**Predicting Airbnb Prices**

Edrick Keane

June 21, 2020

**1. Introduction**

1.1 Background

Jakarta is the capital and largest city of Indonesia where over 10 million people live and it has a population density of 14,464 people per square kilometer. Jakarta consists of five administrative cities and one administrative regency. The administrative cities of Jakarta are: Central Jakarta, West Jakarta, South Jakarta, East Jakarta and North Jakarta. The only administrative regency in Jakarta is the Thousand Islands. In this project, I limit the scope of my research to the districts of Central Jakarta [1].

1.2 Problem

As the capital city of Indonesia, Jakarta is one of the most popular destinations in Indonesia. Thus, there is a lot of potential profit that could be obtained by property owners through listing a home on Airbnb in Jakarta. However, it’s hard for a new host to determine the rate for nightly stay. This research aims to solve this problem by predicting the best rate by using machine learning model which is trained with data from Airbnb listings.

**2. Data Description**

To solve the problem, the data from *Foursquare API* is not enough so I used the *Airbnb API* to get the listings of given districts within Central Jakarta. I extract the listing ID, latitude, longitude, number of guests, bathrooms, bedrooms, beds, number of reviews, room type, rating, minimum nights, maximum nights, rate type and rate. I combine the data obtained through each district query and drop the duplicates according to its listing id. I then evaluate the number of missing data for each features. Since the quantity of missing data on the rating feature is significant, I decided to drop those data. I also exclude the data with room type of shared room, number of guests more than 4, and minimum nights more than 3. I then proceed by removing rate outliers.

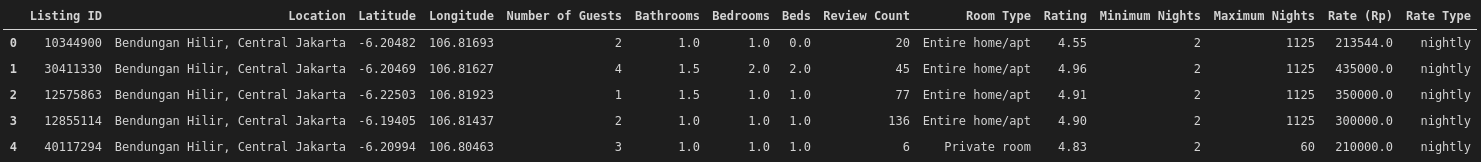


Figure 1. Airbnb Listings

**3. Methodology**

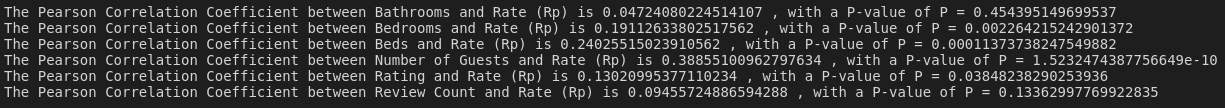
Data obtained from *Airbnb API* is limited to 50 per request. In order to obtain more data, instead of querying with ‘Central Jakarta’, districts within Central Jakarta were used for the queries. I then proceed by dropping the duplicate data. I evaluated the missing data. Because the quantity of missing data on the rating feature is significant, I dropped the rows with missing data. After the missing data were processed, I dropped the data with room type value of shared room, number of guests more than 4, and minimum nights more than 3. I then proceed by removing rate outliers with *IQR*. I then reset the index of dataframe.

After the data has been preprocessed, we remove irrelevant features. Then Pearson Correlation Coefficient was used to calculate the correlation between features.



Figure 2. Pearson Correlation Coefficient Between Features

P-value was also used to calculate the statistical significance between features and Rates.

Figure 3. P-Value Between Features and Label

The features that were used are ‘Latitude’, ‘Longitude’, ‘Number of Guests’, ‘Bedrooms’, ‘Rating’.

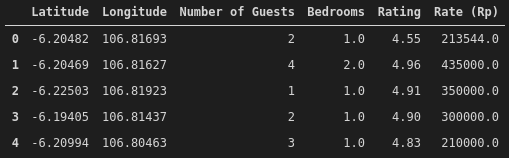


Figure 4. Listings After Feature Selection

Next, we train test split and normalize the data. Then, we check the RMSE value for various K values to get the best K value. Finally, we build the model using the best K value and use it to predict Airbnb rate from a specified host data.

**4. Results**

Our analysis shows that the distribution of the Airbnb rates from the dataset is normal.

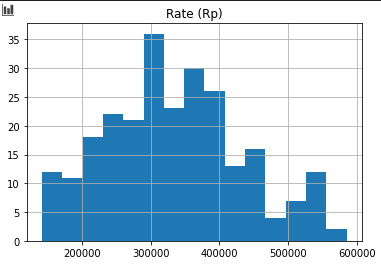


Figure 5. Rate (Rp) Distribution

Because of this, we normalized the features with standard scaler. After we analyzed the RSME for various K values, we found that the best K value is 11. The best rate for the specified host data is Rp 272,000.

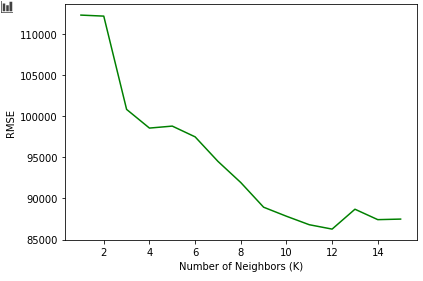


Figure 6. RMSE vs K-Value

**5. Discussion**

We think that the model will become more accurate if we can obtain more dataset.

**6. Conclusion**

The purpose of this project is to predict the best rate for Airbnb host for their listings. Through the implementation of KNN regressor model, we found the best rate to be Rp 272,000.

**7. Reference**

[1] <https://en.wikipedia.org/wiki/Jakarta>