ABES ENGINEERING COLLEGE,UNIVERSITY OF LUCKHNOW,GHAZIABAD,INDIA

**IMAGE CAPTION GENERATOR USING DEEP LEARNING**

Asst. Professor of Department Of Computer Science Data Science

ABES Engineering College , Ghaziabad

-------------------------@abes.ac.in

Somesh Verma

Department Of Computer Science Data science

ABES Engineering College , Ghaziabad

[Pr910647@gmail.com](mailto:Pr910647@gmail.com)

Priyam Rastogi

Department Of Computer Science Data science

ABES Engineering College , Ghaziabad

[Pri](mailto:Pr910647@gmail.com)yam.20b1541026@abes.ac.in

Prakhar Tiwari

Department Of Computer Science Data science

ABES Engineering College , Ghaziabad

[Prakhar.20B1541037@abes.ac.in](mailto:Prakhar.20B1541037@abes.ac.in)

**Abstract**

In this project, we worked on datasets to provide a predictive price of a airbnd service provider. **For those who don't know it yet** for some **reason,** Airbnb is **an Internet** marketplace for short-term **rentals of homes** and **apartments. For example,** you **can** rent out **(advertise)** your home for **his** week or rent out **an** empty **bedroom while you are away.** One **of the challenges** Airbnb hosts face is determining the **best rental price per night.** In many **locations,** renters (hosts) **will see curated** listings **that** can **be filtered** by price, number of bedrooms, room type, and **other criteria.** Airbnb is a **marketplace, so** the amount a host can charge is ultimately tied to **the** market **price.**

**Introduction**

**Data Set**

**Data Cleaning**

**Data Analysis**

**Modelling**

**Conclusion**

1. Introduction-

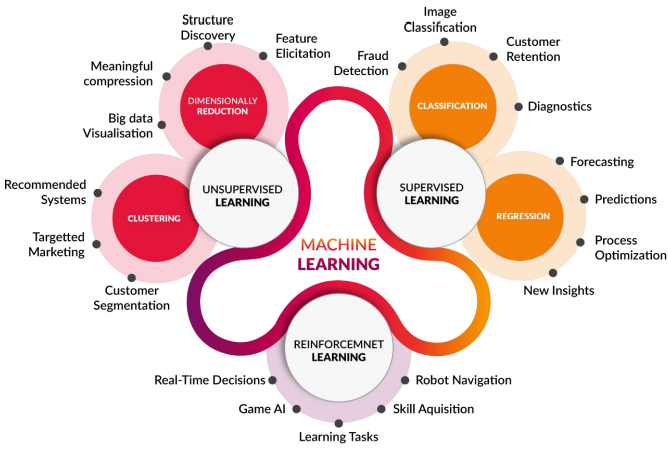
We used different algorithms to get a accuracy in our price prediction. Pricing a rental property on Airbnb is a daunting task for owners as the location has

customers. Customers, on the other hand, should evaluate the offered price with a minimum of knowledge of the property's optimal value. This white paper aims to develop reliable price prediction models using machine learning, deep learning and natural language processing techniques. Rental characteristics, owner characteristics, and

customer ratings include linear regression to tree-based

models, support vector regression (SVR), k-means clustering (KMC), and neural networks (NN)

Includes predictors and different methods. Used to build predictive models.



1. **Problem Statement-**

Airbnb provides general guidance to hosts, but there is no readily available method for determining the best price for renting a space.Third-party software exists, but it is expensive.

One way is to find several properties similar to rental housing, average the property prices, and set the price to the calculated average price. However, since the market is dynamic, prices need to be updated frequently and this method can become tedious.

Other problems? We may miss the competitive advantage our listings have over neighboring listings. B. Proximity to grocery stores, pubs, and other additional services, or how good my photography is.

So in this project we will try to predict the price using multiple listing functions. Additionally, we'll add a space-based predictor variable (property's proximity to a particular venue). This allows the model to implicitly price things like living near a bar, pub, or supermarket.

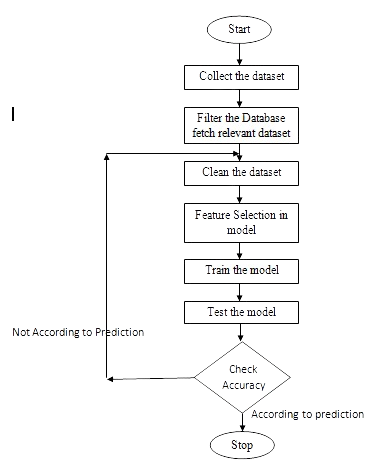
1. **METHODOLOGY-**

The problem statement is to predict the price of Airbnb house based on multiple factors like locality, reviews per month, availability, room type etc .

We implement different machine learning algorithms like ridge , lasso, decision tree etc. to predict the price.

We created a flask web app which takes input as neighbourhood group, room type availability and show predicted result on screen.

We performed exploratory data analysis to analyse (univariate and bivariate) the data set and summarize the data



Linear regression using the entire feature set as model input was used as a base model for to assess the performance of other methods. After selecting a set of features using Lasso Feature Selection , we went through several machine learning models to find the best one.

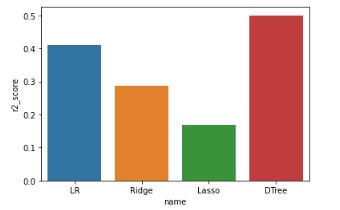
**Ridge Regression**

Linear Regression with L2 regularization adds a penalizing term to the squared error cost function in order to help the algorithm converge for linearly separable data and reduce overfitting. Therefore, Ridge Regression minimizes J(θ) = ||y − Xθ||2 2 + α||θ||2 2 with respect to θ, where X is a design matrix and α is a hyperparameter. Since the baseline models were observed to have high variance, Ridge Regression seemed to be an appropriate choice to solve the issue.

**Decision Tree Classification Algorithm**

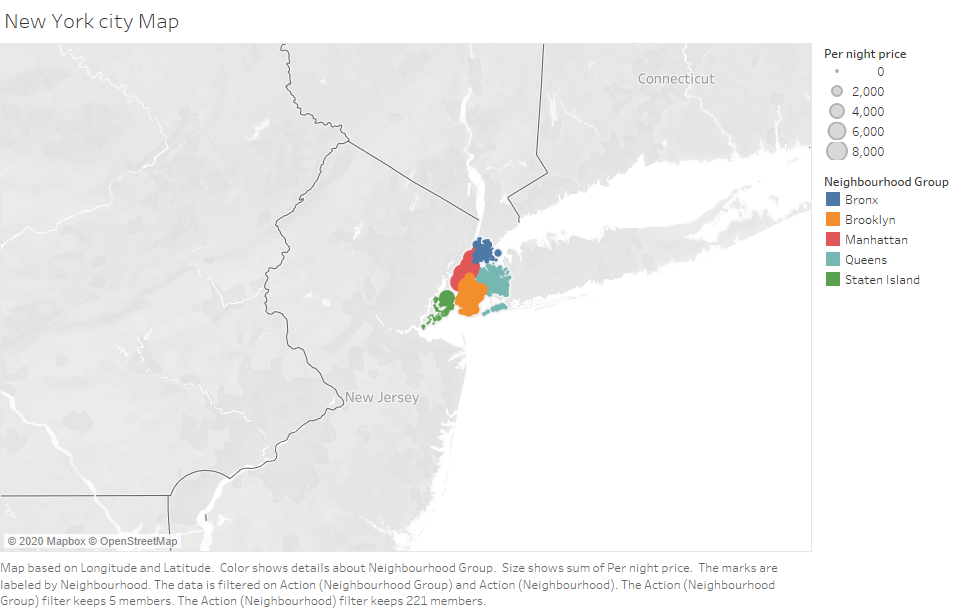
* Decision Tree is a **Supervised learning technique**that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where**internal nodes represent the features of a dataset, branches represent the decision rules** and **each leaf node represents the outcome.**

**Lasso Regression-**

Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination. 

Some of the more important features this project will look into are the following:

* accommodates: the number of guests the rental can accommodate
* bedrooms: number of bedrooms included in the rental
* bathrooms: number of bathrooms included in the rental
* beds: number of beds included in the rental
* price: nightly price for the rental
* minimum\_nights: minimum number of nights a guest can stay for the rental
* maximum\_nights: maximum number of nights a guest can stay for the rental
* number\_of\_reviews: number of reviews that previous guests have left



1. **Conclusion and Future scope-**
2. Through this research we are able to understand the 2019 data and the price distribution of the rental properties.
3. Avg. Price of Airbnb house in Manhatten is 31% which is maximum among all the neighbourhood group. Bronex is the cheapest as its contribution in pi chart is 14%. Most of the persons prefers Entire Room/ Apartment and very less people prefers Shared Room to stay.
4. Average price of Entire Room/Apartment is 160 which is twice of the price of Private Room.
5. Further the average price in both the cases is below 200 $ which shows that in an average the properties are quite affordable.
6. We can also add some state and location specific features to improve the model.
7. For future works, we expect the price prediction model to be improved using a larger dataset with balanced customer reviews since the mean squared error in the Neural Network model is still decreasing at the end of the training phase. Additionally, public Airbnb datasets contain more positive reviews than negative reviews. A well balanced dataset should be helpful to build a more accurate price prediction model. Other than customer reviews, historical prices might be another key factor to evaluate the price. Customers expect a lower price if the price is constantly decreasing.
8. We would like to know if the model performs the same using datasets for different cities.

References-

1. Luo, Y. et al. “Predicting Airbnb Listing Price Across Different Cities.” (2019).
2. Tang, E.. “Neighborhood and Price Prediction for San Francisco Airbnb Listings.” (2015).
3. Nguyen, QuangTrung. “QuangTrungNguyen/Airbnb-Pricing-Prediction.” GitHub, github.com/QuangTrungNguyen/Airbnb-pricing-prediction.
4. Cmdline. “Linear Regression Using Matrix Multiplication in Python Using NumPy.” Python and R Tips, 18 Mar. 2020, cmdlinetips.com/2020/03/linearregression-using-matrix-multiplication-in-python-using-numpy/.
5. Breiman, L. (June 1997). "Arcing The Edge". Technical Report 486. Statistics Department, University of California, Berkeley.
6. Mason, L.; Baxter, J.; Bartlett, P. L.; Frean, Marcus (1999). "Boosting Algorithms as Gradient Descent" (PDF). In S.A. Solla and T.K. Leen and K. Müller (ed.). Advances in Neural Information Processing Systems 12. MIT Press. pp. 512–518.
7. Friedman, Jerome. (2000). Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics. 29. 10.1214/aos/1013203451.
8. Kennedy, Peter (2003). A Guide to Econometrics (Fifth ed.). Cambridge: The MIT Press. pp. 205–206. ISBN 0-262-61183-X.
9. An Introduction to Support Vector Regression, towardsdatascience.com/anintroduction-to-support-vector-regression-svr-a3ebc 1672c2.