

AIRBNB PRICE PREDICTION

MINI PROJECT – II

SYNOPSIS



Department of Computer Science & Application

Institute of Engineering & Technology

SUBMITTED TO: -

Mr. Amir Khan
(Technical Trainer)

SUBMITTED BY: -

Yash Kumar Gupta (201500820)
Vijay Kaushal (201500783)

Acknowledgement

It gives us a great sense of pleasure to present the synopsis of the B.Tech mini project undertaken during B-Tech III Year. This project is going to be an acknowledgement to the inspiration, drive and technical assistance will be contributed to it by many individuals. We owe special debt of gratitude to Mr. Amir Khan Sir, Technical Trainer, for providing us with an encouraging platform to develop this project, which thus helped us in shaping our abilities towards a constructive goal and for his constant support and guidance to our work.

His sincerity, thoroughness and perseverance has been a constant source of inspiration for us. We believe that he will shower us with all his extensively experienced ideas and insightful comments at different stages of the project & also taught us about the latest industry-oriented technologies. We also do not like miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind guidance and co-operation.

Signature of Students: -

1st – Yash Kumar Gupta
201500820

2nd- Vijay Kaushal
201500783

ABSTRACT

Figuring out the price of a listed Airbnb rental is an important and difficult task for both the host and the customer. For the former, it can enable them to set a reasonable price without compromising on their profits. For the customer, it helps understand the key drivers for price and also provides them with similarly priced places. This price prediction regression task can also have multiple downstream uses, such as in recommendation of similar rentals based on price. We propose to use geolocation, temporal, visual and natural language features to create a reliable and accurate price prediction algorithm.

Contents

Acknowledgement

Abstract

1. Introduction

1.1 Objective

1.2 Motivation

1.3 Problem Statement

2. Software Requirement

2.1 Hardware Requirements

2.2 Software Requirements

3. Project Description

4. Working and Implementation

5. Result and Conclusion

6. References

INTRODUCTION

Airbnb has hosted over 60 million people in 34,000 cities across the world and is continuing to grow quickly. Airbnb provides a convenient source of income for people who have otherwise vacant space and for guests looking for affordable and convenient housing options. With any service, trying to monitor and understand the underlying pricing dynamics of the Airbnb market is very important both for hosts and guests. As users continue to grow on both the supply and demand side, homeowners may find it hard to properly price their property. Airbnb has recognized this and conducted considerable research into suggesting pricing from a [supply side standpoint](#).

We seek to analyze over 27,000 listings in the NYC area in order to better understand how the use of listing attributes such as bedrooms, location, ratings, and more can be used to accurately predict the optimal listing price both for the host and guest. Holiday and seasonality is another useful component that can attract more customers and drive higher prices, but it is unclear how much of a premium one should pay per holiday. With better price suggestion estimates, Airbnb home providers can reach an equilibrium price that optimizes profit and affordability. The objective of this project is to build a model that predicts the optimal price of a property taking into account listing features and seasonality. The end goal is so users can understand what features of an AirBnB listing are most important as well as how prices should be fluctuating based on seasonality

Project Objective

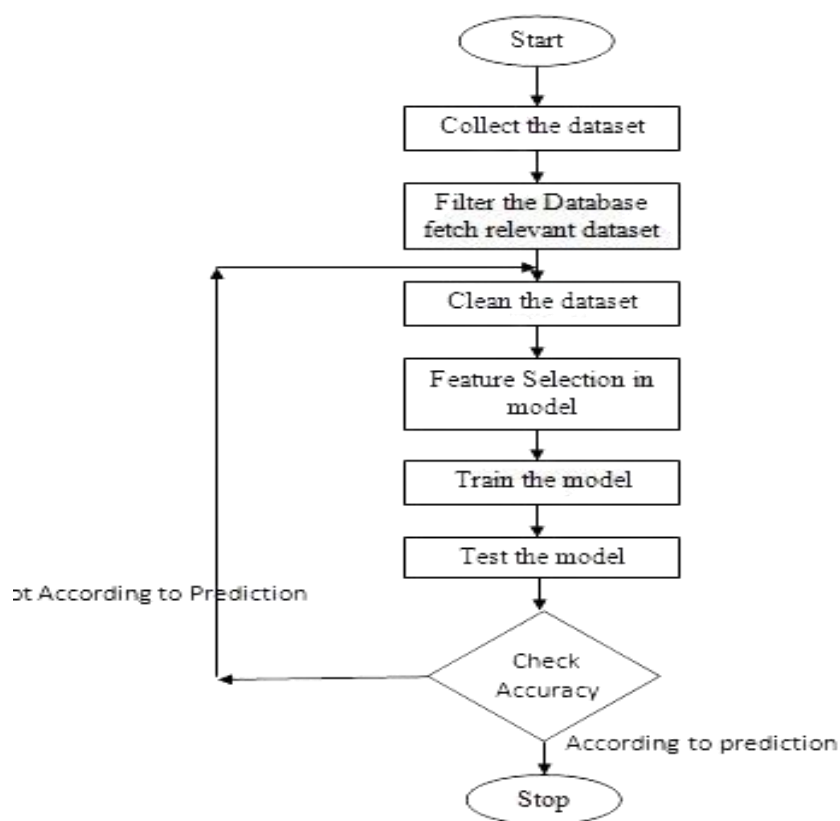
- ♦ The main aim of the research was to predict the price of Airbnb house based on multiple factors like locality, reviews per month, availability, room type etc.
- ♦ To find if there exists a relationship between the name of the property and the price of the properties.
- ♦ To use different ML model and find the best one. This project will analyze different Machine Learning Algorithms and finds the one with best accuracy.
- ♦

SOFTWARE AND HARDWARE REQUIREMENTS

- Anaconda3/ Jupyter
- Reference Book
- Kaggle
- MS Office
- Laptop Device
- 4 GB RAM or above
- Window 10 or above

PROJECT DESCRIPTION

The purpose of this project is to develop multiple algorithm for predicting the price of one stay, two stay or multistay hotels using AIRBNB dataset of year 2015 and finding the best algorithm to predict this by using machine-learning technology. It allows for flexible data format and deliver of its data so that each analysis application can receive only the information it needs and, in the format, required.



Work Flow Diagram

WORKING and IMPLEMENTATION

The proposed methodology related to our project is given below:

Step 1: Data extraction :

The dataset is taken from Kaggle, it is named as Airbnb New York dataset.

Step 2: Data preprocessing :

In this part, first we need to understand the attributes and behavior of data like, how many columns are there, what are the data types and how much data is missing? We treated missing values, then filled mean values in place of null in some columns and simply remove the whole row in some column. Also remove insignificant columns from the dataset. It involves the following steps:

- Removal of null values.
- Removal of links, gifs, emoji, images and special characters.
- Removal of missing data.
- Removal of non-English words.

Step 3: Exploratory data analysis :

It is an approach to analyse the dataset and summarize the data and find patterns in it. We had also analysed categorical and numeric features by univariate and bivariate analysis.

Step 4: Feature engineering :

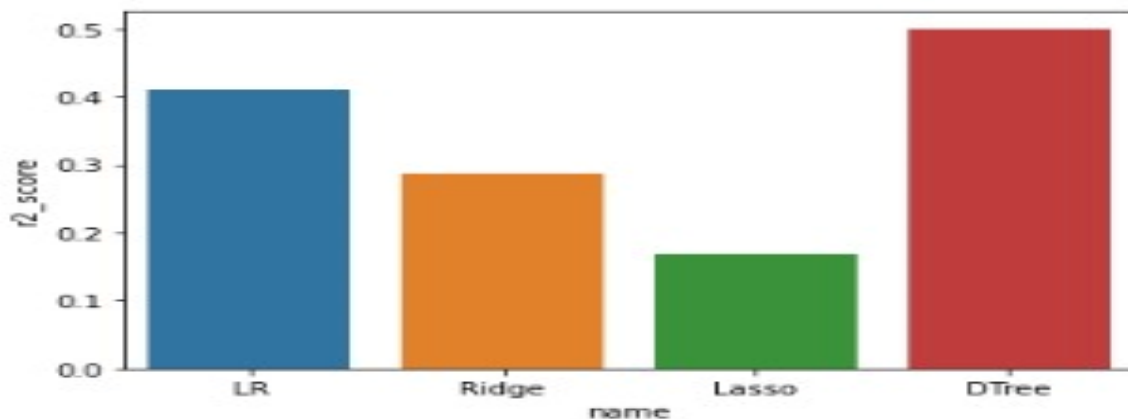
It is a process of using domain knowledge to extract feature from raw data. These features are used to improve the performance of machine learning model. In this step converted categorical features like neighbourhood group, root type and neighbourhood into numerical features because these are not processed in machine learning. Also, I had to perform log transformation on Price attribute because it is left skewed. Giving the step 3 output in different

machine learning algorithms and analyze it to find the algorithm with best accuracy.

Step 5: Model building :

In this step, I had applied Linear Regression, Ridge and Lasso Regression and Decision Tree on pre-processed data. I found Decision Tree for predictive model with the help of R-squared score because its accuracy is the higher than all other models.

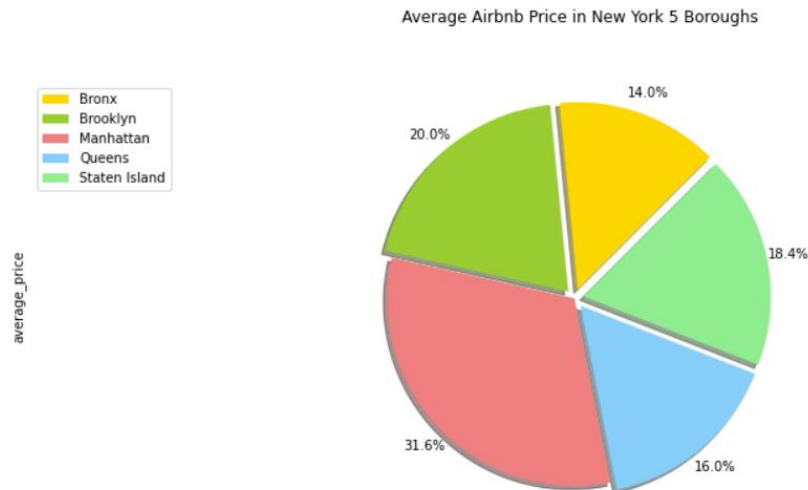
RESULT and CONCLUSION



From the above bar chart, it is clearly seen that Decision Tree gives the best output or it is giving best result with accuracy of 0.78 or 78%.

Some concluded points are as follows:-

- ♦ Avg. Price of Airbnb house in Manhattan is 31% which is maximum among all the neighbourhood group. Bronx 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. Pi chart is depicted below:-



REFERENCES

Books and Paper:

- Digital Image Processing by Rafael C. **Gonzalez**
- Life 3.0_ Being Human in the Age of Artificial Intelligence • DIP by Sridhar, S.
- Paper:- Luo, Y. et al. “Predicting Airbnb Listing Price Across Different Cities.” (2019).

Websites

- www.tensorflow.com
- <https://app.supervised.ly>
- www.deeplearningAI.com
- www.youtube.com
- www.kaggle.com

Faculty Guidelines:

Mr. Amir Khan Sir (Technical Trainer in GLA University)

GitHub Repository link:

<https://github.com/Yash-Kumar-Gupta-0845/Mini-Project-2>