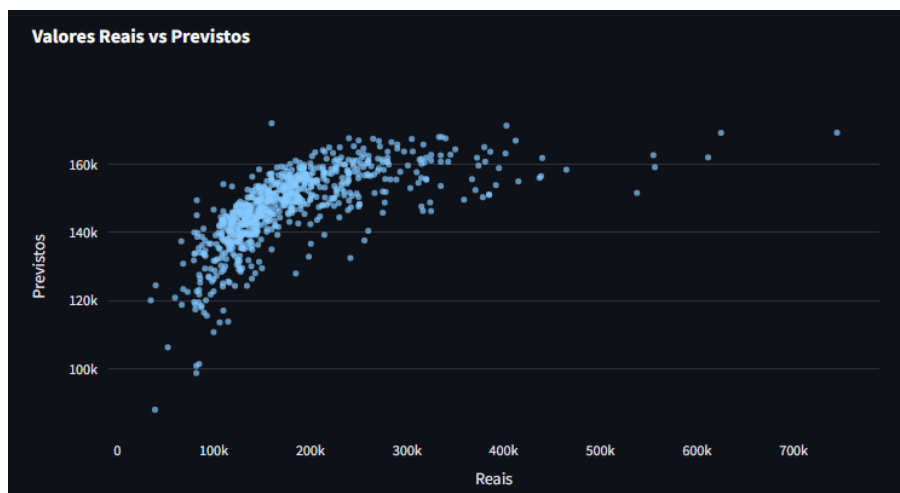
Após o treinamento do modelo:

- O gráfico de perda mostrou uma redução significativa no erro de treinamento e validação, indicando que o modelo aprendeu a partir dos dados sem sinais aparentes de overfitting.



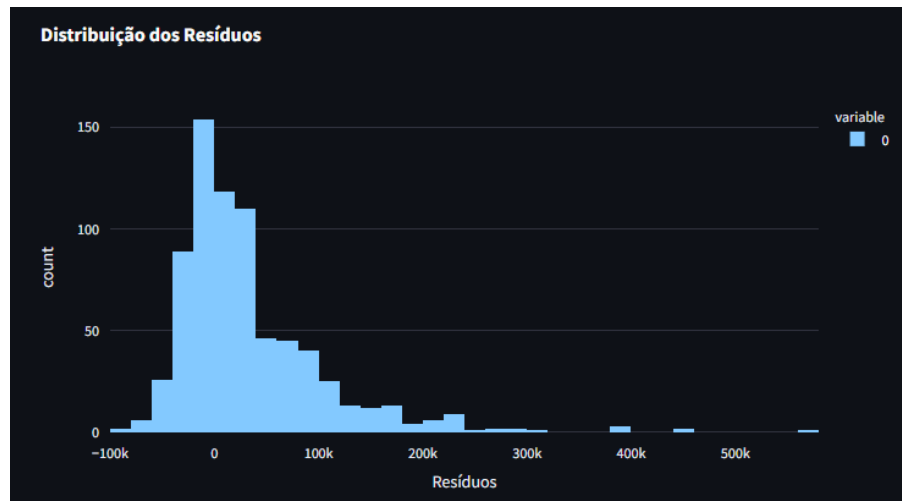
Perda de Treinamento vs Validação

- As previsões geradas para os dados de validação apresentaram uma forte correlação com os valores reais, conforme evidenciado pelo gráfico de dispersão.



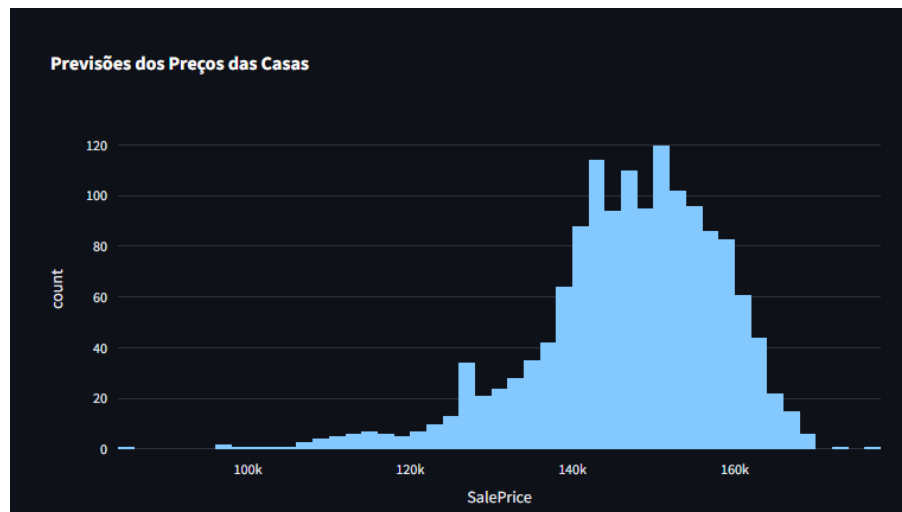
Valores Reais vs Previstos

- A análise da distribuição dos resíduos revelou que o modelo produziu erros relativamente pequenos, centrados em torno de zero, o que reforça a qualidade das previsões.



Distribuição de Resíduos

- No conjunto de dados de teste, as previsões mostraram uma distribuição coerente com os preços de venda esperados.



Previsão dos Preços das Casas

## 5 Conclusão

O projeto demonstrou a aplicabilidade de técnicas de aprendizado profundo na previsão de preços de casas. A abordagem adotada, envolvendo redes neurais com a função de ativação ReLU, apresentou bons resultados, evidenciados pelos gráficos de avaliação e pelas distribuições dos resíduos.

Para trabalhos futuros, seria interessante explorar:

- A otimização dos hiperparâmetros do modelo.
- O uso de arquiteturas mais avançadas, como redes neurais convolucionais ou modelos baseados em gradient boosting.
- A integração de dados externos, como informações sobre economia e infraestrutura local, para enriquecer as previsões.

Esse trabalho reforça o papel da inteligência artificial como uma ferramenta poderosa para resolver problemas complexos do mundo real.