Northeastern University

EAI 6000 - FINAL PROJECT

Airbnb Price Prediction

Group 5

Prof. Sergey Shevchenko

May 16th, 2022

Submitted By :-

Akash Raj Member

Abhigna Ramamurthy

Durga Bhanu Nayak

Harshit Gaur

Jeseeka Shah

Project Description

The Project covers analysis of the vital data set "**Airbnb Business Case**". Visitors and hosts have been using Airbnb since 2008 to expand travel possibilities and give a more unique, personalised way of experiencing the world. Use **LIME** and **SHAP** for model interpretability and predictive modelling to develop market-specific forecasts with various variables. The data used in this research offers information on listing activity and indicators in the United States in 2019. It offers over 800,000 listings and 27 key characteristics, enough to learn more about hosts, geographical availability, and the analytics needed to make projections and draw conclusions.

Airbnb has established their market in accommodation business successfully since 2008 and has provided hosts and visitors with attractive options to visit places across the globe. Using the data collected by Airbnb for their listings, we have employed ML models that predict listing price based on user inputs over the UI. In this process we have used regression models like Linear Regression, Decision Tree, Random Forest, and XGBoost algorithms to calculate the price from the selective input variables and have used LIME and SHAP for model interpretability. For this project, we have limited the scope of the training data for United States listings only.

Business Objectives

Airbnb has stood the test of time and has proved to be a go to site for travel planning. With their newly introduced features of the Airbnb App, we can expect a lot of growth from the company. The dataset is taken from their official website *insideairbnb.com.* This provides a great platform to analyze the factors that contribute to the listings' prices. The project objective is to help customers and hosts get a sense of Airbnb listing prices based on the important features that define the listing property. This is aimed at Airbnb too as they can optimize their search models and service better to their customers.

Predictive Modelling

We start with linear regression to predict the Airbnb listing prices and then dive into the tree models like Decision trees, Random Forest, and Extreme Gradient Boosting. Advantages of comparing multiple models is that we will have an option of selecting the best performing model given the benchmarking like Root Mean Squared Error (RMSE) and accuracy of the model performance. Gradient Boosting is an intuitive algorithm that makes most of its predictions simple to understand. Boosting is a durable and reliable way for preventing and cubs from over-fitting. On medium, small datasets with subgroups and structured datasets with few characteristics, XGBoost excels. It's an excellent strategy to take because most real-world issues involve classification and regression, two tasks for which XGBoost reigns supreme. One of the biggest challenges we faced was with outliers as XGBoost is very sensitive to outliers. Hence outliers were removed to avoid this situation. The biggest drawback is that the method isn't very scalable as the estimators rely on past predictors to determine their accuracy, the technique is difficult to simplify.