

Airbnb Price prediction

Ragib Rohan (170201009)
Nayan- E- Alam (170201018)
Computer Science & Engineering
Bangladesh Army University of Science &
Technology, BAUST

1. Introduction:

Airbnb is a service that lets property owners rent out their spaces to travelers looking for a place to stay. It is the landing platform for providing lodging and tourism experiences. In this platform, the property owners rent out their spaces to travelers looking for a place to stay. Travelers can rent a space for multiple people to share, a shared space with private rooms, or the entire property for themselves. It is based on a peer-to-peer business model. This makes it simple, easy to use and tends to be more profitable for both parties. This model also allows us to customize and personalize our guest's experiences the way we want.

so inside Airbnb's business logic, pricing is probably the most important for hosts and customers. Ensuring fair pricing directly affects booking activities, and also matters to the well-being of the e-commerce environment.

In Airbnb, hosts need guidance when choosing a pricing range when they post property online. Many people will check their neighbor's fees and choose similar prices. people can't decide on the price list because they don't know how much similar bookings cost. Sometimes hosts set prices on external motivations that are not related to the market price on value. thus distorting the list and making it harder for the neighbor's competitors to set price accurately as well.

So in this project, we build a price prediction model of Airbnb listings and make comparisons between different methods. We use dataset from tomslee_airbnb which contains continuous data of Airbnb. Then we use Data Mining methodology to find the main factors influencing a rental price on Airbnb.

2. Relevant Work:

Prior works on rental price prediction based on Airbnb data are deficient in terms of evaluation metrics and performance. Tang and Sanghani [2015] work on the task of price prediction for San Francisco Airbnb listings[12]. They turn the regression problem into a binary classification problem by splitting the price according to the median, which effectively reduces the difficulty

of the task. In more recent work, Kalehbasti et al. [2019] work on price regression for Airbnb listings in New York[13]. They use a range of methods including tree-based models, SVR, KMC, NN, etc and integrate sentiment analysis into their model. While they claim to achieve the highest R2 of 0.7246, they evaluate their metrics (R2 and MSE) on the logarithmic scale of price instead of the original scale.

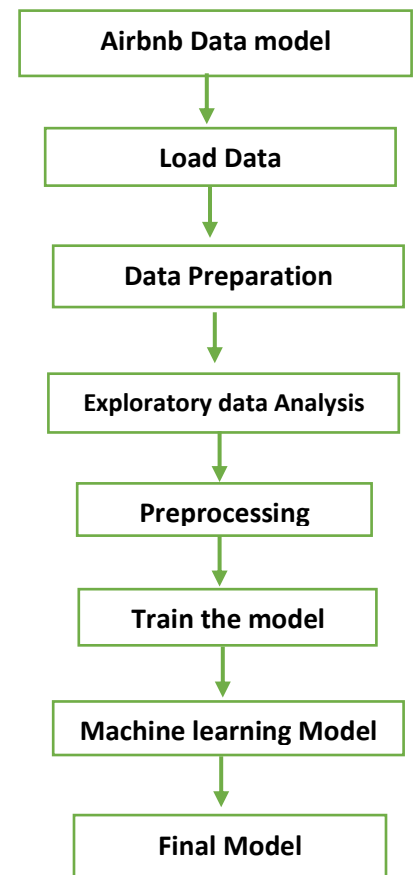
We done our paper following the paper of Airbnb price prediction done by Tiancheng Cai, Kevin Hen, Han Wu of Standard University.

In our project, we work on the original price regression problem by using linear regression.

3. Dataset:

In our project, we use Airbnb datasets for January 21 and June 19 of 2017. In traditional machine learning methods, we use the Linear Regression, lasso and Ridge method in our project.

1. First, we load the dataset and analyze it, and identified that which columns has no value and which columns have the data we need for our project.
2. then we drop the columns that we don't need. after dropping the unnecessary columns, we select the columns we do need for our project and complete our data preparation.
3. then we analyze the exploratory data from the dataset and plot price according to the room type. and we found that maximum customer prefers to get entire home/apartment, then private room and the least number of people get shared room.
4. Then we measure the longitude and latitude according to the price.
5. After analyzing that we create two tables. one with review data and the other is without review data for training the dataset.
6. after training the dataset we go to the machine learning part where we test the accuracy of our dataset.
7. then we test out the machine learning model and compare the accuracy of two datasets we take for our project.



4. Methods:

in this section, we describe the machine learning model that we used to predict.

For prediction, we use Linear regression machine learning model, Lasso regression analysis technique, Ridge regression technique and Elastic.

Linear regression:

Linear regression is a linear approach for modeling the relationship between a scalar response and one or more explanatory variables. In this model, the relationships are modeled using linear prediction functions.

Lasso regression model:

The full form of lasso is least absolute shrinkage and selection operator. It is a regression analysis technique that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model. The main purpose is the feature selection and regularization of data models.

Ridge regression model:

Ridge regression is the method used for the analysis of multicollinearity in multiple regression data. It is used to deal with overfitting and when the dataset is large.

Elastic:

Elastic is a supervised machine learning technique that helps us to find patterns in our data.

5. Experiments and results:

For experimental purpose we use 2 datasets of Airbnb to see their differences of predicting accuracy level.

For dataset 1:

We use the dataset of Airbnb of 21 January 2017.

After training the model we found the accuracy level of:

34% accuracy for Linear regression.

34% accuracy for Ridge regression.

35% accuracy for Lasso regression.

```
LinearRegressionmrse_train:36.71146273111367, mrse_test: 33.68425026394805  
Ridgемrse_train:36.725396947226834, mrse_test: 33.659973639898226best alpha: 100.0  
Lassomrse_train:36.71443387317868, mrse_test: 33.73228546652151best alpha: 0.1
```

For dataset 2:

We use the dataset of Airbnb of 21 January 2017.

After training the model we found the accuracy level of:

35% accuracy for Linear regression.

LinearRegressionmrse_train:2.372076603992182e-13, mrse_test: 35.492582137110276

35% accuracy for Ridge regression.

Ridgemrse_train:0.00021255685194401948, mrse_test: 35.24123145923182best alpha: 0.01

34% accuracy for Lasso regression.

Lassomrse_train:24.274848204891192, mrse_test: 34.220609708348015best alpha: 1.0

6. Discussion and Future Work:

At this time we use three machine learning model for our project.

In the future, we will apply other machine learning models to achieve more accuracy levels in our model and we will identify the best machine learning model for it.

we will work on more data and test this model more and more to achieve a higher accuracy we can.

7. Conclusion:

This project will provide an accurate model for price prediction to help hosts to set the price correctly when they want to post up their house for the online rental market.

We will try our best to improve it for further work.

8. References:

[1]Melbourne airbnb open data. <https://www.kaggle.com/tylerx/melbourne-airbnb-open-data>, 2018

[2]L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux. API design for machine learning software: experiences from the scikit-learn project. In ECML PKDD Workshop: Languages for Data Mining and Machine Learning, pages 108–122, 2013.

[3]F. Chollet et al. Keras. <https://keras.io>, 2015.

[4]Furuta, 2018. URL <https://gist.github.com/furuta/eb2453a0ad13ca51b780caef5f3c95c7>.

[5]T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning. Springer-Verlag, New York, 2009.

[6]P. R. Kalehbasti, L. Nikolenko, and H. Rezaei. Airbnb price prediction using machine learning and sentiment analysis, 2019.

[7]D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014

- [8]J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, 2014. URL <http://www.aclweb.org/anthology/D14-1162>.
- [9]S. Robinson. Predicting the price of wine with the keras functional api and tensorflow, 2018. URL <https://medium.com/tensorflow/predicting-the-price-of-wine-with-the-keras-functional-api-and-tensorflowa95d1c2c1b0>
- [10]E. Tang and K. Sangani. Neighborhood and price prediction for san francisco airbnb listings. CS 229 Final Project Report, 2015.
- [11]R. Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1):267–288, 1996. ISSN 00359246. URL <http://www.jstor.org/stable/2346178>.
- [12]Tang and Sanghani [2015] work on the task of price prediction for San Francisco Airbnb listings.
- [13] Kalehbasti et al. [2019] work on price regression for Airbnb listings in New York