Machine Learning Approach for Predicting Airbnb Prices in NYC

Austin Ressler, Ayden Dillon, Lucas Chough, Aiden Colberg, Antonio Caporossi

Abstract

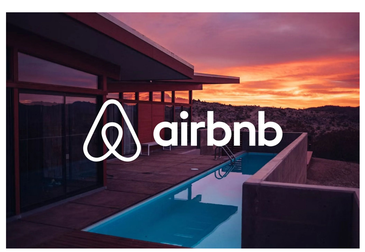
As of January 2024, there were over 38 thousand Airbnb listings in New York City ranging from only $20 to a night up to roughly $10,000 a night. When looking for the right Airbnb, it would be nice to look at the predicted cost to make a trip planning easier, comparing neighborhood prices, and many more benefits. Out of 38,734 initial Airbnb listings, only 5,841 were analyzed after shaving outliers to develop a predictive model. Using a fully connected neural network using regression, the model on average was able to achieve roughly 78% accuracy of the total cost for all airbnb listings predicted. This model does not only assist in predicting the price of an Airbnb in New York City, but also helps anyone using the app to plan a vacation or even listing their own.

Introduction

This project seeks to develop an advanced predictive model to estimate Airbnb rental prices in the vibrant and competitive market surrounding New York City. By harnessing the capabilities of machine learning, the project employs a fully connected neural network tailored specifically for regression tasks. This type of neural network is well-suited for handling complex, multidimensional data, making it an ideal choice for analyzing the diverse factors that influence rental pricing.

Key elements considered in the analysis include geographic location, property characteristics such as size and amenities, and host-specific features like ratings and reviews. By incorporating these variables into the model, the goal is to uncover patterns and relationships that drive pricing decisions in the Airbnb market.

The outcome of this approach will not only provide a robust framework for accurately forecasting rental prices but also offer valuable insights into the dynamics of short-term rental markets in one of the world's most iconic cities. This predictive capability has practical applications for hosts seeking to optimize their earnings, guests looking for fair pricing, and market analysts studying urban rental trends.



Method

Data Preprocessing

- **Data Cleaning:** Addressed missing values, removed duplicates, and encoded categorical variables (e.g., boroughs, property types).
- **Outlier Removal:** Eliminated extreme values in variables like reviews per month and prices using interquartile range (IQR) analysis.
- **Feature Scaling:** Applied MinMaxScaler to standardize input features for improved model performance.

Model Architecture

- **Structure:** Fully connected neural network with 5 layers, progressively decreasing in size:
  - 512 → 256 → 128 → 64 → 1 neurons (linear activation for regression).
  - Relu activation for each hidden layer.
- **Training Configuration:**
  - *Loss Function:* Mean Absolute Error (MAE).
  - *Optimizer:* Adam with a learning rate of 0.0015.
  - *Training Epochs:* 60, with a batch size of 128.
  - *Early Stopping:* Callbacks to prevent overfitting.

Evaluation

- **Train-Test Split:** 80:20 ratio.
- **Performance Metric:** Mean Absolute Percentage Error (MAPE).

Results & Conclusions

Conclusions

- The model demonstrates **78% accuracy** in predicting Airbnb rental prices, highlighting its practical application for both users and hosts.
- By leveraging data-driven insights, the model can assist in **trip planning** and **pricing strategies**, improving the user experience on platforms like Airbnb.

Future Work

- Incorporate additional features such as:
  - **Seasonal trends** to capture temporal variations.
  - **Sentiment analysis** from user reviews for enhanced predictions.
- Expand the model's integration into platforms like Airbnb to provide real-time pricing recommendations.

Impact

- This tool has the potential to **transform interactions** within the Airbnb ecosystem, benefiting both guests and hosts through better decision-making and pricing optimization.

Epochs	Loss ($)	Mean_Absolute_Percentage_Error	Val_Loss ($)	Val_Mean_Absolute_Percentage_Error
1/23	32.4553	43.4018	24.4866	36.1884
6/23	19.3903	25.3292	19.1693	24.0987
11/23	18.8021	24.2544	18.5866	24.5326
16/23	18.7118	23.9309	18.7156	25.3867
21/23	18.3617	23.5043	19.3321	22.9329
23/23	18.4531	23.6280	18.7657	22.8828