International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET229228

A Survey On Price Prediction Model for Airbnb listing using Machine Learning

Prajapati Pradip[1] , Prof. Monali Suthar[2]

[*1] M.Tech Computer Engineering Department, Silver Oak College of Engineering and Technology-Ahmedabad, Gujarat, India

[2]Assistant professor ,Computer Engineering Department , Silver Oak College Of Engineering And Technology, Ahmedabad, Gujarat, India

Article Info
Volume 9, Issue 2
Page Number : 167-171

Publication Issue :
March-April-2022

Article History
Accepted : 20 March 2022
Published: 30 March 2022

ABSTRACT

People who rent their homes through Airbnb find it hard to figure out how much they should charge for them. They have to figure out how many people they can put in each home. Other than that, customers have to figure out the price of a home or apartment based on what they know about that home or apartment and what it has to offer them. A price prediction model is used with machine learning algorithms to add value for both the customer and the person who owns it in this study. To figure out the right price, I'll look at different features of the rentals and the traits of the owner. To make better predictions, I will also look at the reviews of different customers. To do this, I will use a lot of different techniques, like linear regression, tree-based models and the SVM technique.

Keywords: Machine Learning, LASSO Regression, MSE, Tunning, Prediction, Gradient Boosting Algorithm

I. INTRODUCTION

Home owners can list their residences for rent on the website of the home-sharing company Airbnb. Hosts are free to set their own rates for their listings. A vast range of data points offered by Airbnb and other sites now makes it impossible for hosts to appropriately price their apartments.

A "base price" must still be entered into any third-party pricing software, even if you utilise it. The algorithm will adjust the daily rate to be as close as possible to this "base price" based on the day of week, seasonality and distance from the future date.

Because of the intense competition and the potential impact of even little price variations, Airbnb pricing is essential in cities like London. Too low a price and you won't receive any bookings, which is a tricky one to get right. If you under-price your product, you'll miss out on a lot of sales.

II. LITERATURE STUDY

Owners of various properties can use the Airbnb web app to list their properties for rent. Many academics are

drawn to the rental business because of the importance of determining the value of a particular property. [1].

To accurately price a property, a variety of third-party pricing simulators are available to various sellers, allowing them to accurately anticipate a property's typical market worth. It is possible for the owner of the property to set a price for their property that is convenient for them as well as for the host of the property. With Airbnb, you may set the pricing of your property in a variety of ways, including per-day prices, weekly rates, and limits on how long guests can stay. [2].

Researchers have used a variety of methods to arrive at their estimates of the property's value. Researchers employed a regression method to accurately forecast the pricing based on the desired cost of a night's stay and other binary classification algorithms to determine the price. Various reasoning techniques were used to determine the final cost estimates [2].

Various value estimate techniques based on geographical weighted regression are used to evaluate issues with value and can find variables according to the listing cost of the model. On the typical income expansion, Airbnb presents a novel solution to the interest bend that assists the evaluation technique in a greater degree [3].

Machine learning algorithms, such as the random forest regressor, linear regression, and decision tree regressor, are used to better predict the price. For forecasting as well as averaging the errors in terms of the fitness, machine learning algorithms are useful. They also provide a variety of correlations between distinct variables [4].

The frequency stability assessment provided by the ELM approach is beneficial for both system prediction and modelling [5], Sales forecasting and power system evaluation can also benefit from this software tool [6].

When it comes to predicting prices, the ELM model is no longer widely used. Recently, the ELM model and a single-layer feedforward network have been shown to be more accurate at predicting future price fluctuations. ELM models employ rapid learning and generalisation tactics to improve their capacity to anticipate outcomes. [7].

A. Linear Regression

Using a predictor and a response variable, linear regression can assist describe the relationship between two continuous variables. Independent of the response variable, the predictor variable is a factor. If one variable can be stated by another, the relationship is considered to be deterministic. However, statistical relationships, such as the one between height and weight, rarely operate this way.

The best fit line shows that the total error (the distance between the data points and the regression line) in the forecast is too minimal when considering all the data points in the regression model.

The more examination and observation models there are, the more accurate the results will be. Before applying the gradient technique to the regression line, the model needs at least a four- to five-field visual field for the best accuracy of prediction.

Because the linear regression can only determine the link between the dependent variable's mean and its independent factors, it has a severe drawback. However, it is impossible to describe the relationship's whole characteristics using a single variable definition.

B. Gradient Boosting Algorithm

Boosting is a technique that can be used to improve the performance of weak students. For example, each new tree is defined as an updated version of data from the previous one. Predictive models can be built using gradient boosting techniques. In order to minimise the prediction error, this method relies on the best possible combination of the prior models.

C. Recursive Feature Extraction

The recursive feature elimination approach is an effective strategy for selecting features from the training dataset and also an excellent method for predicting the target variable in the dataset.

The RFE approach uses the training dataset to conduct the search and removes the unwanted features iteratively until it obtains the desired features. A machine learning algorithm is fitted to the data based on the ranking of the features and undesired characteristics are discarded during the process.

D. Wrapper and Embedded Methods

As a black box, the learning machine provides a variety of software packages that may be used to do a wide range of tasks. The machine learning for predictive performance is included in the wrapper methods to access the subset of variables from the dataset.

It is necessary to employ other search techniques such as Greedy search in order to reduce the computational time required by the wrapper approach because it uses a "brute-force" method for searching.

E. LASSO Regularization

A simple lasso regression can be used to select features from a wide variety of datasets. We only need to look at features with coefficients that are different from zero in order to select the proper Lasso regression.

III COMPARATIVE STUDY

Title	Journal Name and Year	Methods	Limitations
Do Airbnb host listing attributes influence room pricing homogenously ?	International Journal of Hospitality Management, 2019	Random forest, Decision tree	This means that even a little change in the data can have a significant impact on the optimal decision tree's shape. They tend to be off by a wide margin. Many other predictors do better with the same data as you do.
An Extreme Learning Machine Model Approach on Airbnb Base Price Prediction	International Journal of Advanced Computer Science and Applications, Vol. 11, No. 11, 2020	single hidden layer feedforward neural network (SLFNs), root mean squared error, mean absolute percentage error (MAPE).	Still, the procedure of determining the number of hidden layers and the number of neurons in each hidden layer remains baffling to many scientists. When the actual values are zero or close to zero, it generates infinite or undefined values.
Hybrid-Recursive Feature Elimination for Efficient	MDPI	Recursive Feature Elimination	Due to their ability to train and test in feature space wrapper tend to do better when it comes to picking features.

Predicting Airbnb Listing Price Across NewYork.	Research Gate 2020	Linear Regression, XGBoost.	A linear relationship between the dependent and independent variables is taken as a given. Noise and overfitting are common problems Often necessitates a large number of trees (more than 1000), which can be time and memory consuming.
Price Prediction in the Sharing Economy: A Case Study with Airbnb data	University of New Hampshire Scholars' Repository, 2020	Mean absolute percentage error (MAPE), Anova test.	When the actual values are zero or close to zero, it generates infinite or undefined values. It is capable of reducing the prevalence of type I errors in general
Mathematical modeling of microwave dried celery leaves and determination of the effective moisture diffusivities and activation energy	Food and Science technology.	Extreme Learning Machine	Calculations based on weight must be made.
Airbnb Price Prediction in the Age of Social Distancing	UNIVERSITY OF CALIFORNIA, 2021	Random forest, Linear Regression, XGBoost	An overly large number of trees in a random forest algorithm can slow it down, making it unsuitable for real-time forecasts. For the most part, these algorithms are quick to learn but they take a long time to provide predictions once they have been trained.

III. OBJECTIVES OF THE RESEARCH

The main objectives to design a proposed system are given below

1) To choose the features for the model, you can use an Embedded method like lasso. These methods combine the advantages of both the wrapper (forward feature elimination) and the filter (variance threshold) methods. They include interactions of features but keep the computational cost low. In this case, embedded methods are iterative in that they take care

of each step of the model training process and pick out the features that are most important for that step. Lasso (Least Absolute Shrinkage and Selection Operator) is used to pick the feature.

2) To address data overfitting via boosting and bagging, as bagging seeks to minimise the likelihood of overfitting complex models.

- It simultaneously instructs a huge number of "strong" learners.
- A powerful learner is an unconstrained model.
- Bagging then aggregates all the strong learners' predictions in order to "smooth" them out.
- Increasing efforts to enhance the predictive versatility of simple models.
- It sequentially instructs a large number of "weak" learners.
- A model that is unable to learn is said to be a weak learner (i.e., you could limit the max depth of each decision tree).
- Each one in the sequence focuses on the importance of learning from the previous one's mistakes.
- While both bagging and boosting are ensemble methods, they tackle the problem in diametrically opposed ways.

3) To reduce model error, gradient boosting should be employed, as it sets the target outcomes for the next model to choose the best from, as well as creating a collection of predictors and handling outliers.

4) To determine the model's performance, we will use Adjusted R-Squared, which is a corrective for adding too many terms to the model. It will always be less than R-squared and is therefore a preferable option. However, it shares many of the same flaws as plain old r2, and it is also non-predictive, dealing only with the data you feed it.

IV. CONCLUSION

The usage of data processing and feature selection is a critical task when developing any predictive model, as it directly influences the model's overall accuracy. While developing the proposal, many approaches and machine learning algorithms are taken into account.

Vendors can quickly build marketing strategies and price configurations for multiple assets by leveraging past data from various clients. Machine learning algorithms assist in determining the pricing of various properties based on the consumers' accessible data.

V. REFERENCES

[1]. Manojit, "Do airbnb host listing attributes influence room pricing homogenously?," S. K. M., pp. 1-8, 2019.

[2]. Priambodo, "An Extreme Learning Machine Model Approach on Airbnb Base Price Prediction..," International Journal of Advanced Computer Science and Applications, vol. 10, p. 11, 2020.

[3]. Zhang, "Key Factors Affecting the Price of Airbnb Listings.," MDPI, pp. 1-11, 2017.

[4]. Zulkifley, "House Price Prediction using a Machine Learning Model: A Survey of Literature..," International Journal of Modern Education and Computer Science (IJMECS), pp. 46-54, 2020.

[5]. Xu, "Extreme learning machine-based predictor for real-time frequency stability assessment of electric power systems.," Neural Computing and Applications, pp. 3-4, 2013.

[6]. Alibas, "Mathematical modeling of microwave dried celery leaves and determination of the effective moisture diffusivities and activation energy.," Food and Science technology., pp. 22-26, 2012.

[7]. Huang, "Extreme learning machines: a survey.," International Journal of Machine Learning and Cybernetics volume, pp. 107-222, 2011.

Cite this article as :

Prajapati Pradip , Prof. Monali Suthar, "A Survey On Price Prediction Model for Airbnb listing using Machine Learning ", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 2, pp. 167-171, March-April 2022. Available at doi : https://doi.org/10.32628/IJSRSET229228

Journal URL : https://ijsrset.com/IJSRSET229228