HOUSE RENT PRICE FORECASTING USING MACHINE LEARNING

A.Annapurna Asst Professor,

Humera Fathima, Y.Ramya Sree , N.Bala Yaswanth, K.Deekshitha, Students

Dept Of Computer Science And Engineering

Sri Vasavi Institute Of Engineering And Technology,Pedana,A.P,India

ABSTRACT:

The real estate industry is one of the most price-oriented industries and tends to fluctuate. The objective of the project is predicting the rental price for a house. In this study, a predictive model based on the factors that influence the rental price has been constructed. The dataset has thirteen features. Regression techniques such as Gradient Boosting regressor, Random Forest regressor and linear regression were applied**.** A predictive model is built using the regression techniques, and to pick the best performing model by performing a comparative analysis on their performance scores obtained. The expected outcome of the models is measured using performance metrics such as Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE) and R-square score (R^2) metric. This project explains house rental price prediction model with the help of machine learning and the dataset used in our proposed model.

Keywords:

Regression, Gradient Boosting regressor, Random Forest regressor , linear regression,predictive, Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE) and R-square score (R^2) metric

INTRODUCTION:

Nowadays, traveling becomes much easier and more diversified. People are given a lot of options for accommodations during traveling. Where and what property type do you want to stay in when you are planning for your trip?

This may be the first concern for most people when they are making plans for the trips. Years ago, while the hotel used to be the first choice for many of us, now, people have more choices on online marketplaces, such as Airbnb.

Which is a platform connecting guests who need accommodations with hosts who want to rent their properties. By 2020, Airbnb has more than 150 million users, 0.65 million hosts, and 7 million listings globally [1].

According to the current growth, we may expect that the total number will continually increase in the next few years. Given a large number of hosts and listings, reasonable and competitive rental prices are important to maximize the benefits for both hosts and guests.

LITERATURE SURVEY:

we have to analyse the different Machine Learning algorithms for better training Machine Learning model. Trends in housing cost show the current economic situation and as well as to

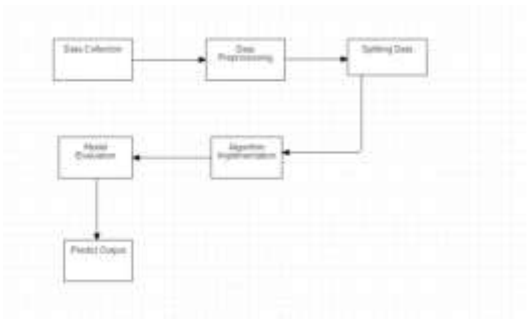
ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

directly concern with buyers and sellers. Actual cost of house is depending on so many factors. They include like no of bedrooms, number of bathrooms, and location as well.in rural area cost is low as compare to city. The house price grate with like near to highway, mall, super market, job opportunities, good educational facilities etc. Over few years ago, the real estate companies trying to predict price of property by manually. In company there is special management team is present for prediction of cost of any real estate property. They are decide price manually by analysing previous data. But there 25% of error is occurred on that prediction.so there is loss of buyers as well as sellers. Hence there are many systems are developed for house price prediction. Sifei Lu, Rick Siow had proposed advance house prediction system. The main objective of this system's was to make a model which give us a good house price prediction based on other features

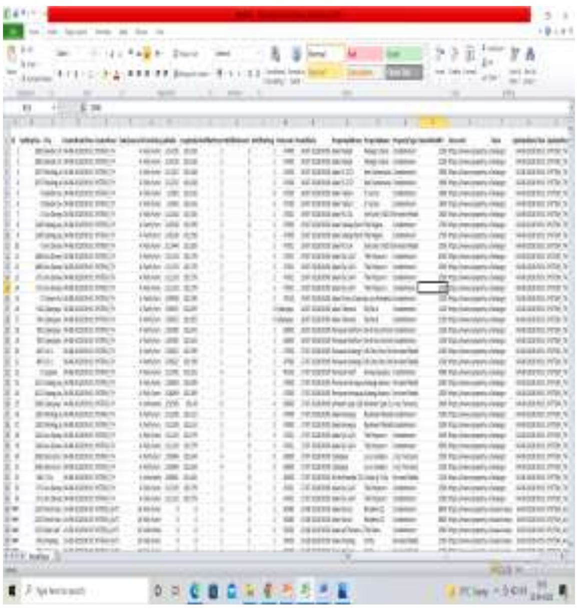
PROPOSED:

We proposed the system "**House Rent Price Forecasting Using** Machine Learning" like Gradient Boosting regressor, Random Forest regressor and linear regression were applied. we have predict the house price using multiple features

Advantages:
High accuracy
Less error rate
SYSTEM ARCHITECTURE:



Dataset:



ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

IMPLEMENTATION:

This repo contains the Predict Rental Prices project as part of my data science portfolio. The rental prices were scrapped from a property website based on properties located in Malaysia. Four different types of algorithms were used to train the models which consist of three machine learning (ML) algorithm. This project focuses on comparing the performance and results of these four algorithms instead of explaining the theory behind the algorithms as it assumes that readers already familiar with these algorithms. For readers who are in a hurry, feel free to dive straight into section 4. Discussion and Conclusion.

1. Linear Regression
2. Random Forest
3. Gradient Boosting Regressor
4. Discussion and Conclusion

LINEAR REGRESSION ALGORITHM:

Step:1 Reading and understanding the data

Step:2 Visualizing the data (Exploratory
Data Analysis)

Step:3 Data Preparation

Step:4 Splitting the data into training and
test sets

Step:5 Building a linear model

Step:6 Residual analysis of the train data:

Step:7 Making predictions using the final
model and evaluation:

RANDOM FOREST:

We know that error can be composited from bias and variance:

- error = bias + variance

A too complex model has low bias but large variance, while a too simple model has low variance but large bias, both leading to a high error but two different reasons. As a result, two different ways to solve the problem come into people's mind (maybe Breiman and others), variance reduction for a complex model, or bias reduction for a simple model, which refers to random forest and boosting. These are two opposite way to achieve a low error.

A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap Aggregation, commonly known as bagging.Bagging, in the Random Forest method, involves training each decision tree on a different data sample where sampling is done with replacement. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees, thus helps to tackle the

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

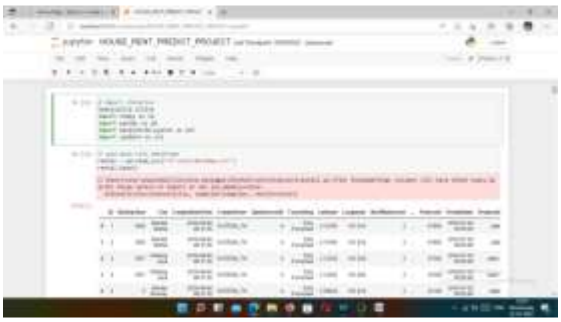
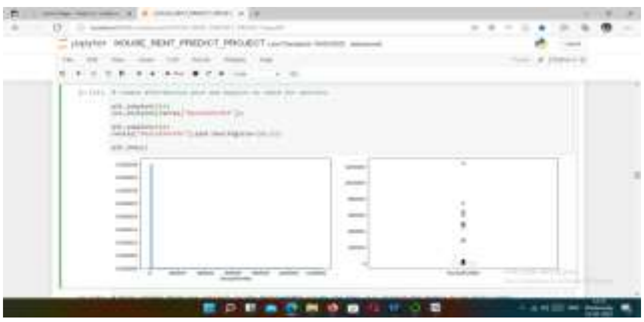
problem of overfitting in decision trees. Random forest reduces variance of a large number of "complex" models with low bias. We can see the composition elements are not "weak" models but too complex models.

Baseline Model - Random Foresst
n_estimators : integer, optional (default=10Changed in version 0.20: The default value of n_estimators will change from 10 in version 0.20 to 100 in version 0.22.
max_depth : integer or None, optional (default=None)
The maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples.
Baseline Model - Gradient Boosting Regressor
learning_rate : float, optional (default=0.1)
learning rate shrinks the contribution of each tree by learning_rate. There is a trade-off between learning_rate and n_estimators.
n_estimators : int (default=100)
The number of boosting stages to perform. Gradient boosting is fairly robust to over-fitting so a large number usually results in better performance.
max_depth : integer, optional (default=3)
maximum depth of the individual regression estimators. The maximum depth limits the number of nodes in the tree. Tune this parameter for best performance; the best value depends on the interaction of the input variables.
max_features : int, float, string or None, optional (default=None)
The number of features to consider when looking for the best split:

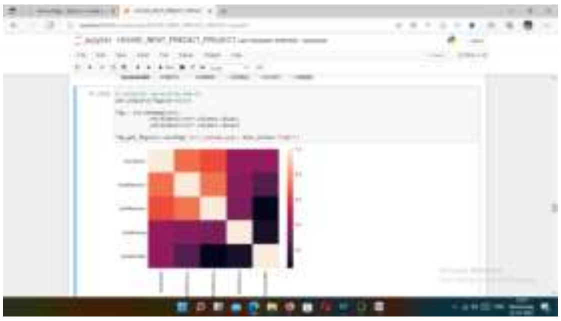
- If int, then consider max_features features at each split.
- If float, then max_features is a fraction and int(max_features * n_features) features are considered at each split.
- If "auto", then max_features=n_features.
- If "sqrt", then max_features=sqrt(n_features).
- If "log2", then max_features=log2(n_features).

- If None, then max_features=n_features.

RESULTS:

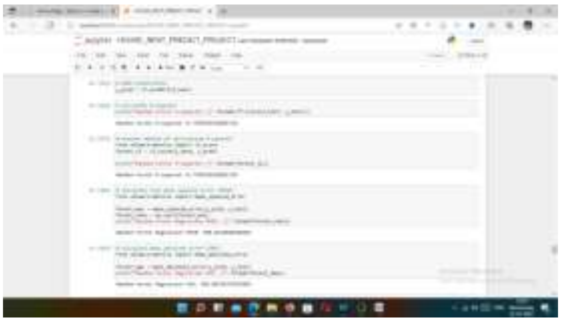



ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

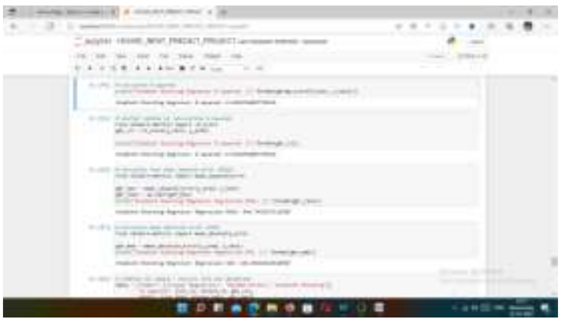


Linear Regression:

Randome forest:



Gradient boost:



Comparision:

CONCLUSION:

Predicting the rental price of house property using machine learning classifier is a challenging problem. The selection and training of a suitable machine learning model for this purpose depends on many factors including but not limited to the type of data, influencing features, accuracy and classifier's structure. Our study shows that the influencing parameters for rent prediction highly depends on the type of a housing property. HIGHER R-SQUARED (R2) VALUES AND LOWER MEAN ABSOLUTE ERROR (MAE) SHOWS HIGHER ACCURACY. Finally randome forest is give high accuracy. Our future work will investigate these challenges among them predicting future rental price of a property and study it as a time-series problem. **REFERENCES:**

[1] Antonio Bella Sanjuan, Cesar Ferri Ramırez, Jos ´ e Hern ´ andez Orallo, and ´ Marıa Jose Ramırez Quintana. Model integration in data mining: from ´ local to global decisions, Universitat Politecnica de Val ` encia, 2012. `

[2] Hujia Yu, Jiafu Wu, Real Estate Price Prediction with Regression and Classification, 2016.

[3] Ines M. Galv ´ an, Jo ´ s M. Valls, Miguel Garc ´ ´ıa, and Pedro Isasi. A lazy learning approach for building classification models. International journal of intelligent systems 26, no. 8 (2011): 773-786.

[4] Jerome H. Friedman, Ron Kohavi, and Yeogirl Yun. "Lazy decision trees." In AAAI/IAAI, Vol. 1, pp. 717-724. 1996.