Review on House Price Prediction through Regression Techniques

Mrs. Sonal Naik[1], Mrs. Aarti Puthran[2]

[1]Student,Dept .of Computer Engineering, Shree L.R.Tiwari college of Engineering, Maharashtra, India

[2]Assistant Professor, Dept. of Computer Engineering, Vidyavardhini's College of Engineering ,Vasai Road, Maharashtra, India

---***---

Abstract – *Housing price ranges are of excessive interest for both purchasers and vendor. First consider a situation where a person needs to purchase a house. The person will look for his/her chosen house for a price tag. The person will have some structures decided what he/she wants to have in the house. The person will be able to decide the type of house he/she is looking is good of the price or not. Similarly, consider a situation where a person needs to sell a house. They used the house price prediction system; the seller would be able to decide what all structures he/she could add in the house. So that the house can be sold-out for a higher price. Hence, from both the above states we can confirm that house price prediction is useful both for the buyer and seller. This paper present various algorithms while predicting house prices with good accuracy such as Linear Regression, Support Vector Regression, Random Forest Regression, Decision Tree Algorithm and selected the best fit among the algorithm. This paper guides that it can be best application of machine learning models in order to optimize the result. This paper is comparing different algorithm for house price predication.*

Key Words: *Decision tree regression, House Price Prediction, Linear regression, Machine Learning (ML), Random forest regression, Support vector machine .*

1.INTRODUCTION

House is one of human life's most essential needs, along with other fundamental needs such as food, water, and much more. Demand for houses raised rapidly over the years as people's living standards improved. There are people who make their house as an investment and property; most people around the world are buying a house as somewhere to live or as their income. Buying a house is certainly one of the most important decisions one makes in his life. The price of a house may depend on different factors from the house's location, its features, as well as the property demand and supply in the real estate market. Machine learning develops algorithms and builds models from data, and used that data to predict on new data. Supervised learning is used to train labeled data, while unsupervised learning is used to train unlabeled data. There are a few common machines learning algorithms, such as classification, linear regression, neural network and deep learning.

How to use machine learning algorithms to predict House price? This paper will used linear regression algorithm, Support Vector machine, Random Forest Regression, Decision *Tree* to predicate the house price. It is resolute by location, size, house type, city, country, tax rules, economic cycle, and population, movement, interest rate, and many other factors which could affect demand and supply. So we can use linear regression algorithm to predicate the house price. Linear regression is a predictive modeling technique that finds a relationship between independent variables and dependent variables (which is a continuous variable). The independent variables can be categorical or continuous while dependent variables are continuous. It can predict house prices more accurately based on their attributes, regardless of the data. Underlying function mapping can be linear, quadratic, polynomial or other non-linear functions but this article is on linear technique. Regression techniques are heavily used in making real estate price prediction, financial forecasting, and predicting traffic arrival time [2].

The goal of this study is through examining a real past Transactional dataset to derive valuable insight into the housing market. The house price prediction of the house is prepared using different Machine Learning algorithms. 70% of data form knows dataset is used for training purpose and remaining 30% of data used for testing purpose. [11] Accurate prediction of house prices has been always an interest for the buyers, sellers and for the bankers also. Many researchers have already worked to unravel the mysteries of the prediction of the house prices [1] [2].

2. LITERATURE SURVEY

Debanjan Banerjee, Suchibrota Dutta proposed Predicting the Housing Price Direction using Machine Learning Techniques the issue of changing house rate as a categorization issue and relates ML procedures to expect whether house rates will rise or fall. This work applies different component choice schemes, like variation impact factor, Information approval, and standard part examination and information change procedures, for example, inconsistency and missing price treatment just as box-cox change methods. This paper used three techniques Support Vector Machines (SVM), Random Forest and artificial neural network (ANN).Random Forest gives more precision anyway simultaneously this specific sort of classifier also motivated to over fitting along these lines the presentation of Support Vector Machine classifier can have supposed to be hard and stable over the remainder of the two classifiers [1].

Reference 2 explains the prediction of housing prices for the customers is since their commercial plans and requirements. The examination of land regard is felt basic to help the decisions in urban planning. The land system is a shaky stochastic phase. Commercial authority's judgments rely upon nearby examples to protected most extreme returns. Originators are interested in knowing the future arrangements for their simple management. Too totally check land prices what's more, upcoming cases, huge amount of data that things show up cost is required for valuation, exhibiting, and determining [2]. Reference 3 explains the prediction the intelligible house prices is for non-house containers based on their economic requirements and their ambitions. Investigating the preceding merchandise, fare choices and cautions expansions, speculated prices will be predictable. This paper is to help the vender to approximation the selling price of a house perfectly and to service persons to predict the exact time slap to accumulate a house. [3] Wan Teng Lim proposes two processes, to expect Singapore covering market and to think about the perceptive performance of artificial neural Network (ANN) model, i.e., the multilayer perceptron, with autoregressive integrated moving average (ARIMA) model. To expect the future cooperative price index (CPI) the more established model is developed. The lower mean square error (MSE) of the artificial neural Network models indicated the occurrence of ANN over other perceptive contraptions. The estimates depended on time preparation information of causes that are acknowledged to impact the townhouse costs in Singapore. These reasons and the CPI were the sources of information and produce to the models, individually. Housing rates are a form of time sequence. There are various techniques ANN, ARIMA and MRA, have been used in predicting many types of time sequence with housing rate in other amounts of the world and monetary markets. [4] Reference 5 presents how to utilize Linear Regression, Forest regression and Boosted regression for expectation and efforts to provide an examination of the result. The ability of the calculation has been additionally extended with consumption of neural organizations. The structure will realize customers by giving exact produce and prevention the danger of putting funds into an inappropriate house. Extra highlights for the customer's advantage can likewise be added to the structure without upsetting its center effectiveness. They followed an exclusive method to increase the accuracy, examination led to an assumption that the real estate value also depends on close local services and now they propose exclusive methodology that can counter this requirement. If they find any community places in the group they growth the value of the property individually. They accepted this available with physical patterns and this provided us wonderful results in terms of accuracy in prediction. [5] Reference 6 presents restricted dataset and information consists of a functional and combination of data pre-processing, creative element and designing approach is examined. The papers propose hybrid Lasso and Gradient boosting regression model to predict separate house value.

They catch that the data quality is a key reason to calculate the house costs. Data input feature concentration approximation is main for regression. Relapse designs with boundary in extra of 10000 iteration are useful by thinking about the homoscedasticity verification. The outcome is determined by the homoscedasticity between training information and test information. Linearity of each section is the dimension key of relapse design, subsequently various changes are useful to upgrade the linearity of information highlights. [6] The Danh Phan explain Machine learning methods are applied to study demonstrable property exchanges in Australia to find respected models for house purchasers and dealers. There is the high inconsistency between house price in the most expensive and most reasonable suburbs areas in the city of Melbourne. Additionally, tests show that the combination of Stepwise and Support Vector Machine that based on mean squared error measurement is a competitive methodology. This paper looks for useful models for house value expectation. It similarly gives abilities into the Melbourne Housing Market. The first data is prepared and improved into a cleaned dataset ready for examination. Data reduction and change are done then applied by developing Stepwise and PCA strategies. Various strategies are then realized and evaluated to achieve a perfect arrangement. The assessment stage establishes that the combination of Stepwise and SVM model is a serious approach. It could be utilized for additional sending. [7] Reference 8 proposed that they will define explanation for House Value Progressive Regression.

Methods machine learning competition, which was held on Kaggle platform. This paper was proposed to predict house's sale price by their elements for example house area, year of building. In solution, they use classic machine learning processes, and our unique methods, which will be described in this paper. They apply data imputation, feature engineering and machine learning modeling to achieve a better predictive accuracy on the housing price. [8] Reference 9 proposed machine learning algorithms are used to construct the calculation model for houses. This paper propose machine learning algorithms such as logistic regression and support vector regression, Lasso Regression technique and Decision tree are employed to construct a predictive model. They have reflected housing data of 100 properties. [9] Reference 10 proposed the house price prediction of the house is done using different Machine Learning algorithms like Linear Regression, Decision Tree Regression, K- Means Regression and Random Forest Regression. 80% of data from dataset is used for training purpose and remaining 20% of data used for testing purpose. This work applies various methods such as features, labels, reduction techniques and transformation methods such as attribute combinations, set missing attributes as well as looking for new correlations. This all specifies that house price prediction is a developing research area and it requires the understanding of machine learning. [10] Reference 10 proposed house prices prediction will

help based on various considerations. The customers will be able to input the type of house they wish to purchase and with the help of machine learning algorithm the house price predictor will show the predictable price of the chosen house. There are many stages which help the customers and retailers to predict the price of the property they are need and the property they are looking for. They permit the user to enter the area of the house anywhere in India along with all the other features making the house price prediction system more effective. [11]

3. COMPARISON OF LEARNING TECHNIQUES

Table 1.1 Shows the Comparison of linear regression, random forest regression, decision tree regression and support vector regression based on certain parameter as feature, advantage, disadvantage, training Speed

Predication speed and understandability achieved through this technique. The ML Algorithm is the technique by which an AI system executes its task, and is most commonly used to predict output values from given input values. Regression is the central processes of machine learning. The regression model contains a set of machine-learning techniques that allow us to predict a label variable (y) based on the values of one or more element/feature variables (x). The aim of a regression model is to construct a mathematical calculation that defines y as a function of the x variables.

Linear regression is the simplest method for prediction. The linear regression are used the dependent variable in situation of a change in independent variables. It uses two things as variables which are the predictor variable and the variable which is the most crucial one first whether the predictor variable. These regression approximations are used to describe the relationship between one dependent variable and one or more independent variables. Training speed and predication speed are faster than random forest and support vector regression. Linear regression works simply and easy to understand. The equation of the regression equation with one dependent and one independent variable is defined by the formula [8]. Decision Tree Regression, as the name suggests it uses tree like structure to build regression and classification models. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final outcome is a tree with decision nodes and leaf nodes. A decision node has two or supplementary branches, each demonstrating values for the attribute tested. Leaf node represents a decision on the numerical target. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data. The decision tree repressor detects features of an attribute and trains a model in the form of a tree to predict data in the coming to produce significant output. Decision tree regression studies from the max depth; min depth of a graph and giving to system analyzes the data. Support vector machines are used for classification and regression. Support Vector Machine is typically used classification and regression calculation tool, that implements machine learning ideas to maximize predictive accuracy, which avoids over suitable to the data. A better learning technique must permanently avoid over fit of the data. Support vector regression deals huge benefits in so many features as this model can avoid over-fitting problems, while confirming a single optimum solution by reducing structural risks and empirical risks. The random forest regression detects features of an attribute and trains the model by examining given features. Random Forest regression from the graph, attribute combination, labels including structures and permitting to system analyses the data. Random Forests are cooperative classifiers built from of a set of Decision Trees, with the output of the classifier being the mode of the output of the Decision Trees. Random forest is a Supervised Learning algorithm which uses collective learning method for classification and regression. Random forest is a bagging method and not a boosting technique. The trees in random forests are executed in parallel. There is no communication between these trees while construction the trees. It operates by building a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

Algorithm	Linear regression	Decision tree regression	Random forest regression	Support vector machine
Features	It is used to approximation the dependent variable in situation of a change in independent variables.	It is used to construct tree Rules to order the data items.	It provides partial describe ability works well with both continuous and categorical predictors	Work well if information is not linearly separable in the base feature space.

Advantage	Works well irrespective of dataset size. Provides information about the relevance of features.	Works well on both linear and nonlinear problem. No essential to apply feature scaling.	Good performance on many difficulties including nonlinear.	Easily adjustable works very well on nonlinear problems.
Disadvantage	It is easy to use. The assumptions of linear regression.	Poor results on small datasets over fitting can simply occur.	No interpretability over fitting can easily occur. We need to select the no of trees.	Necessary to apply feature scaling. Not well-known difficult to understand.
Training Speed	Fast	Fast	Slow	Slow
Predication Speed	Fast	Fast	Moderate	Slow
Understandability	Simple to understand and construct.	Simple to understand and generate.	It is more difficult to interpret.	It is difficult to interpret.

Table -1.1: Comparison of learning technique

4. SUMMARY

This paper examines different models for housing price prediction. Four different types of Machine Learning methods including linear regression, Decision tree regression, Random Forest, and support vector machine are compared and analyzed for optimal solutions. Yet all of those methods achieved appropriate results, different models have their own accuracy and predication. The Random Forest and support vector machine has the lowest training set. Its time complexity is high since the dataset has to be fit multiple times. The result from linear regression algorithm reaches a better predictive accuracy on the housing price.

5. CONCLUSION

This paper mainly focus on the comparison between different machine learning algorithms (Linear Regression, Decision tree regression, Random forest regression, Support vector machine) about House price prediction analysis. Different machine learning algorithm is investigating and evaluated to achieve an optimal solution. The comparisons of different regression algorithm, linear regression are relatively faster. Linear Regression displayed the best performance, fast training speed and more accuracy for this Dataset and can be used for deploying purposes. From the above results, linear regression algorithm has high accuracy value when compared to all the other algorithms. This paper seeks useful models for house price prediction using linear regression

ACKNOWLEDGMENT

With the completion of this paper, I would like to take this opportunity to express my gratitude and deep regards to

REFERENCES

[1]Debanjan Banerjee, Suchibrota Dutta, "Predicting the Housing Price Direction using Machine Learning Techniques" 2018

[2]Nehal N Ghosalkar, "Real Estate Value Prediction Using Linear Regression", 2018.

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

[4]Wan Teng Lim, Lipo Wang, Yaoli Wang, and Qing Chang, "Housing Price Prediction Using Neural Networks", 2016

[5]Ayush Varma, Abhijit Sharma, Sagar Doshi, Rohini Niar, "House Price Prediction Using Machine Learning and Neural Networks", 2018.

[6] Sifei Lu, Zengxiang Li, Zheng Qin, Xulei Yang, Rick Siow Mong "A Hybrid Regression Technique for House Prices Prediction", 2018.

[7] The Danh Phan, "Housing Price Prediction using Machine Learning Algorithms: The Case ofMelbourne City, Australia", 2018

[8] Parasich Andrey Viktorovich, Parasich Viktor Aleksandrovich, Kaftannikov Igor Leopoldovich, Parasich Irina Vasilevna," Predicting Sales Prices of the Houses Using Regression Methods of Machine Learning", 2018

[9] Atharva Chouthai, Mohammed Athar Rangila , Sanved Amate, Prayag Adhikari, "HOUSE PRICE PREDICITION USING MACHINE LEARNING", Mar 2019

[10]ANAND G. RAWOOL , DATTATRAY V. ROGYE , SAINATH G. RANE, "House Price Prediction Using Machine Learning", MAY 2021

[11]Anirudh Kaushal, Achyut Shankar," House Price Prediction Using Multiple Linear Regression", Apr 2021