JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

HOUSE RENT PREDICTION USING MACHINE LEARNING ALGORITHMS – A METHODICAL REVIEW

[1]R.Geethamani, [2]B.Karthika
[1]Assistant Professor, [2]II M.Sc (CS),
[1]Department of Computer Science, [2]Department of Computer Science,
[1,2] K.R College of Arts and Science, KovilpattiTamilnadu

Abstract

House rent prediction is crucial for a common type of person because the rent of a property rises every year. Some people desire a home with a low rent cost and useful room, thus our article provides useful information about rental costs for needed customers. Hence, a methodical approach to predicting future home rentals is necessary for our demographic. The buyer who rents the home for their specific requirements is aware of the ideal rental period. Numerous factors, including location, area, square footage, rent, and house size, affect a house's rent. This research explores and implements several regression methodologies, such as linear regression, random forest regression, gradient boosting regression and XG boost regression algorithms in machine learning are used to predict the rental value of a residence.

Keywords: House Rent, Linear regression, Random forest, XG boost regression, Gradient boosting regression.

1. INTRODUCTION:

Many real estate platforms use the right data and machine learning algorithms to identify house rent possibilities that fit the customer budget. In all over the world, the rental housing industry is one of the most significant real estate markets. The academics and practitioners can benefit greatly from research into rental price patterns. Tenants find it most cost-effective to discover suitable pricing based on projected location and housing quality, but rental property owners must establish appropriate prices in order to attract the interest of possible renters. Predicting the housing Rent is a good indicator for urban researchers.

Regression models are popular for examining house/rental prices and their factors; however their use is opposed due to errors in basic model assumptions, feature selection, and the use of unstructured data. In this paper, the predicted rent of the house using the comparison of regression algorithms and its gives the best performance for random forest regression. The price per square foot, rent price and number of bedrooms are some classic factors used to forecast rent. Other characteristic does not affect the rental prices.

2. DATASET: Annual rent growth rate hit 11.6% at the end of 2021 and start of 2022, about three times what it was in the five years prior to the pandemic, according to the Harvard Joint Center for Housing Studies.The dataset was taken from the Kaggle and the values of the data cleaned dataset is belongs to united states, there are 26 attributes and all the values are only in numerical. Some important Attributes are Region- A formal region could be any country in the world, like the United States, or the linguistic region of a state. Price describes the rent of particular houses. Each value of the attributes has their own identities and gives the specified information of the houses. There were far more people looking to rent new homes than there were homes available. So that rent will increases over upcoming year. The past three months have seen an annual rise in asking rent of 15.8 percent. The latest government statistics, which cover all rents including existing tenancies, paid a high amount.

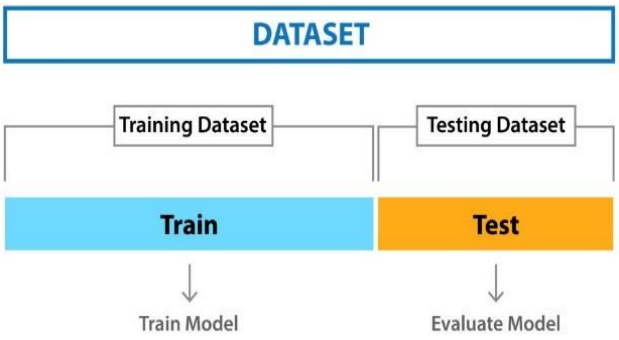

Fig: 2.1 Dataset preprocessing

3. METHODOLOGY: The data has to be pre-Processed, first of all we have to import the libraries, load the dataset and the split the data for training and testing process. The fit Transform has to be required for the dataset for predictions, evaluation and performs a performance metrics like MAE, RMSE, R^2_Score for Linear regression, random forest, Gradient boosting and XG boost regression algorithms.

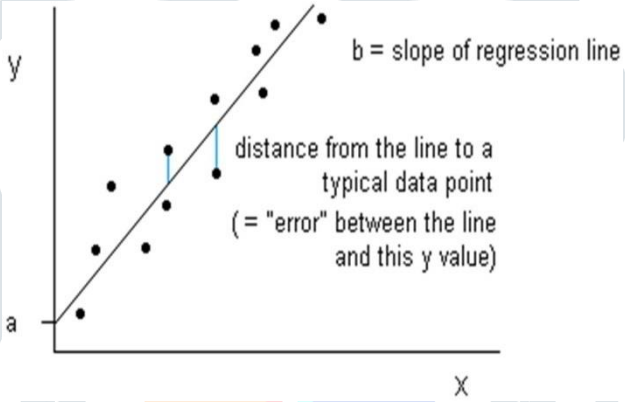

Fig: 2 Regression Technique

MAE: Mean absolute error (MAE) is a loss function used for regression. The loss is the mean Over the absolute differences between true and predicted values, deviations in either direction from the true value are treated the same way.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Where,

MSE = mean squared error
n = number of data points
Y_i = observed values
$\hat{Y}_i$ = predicted values

RMSE: Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). The RMSE estimates the deviation of the actual y-values from the regression line. Another way to say this is that it estimates the standard deviation of the y-values in a thin vertical rectangle.

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \hat{x}_i)^2}{N}}$$

Where,

RMSD = root-mean-square deviation
i = variable i
N = number of non-missing data points
x_i = actual observations time series
$\hat{x}_i$ = estimated time series

R 2_ score: Coefficient of determination also called as R 2 score is used to evaluate the performance. It is the amount of the variation in the output dependent attribute which is predictable from the input independent variable(s).

$$R^2 = 1 - \frac{RSS}{TSS}$$

Where,

R^2 -Coefficient of determination RSS -Sum of squares of residualsTSS – Total sum of squares

4. REGRESSION ALGORITHMS

To enable machines to learn the relationships within the provided data and create predictions based on patterns or rules discovered from the dataset. In this study, used a variety of algorithms in machine learning, Regression is a machine learning technique that uses continuous numerical values to predict output.

Linear Regression: It is a machine learning algorithm based on supervised learning. It executes a regression operation. Linear Regression uses independent variables to model a goal prediction value. It is mostly used to determine how variables and forecasting relate to one another. The association between the data points is used in linear regression to create a straight line connecting all of them. Future values can be predicted using this line. Future prediction is crucial in machine learning. A linear regression line has an equation of the form ***Y = a + bX*** where *X* is the explanatory variable and *Y* is the dependent variable. The slope of the line is *b*, and *a* is the intercept (the value of *y* when *x* = 0). More precisely, linear regression is used to determine the character and strength of the association between a dependent variable and a series of other independent variables. It helps create models to make predictions, such as predicting a company's stock price etc.

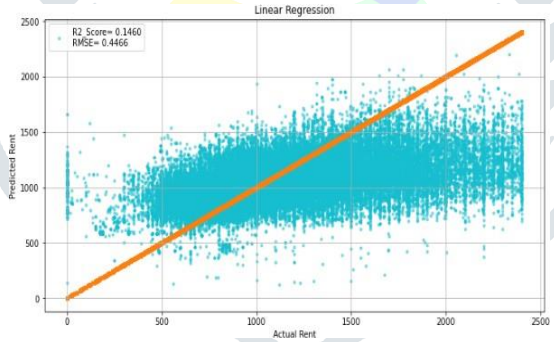

Fig:4.1 Linear regression

Random forest: Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model. Random Forest is an ensemble technique capable of handling both regression and classification tasks. The primary idea behind this is to combine numerous decision trees in selecting the final output instead of depending on individual decision trees. The underlying learning models of Random Forest include a variety of decision trees. For each model, sample datasets are created by randomly selecting rows and features from the dataset.

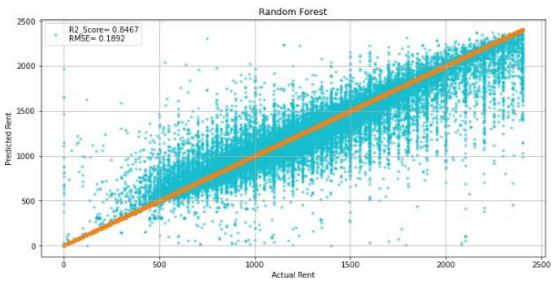

Fig: 4.2 Random forest

Gradient Boosting: Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees. Both

continuous and categorical target variables can be predicted using the gradient boosting approach (as a Regressor) (as a Classifier). The cost function is Mean Square Error (MSE) when it is used as a regressor, while it is Log loss when it is used as a classifier.

$$L_{\text{MSE}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - F(x_i))^2$$

The observation that residuals for a given model are proportional to the negative gradients of the (MSE) loss function leads to a development of this concept to loss functions other than squared error as well as to classification and ranking problems.

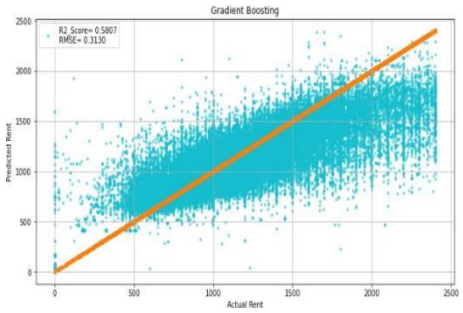

Fig:4.3 Gradient Boosting

XG Boost: The method of XGBoost is effective for creating supervised regression models, Knowing about its objective function and base learners allows one to infer the truth of this assertion. Loss function and a regularization term are both part of the objective function. Decision trees and linear regression are two widely used regression algorithms. Regression involves many measures, such as the root-mean- square error (RMSE) and mean-square error (MSE). MAE: It is the sum of actual and expected differences in absolute terms, but it is not mathematically sound, which is why it is less frequently employed than other metrics.

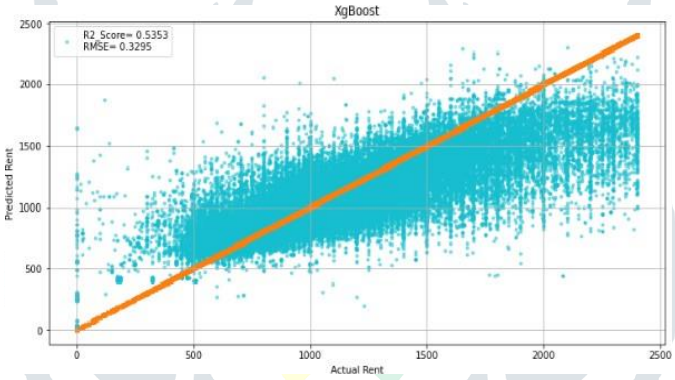

Fig:4.4 XG Boost

RESULT

The process of developing machine learning models must include evaluating the model accuracy in order to characterize how well the model is performing in its predictions. The type of the problem affects the evaluation metrics. In regression analysis, the MSE, MAE, RMSE, and R-Squared metrics are primarily used to assess prediction error rates and model performance.

Finally predicted the outcomes of the following regression algorithms described in Table.

Table 1: Model Performance Based on MAE, RMSE and R^2 Score

Algorithms	MAE	RMSE	R^2 SCORE
Random Forest Regression	0.0615	0.1892	0.8557
Gradient Boosting Regression	0.1774	0.3129	0.5806
XG BoostRegression	0.1774	0.3294	0.5353
Linear Regression	0.2658	0.4466	0.1458

5. CONCLUSION

In this research paper, the primary goal is to estimate the rent of a residence by examining the demands of the client and their financial situation. This study paper will assist the clients in learning the actual rent for the residence, and it will assist the owners in

learning the rental rate that best suits the client's requirements. The comparison of all the algorithms gives the results but Random forest regression gives the best results among all the algorithms. Because random forest have less mean absolute error and root mean square error compare to other algorithm and Random forest have 84% of R^2_score, so it is consider to the best one.

6. FUTURE SCOPE

According to a recent analysis, homes of the future will help to integrate with local energy networks, have adaptable floor plans, and employ improved technology to make decisions about heating, security, and even mail deliveries automatically. By the end of 2022, rents will have increased by 6%, but due to the government's budgetary plans' sudden increase in mortgage rates, rents will drop by 5% in 2023 and by another 5% in 2024. The city's rent for apartments is rising, and estimating the actual rent for a home is very risky. Everyone can accurately predict the rent of the residence with the guidance of this prediction methods using machine learning.

REFERENCES

[1] ´zar CF, Ceriani L, Olivieri S, Ranzani M. Rent-Imputation for Welfare Measurement: A Review of Methodologies and Empirical Findings. Rev Income Wealth. 2017; 63(4):881–9

[2] Johan Oxenstierna. Predicting house prices using Ensemble Learning with Cluster Aggregations. 2017.

[3] P. Rezazadeh, L. Nikolenko, and H. Rezaei. Airbnb Price Prediction Using MachineLearning and Sentiment Analysis. 2019.

[4] Nor Hamizah Zulkifley, Shuzlina Abdul Rahman, Nor Hasbiah Ubaidullah - House Price Prediction using a Machine Learning Model.

[5] Bharatiya, Dinesh, et al. “Stock market prediction using linear regression.” Electronics, Communication, and Aerospace Technology (ICECA), 2017 International conference of. Vol. 2. IEEE, 2017.

[6] Sifei Lu, Zengxiang Li, Zheng Qin , Xulei Yang , Rick Siow Mong Goh - “A hybrid regression technique for house prices prediction” 2017,IEEE.

[7] CH. Raga Madhuri, G. Anuradha, M. Vani Pujitha -” House Price Prediction Using Regression Techniques: A Comparative Study” 2019 in (ICSSS), IEEE.

[8] G.Naga Satish, Ch.V.Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu “House Price Prediction Using Machine Learning”. IJITEE, 2019.