
AMSTERDAM-AIRBNB PRICE PREDICTION

Namrata Paul
Department of Mathematics
Birla Institute Of Technology,Mesra
Ranchi 835215,Jharkhand,India
nampaul1999@gmail.com

Soumyajit Behera
Department of Mathematics
Birla Institute Of Technology,Mesra
Ranchi 835215,Jharkhand,India
ashish454570@gmail.com

Ankit Tewari
Artificial Intelligence Engineer
Knowledge Engineering and Machine Learning Group
ankit.tewari@estudiant.upc.edu

July 6, 2019

ABSTRACT

Price of rental rooms in Airbnb varies according to different filters,so there is a need for a system to predict price according to different features.Price prediction can help the developer determine the price at which host can rent their property. The factors that influence the price are accommodations,number of bathrooms and bedrooms one is booking,number of guests,minimum night stay and where the room is located,whether it is near city centre or far away.The aim of this research is to predict price of property available in Airbnb.

Keywords Price Prediction · KNN regression · More

1 Introduction

The data which is analysed in this project is data of Amsterdam-Airbnb.Airbnb is an American online marketplace and hospitality service brokerage company. Member can use the service to arrange or offer lodging,primarily homestays or tourism experience. Setting a reasonable price for rental property on Airbnb is quite challenging for the hosts to determine the optimal nightly rent price, as well for the customers to get rental according to their preference with an offered price. This project's motive is to develop a price prediction model using KNN regression to tackle this problem.Features that affect the price will be used to predict the price which includes accommodates,number of bathrooms and bedrooms, number of guests to be included,minimum night stay,distance of rental rooms from the city centre.

2 Dataset

In this project we take data of airbnb listings in Amsterdam and try to predict the price of stay in that listing. The source of the dataset: <https://www.kaggle.com/adityadeshpande23/amsterdam-airbnb> The data 15181listings and 13 columns including price,what we are trying to predict.Each row in the data set is a specific listing that's available for renting on airbnb in Amsterdam.Here are some of the more important columns:

- accommodates: Number of guests the rental can accommodate.
- bathrooms: Number of bathrooms included in the rental.
- bedrooms: Number of bedrooms included in the rental.
- guests included: Number of guests

- minimum nights: Minimum number of nights a guest can stay for the rental.
- Number of reviews: Number of reviews the previous guests have left.
- d centre: Distance of rental rooms from the city centre

Some columns will not be used as features such latitude,longitude,instant bookable f, instant bookable t, room type Entire home/apt, room type Private room, room type Shared room.So we are left with 7 columns to process and consider as features. There was no information about this data so we assume that since all the listings are in Amsterdam,the price that we are trying to predict is the general pricing for minimum night stay of the listing in USD,not for specific dates/seasons and not including additional fees, i.e, cleaning and airbnb service fees.

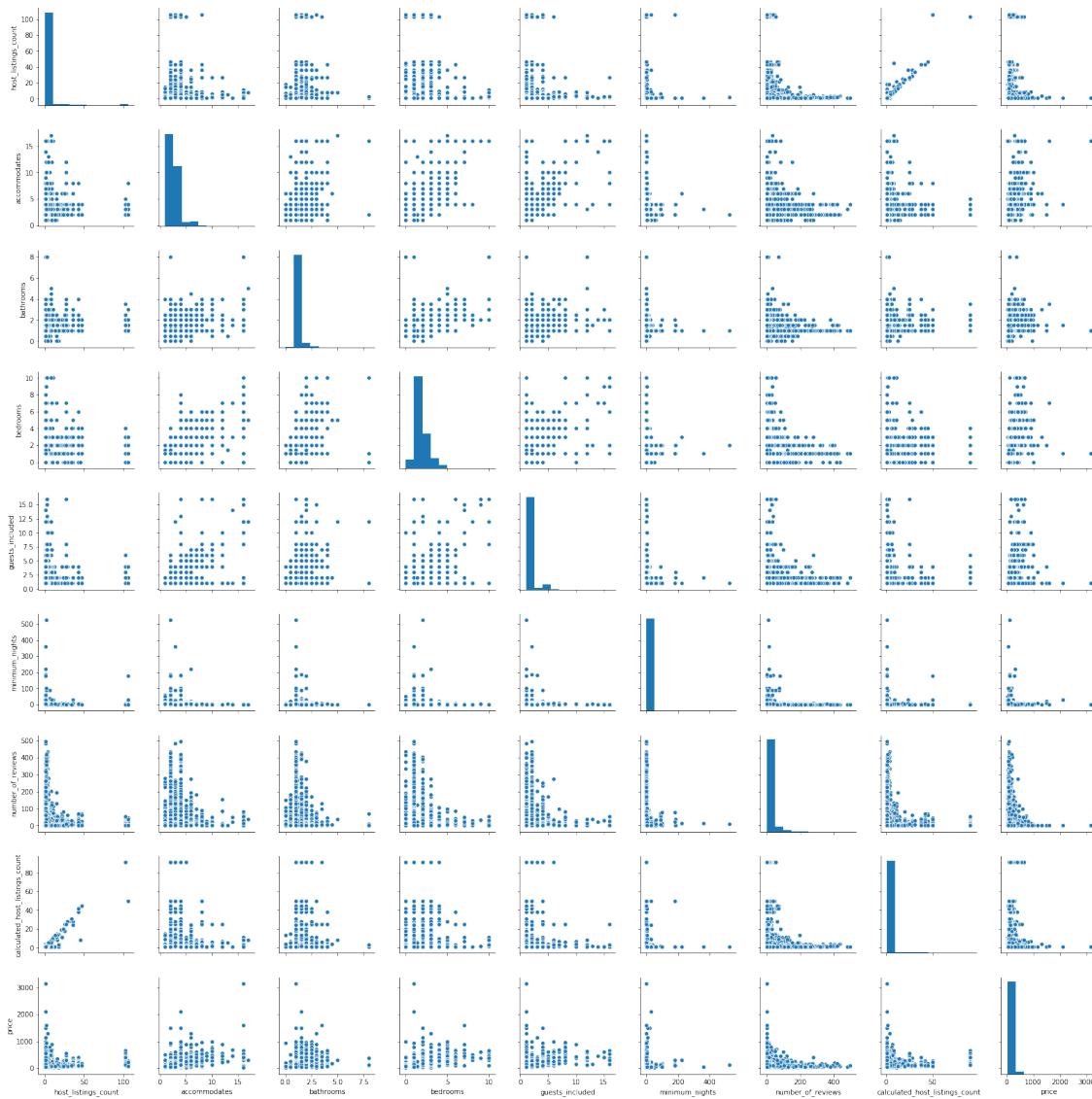


Figure 1: Pairplot

3 Methods and Experiments

3.1 KNN Regression

KNN regression is used on the dataset with all numerical features as model inputs. The model is implemented using scikit-learn library. Scikit-learn library provides a range of supervised learning algorithms via a consistent interface in Python. 3.

3.2 Scikit-learn

We've been writing functions from scratch to train the k-nearest neighbor models. The workflow consists of four main steps:

- Instantiating the specific machine learning model that we want to use.
- Fit the model to the training data.
- Use the model to make predictions.
- Check the accuracy of the predictions.

Now use the rows in the training set to predict the price value for the rows in the test set.

3.3 KNeighborRegressor

Sklearn.neighbors provides functionality for unsupervised and supervised neighbors-based learning methods and supervised neighbors-based learning comes in two flavours: classification for data with discrete labels and regression for data with continuous labels. In this project KNeighborRegressor is being used which implements learning based on the k nearest neighbors of each query point, where k is a user defined integer value. The principle behind nearest neighbors method is to find a predefined number of training samples closest in distance to the new point and predict the label from these. The number of samples being set in this project is a user-defined constant.

3.4 Evaluate the model using r2_score

The model developed in this research will be tested using r2. R2 is statistical measure that represents the proportion of the variance for a dependent variable that's explained by a independent variable(s) in a regression model. If R2 increases, the model gets better.

$$R^2 = \frac{TSS - RSS}{TSS} \quad (1)$$

TSS is the total sum of squared difference of each observation from the overall mean. RSS is the residual sum of squares.

Sklearn.metrics.r2_score is used to calculate R2 for test set and train set. R provides an indication of the goodness of fit of a set of predictions to the actual values. This is a value between 0 and 1 for non fit and perfect fit respectively.

4 Results

In this project we have tried to predict the rental price of Amsterdam airbnb using KNN regression.

Table 1: R2 for train set

Value of k	r2_score
1	1.0
2	0.75197
3	0.66335
4	0.61151
5	0.58466
6	0.56340
7	0.54481
8	0.53381
9	0.52504
10	0.51789

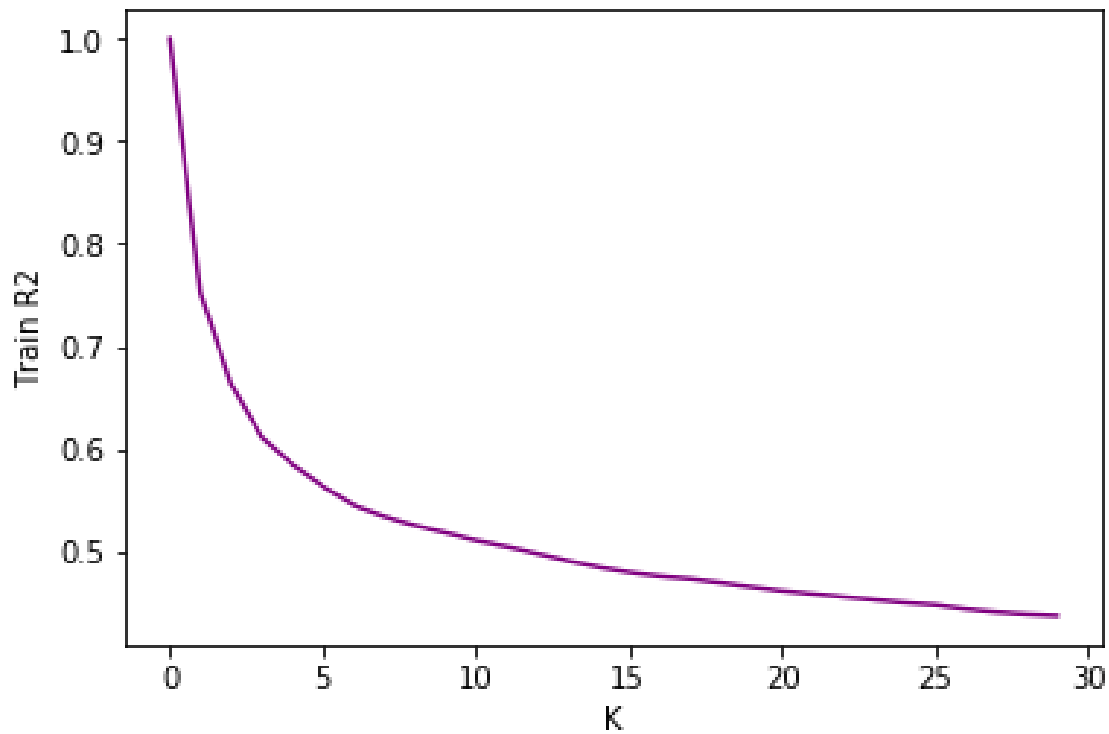


Figure 2: Train set r2_score Vs Neighbors

In figure 2, the relationship between training set and neighbors can be seen. We can see that the r2_score rapidly decreases as the value of k increases.

See awesome Table 1.

The table 1 shows the r2_score for different values of k upto 10. We can clearly see that r2_score is maximum for k=1. We used this k to predict target value in test set and to evaluate r2_score for test test which is found to be 0.017.

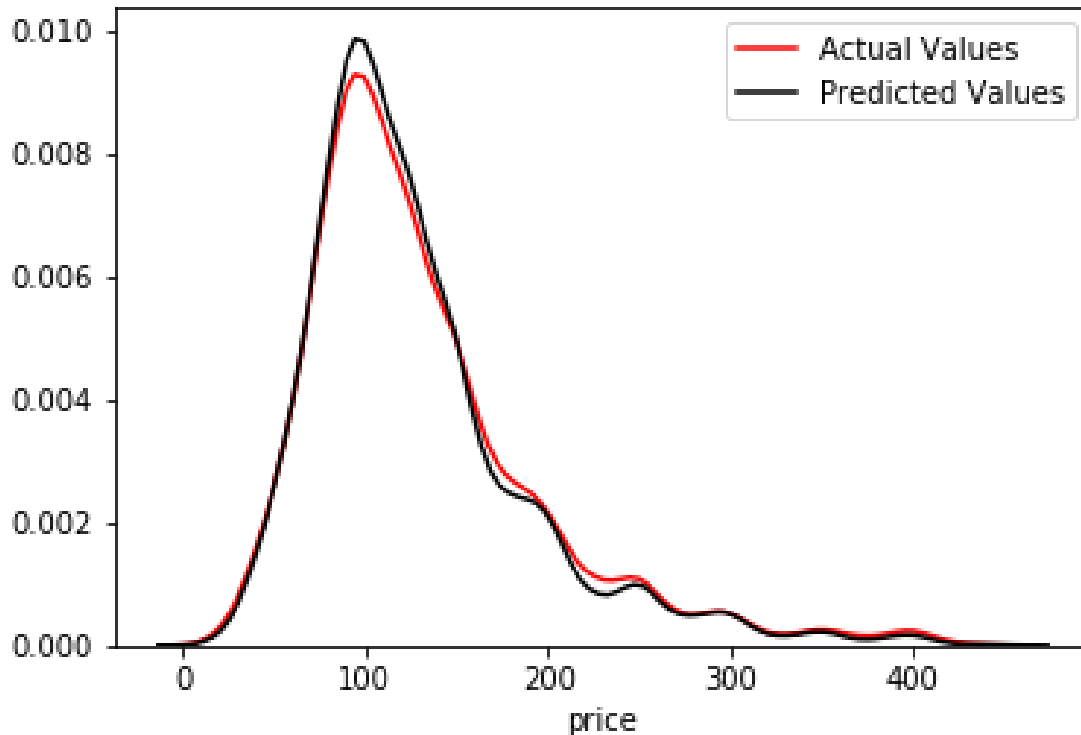


Figure 3: Actual value Vs Predicted value

Figure 3 shows the accuracy of predicted value. The curve for predicted value fits approximately well with the curve of actual value.

5 Conclusions

We have predicted the price of rental properties available in airbnb of Amsterdam using the features which included number of bedrooms, bathrooms, reviews, guests, accommodates and distance from city centre. There are many future applications for this project. New variables can be added to help increase the accuracy of the model. Seasonality can be the next big influencer to be added. As it is obvious that there are months with more traveller than others, but our model do not account for that yet. In future we can predict the price on week to week basis or month wise.

6 Acknowledgement

This is our first project under the guidance of our mentor Ankit Tewari who guided us how to use KNN algorithm to predict the target value. He advised and helped us to overcome with the obstacles and problems we faced during the accomplishment of the task.

References

- [1] An Introduction to Statistical learning: with applications of R (Springer Texts in Statistics)
- [2] Notebook: <https://www.kaggle.com/nampaul/practice>
- [3] <https://necromuralist.github.io/machine-learning-studies/posts/knn-regresion/>
- [4] <https://www.analyticsvidhya.com/blog/2018/08/k-nearest-neighbor-introduction-regression-python/>