

Predicting Airbnb Prices in New York City

Ian MOON, Felipe GARVANO, Yasmina HOBEIKA, Akshay SHASTRI, Pallav SAHU

January 23, 2023

1 Introduction

Airbnb has become a popular alternative to traditional hotels and has disrupted the hospitality industry. Through Airbnb, individuals can list their own properties as rental places. In New York City alone, there are around 40,000 listings. While traditional hotels have teams that carefully measure demand and supply to adjust pricing, as a host, it can be challenging to determine the optimal price for a listing. The variation in types of listings can also make it difficult for renters to get an accurate sense of fair pricing. In this project, we will use ensemble learning methods to predict the price of Airbnb listings in New York City.

2 The Data Set

2.1 Description

The data set for this project is obtained from Kaggle and contains the listings in New York City in 2019. The data set includes 15 features on listings, including:

- Name of the listing
- Neighborhood
- Price
- Review information
- Availability

The data set contains around 47,000 listings.

2.2 Access to the Data

The data set can be downloaded on Kaggle [here](#).

2.3 Pre-processing

Before starting the project, the data should be pre-processed by cleaning and transforming it into a format that can be used. This will include removing missing values, converting categorical variables to numerical variables, and normalizing the data.

2.4 Task for the Project

1. Apply all approaches taught in the course and practiced in lab sessions (Decision Trees, Bagging, Random forests, Boosting, Gradient Boosted Trees, AdaBoost, etc.) on this data set. The goal is to predict the target variable: price of the listing.
2. Compare performances of models using various metrics learned in class.
3. Write a report in .tex format that addresses all these guidelines with a maximal page number of 5 (including figures, tables, and references). The quality of writing and presentation of the report will be taken into account.