



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact Factor: 6.078

(Volume 7, Issue 6 - V7I6-1168)

Available online at: <https://www.ijariit.com>

Real Estate Price Prediction using Ensemble Methods

Nidamaneni Rajyalakshmi

nidamanenirajyalakshmi@gmail.com

LakiReddy Bali Reddy College of Engineering,
Mylavaram, Andhra Pradesh

Sai Pavan Kumar Kokkiri

saipavankokkiri4@gmail.com

LakiReddy Bali Reddy College of Engineering,
Mylavaram, Andhra Pradesh

Chandra Shekar V.

luckychandulu143@gmail.com

LakiReddy Bali Reddy College of Engineering,
Mylavaram, Andhra Pradesh

Dr. B. Srinivas Rao

drbsrinivasarao@lbrce.ac.in

LakiReddy Bali Reddy College of Engineering,
Mylavaram, Andhra Pradesh

ABSTRACT

Real estate is one of the popular fields in current society. Customers should be very careful when they are buying or selling the property. There exist several reasons for increasing in the demand of land and houses. The factors like locality of house, the rooms it consists of and cost of living at that particular place plays an important role for deciding the value of house. Automated house price prediction can be done using linear regression, random forest and linear classification methods which are also termed as Ensemble methods. In the system, the data will be cleaned initially like removal and detection of extreme data from datasets and then the algorithms be applied.

Keywords— Machine learning, Ensemble Methods, Linear Regression, Random Forest

1. INTRODUCTION

For any system that to be developed, we consider data as crucial. The prediction of can be done using various regression algorithms. Without data we can't train the model. Machine learning is used for building these models for data and used to predict the new data. These approach gives accurate outputs. Machine learning is a subfield of artificial intelligence. It can be used for constructing computer programs. It gives system capability to learn automatically and improve its performance without being explicitly programmed. Machine learning closely related to statistics. It contains a verity of applications such as filtering of emails. Data will be the important part when it comes to machine learning algorithms. Algorithms in Machine Learning can be divided into four types. They are unsupervised, supervised semi-supervised and reinforcement algorithms.

In Supervised learning algorithm. It is a type of machine learning method in which we provide the data to it in order to train it, and that basis, it provides output. The main goal for supervised learning is to map input data with output data. spam filtering is the example of supervised learning. Supervised learning algorithms are Regression, Classification, Random Forest, Decision Tree, and Logistic Regression.

In algorithm of Unsupervised learning, it does not contain any destination variable value. It is used for clustering and association into different groups. KNN, K-means, Apriori algorithm, Neural networks, Singular value decomposition, Independent Component Analysis are some examples of Unsupervised learning algorithms.

The background work that has done earlier is studied for the reference. The paper also explains about the dataset that taken into account and pre-processing techniques that had applied. Taking the consideration different algorithms are applied and the one with highest accuracy is considered. Finally, the result, conclusions and references are discussed in the paper.

2. LITERATURE SERVEY

Housing price trends are helpful for the buyers and sellers. House price also based on various functions like number of Bedrooms and Bathrooms, Location. The house which is near by malls, colleges, highway, schools has higher value compared to other houses. Keeping all these in mind and price prediction of house without automation is not an easy task and results obtained are also not

accurate. So, a system should be developed to predict the house price.

The paper proposed by Adair, J. Berry, W. Mc Greal [1], In this system they suggested that Boosted regression. By using Neural networks algorithm, the efficiency of the algorithm be increased for further implementation. By using this system risk of investing can decreased in the wrong house and it provide accurate results. In this system customers can also add the features to the existing system without any disturbance of its functionalities. The major useful of this project is the customer can easily add or update the information in this database withoutany interruption to get more accuracy and finite decision.

The paper proposed by, Muhammad Fahmi, Mukhlshin, Adi Wibowo and Ragil Saputra [2] shows the prediction model is based on regression techniques like Decision Tree, Logistic regression, Ada Boost and Naïve Bayes. Pricing is implemented using regression algorithms to predict the particular city based on areas. This system predicts the prices of houses most accurately according to the requirements of customer. This project also analyses price ranges and past industry trends to predict future prices. In this system finally, Ada Boost and Decisiointree gets the high accuracy. Using this rules, value of the house can be predicted.

The paper proposed by Priyanka Khaire, Atharva Chogle Akshata Gaud, Jinal Jain [3] This is a web app which help us to create new documents and text. It contains tools which helps us to clean and transform the data. Re-transforming of numeric values, development of models using numerical values, development of visuals using data and tools. In this paper USA dataset can be taken and can apply regression techniques like Ridge Regression, Multiple Linear Regression, Elastic Net Regression, LASSO Regression, Ada Boosting Regression, Elastic Net Regression gradient boosting comparing all this algorithms Gradient boosting gives high accuracy then compared to other algorithms.

The paper proposed by B. Case, et al., [4], shows estimating price of house based on historical data of property markets. In this system using techniques of machine learning are implemented to detect the historical properties of Australia to know the methods for customers and dealers. In this paper Melbourne dataset is taken for estimate the price of houses. By using SVM (Support Vector Machine), Naïve Bayes, Regression Tree and Neural Network technique estimate the price of house and this depends on measurement of mean squared error is a competitive approach. these all algorithms can be used to find the useful one and check the accuracy.

3. PROPOSED SYSTEM

Now a days everything gets automated. The main aim of our project is to minimize the manual work of customer. In manual process, the customer should enquiry a real estate agent to get the suggestion for suitable properties which meet their requirements. But the drawback of the system, it has high risk of cheating by agent and waste of time. So, to overcome drawback we propose an automated and new system. First step of the system is scraping of data. structured data is extracted in this step from application like web and later it saved to csv file or database. RPA flow chart is developed using UiPath studio platform. It also used to provide the power data scraping. In next step, we perform data cleaning. Which involves modification applied before applying algorithm, in this process raw data into clean dataset which deals with missing data, categorical data as per requirement. In the next step, algorithms will be applied. There exist various algorithms that we use to estimate price of house. Regression Techniques like Random Forest, Linear regression, Logistic regression, Decision Tree and Ensemble methods like Ada Boost, XG Boost and Gradient boosting we will use these algorithms for prediction.

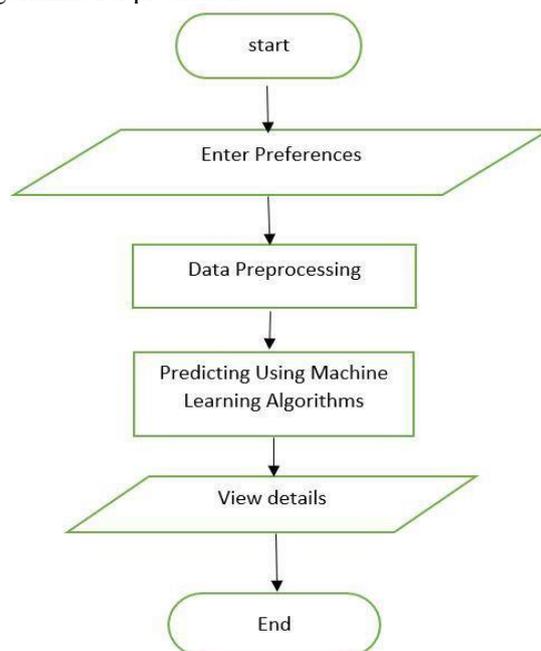


Fig 1: Flow Diagram of the Proposed System

XG Boost is one of the highly effective Machine Learning algorithm, it is an updated implementation of GBM. XG Boost can predicts more than other algorithms and predicting power is almost ten times higher and it is used to improves overall performance of algorithm. XG Boost also called as regularized boosting techniques. Ridge regression gives the better long term predictions this technique reduces the model complexity.

Lasso (Least Absolute and Selection Operation) regression is one of the regularized techniques. It can be useful to reduction the model complexity. When compared to the ridge regression it exhibits all properties excluding the penalty term properties. It contains absolute weight where in ridge regression it contains square of weights. The extension of bagging estimator algorithm is random forest. In random forest base estimators are decision trees and it selects features set in any order by which we can able to decide the decision tree's best split at every node.

4. RESULTS AND DISCUSSION

After analysis of Regression techniques and Boosting algorithms. In our proposed system XG Boost algorithm and Linear regression give the accuracy as show in the table. The proposed system will be easy to use and will contain simple operations. The data set after pre-processing looks like:

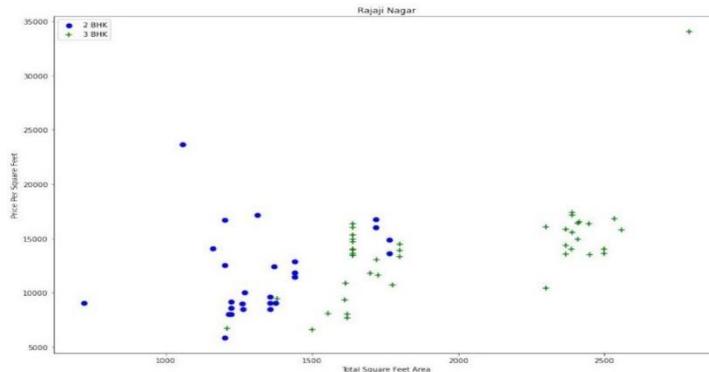


Fig. 2: Data set before pre processing

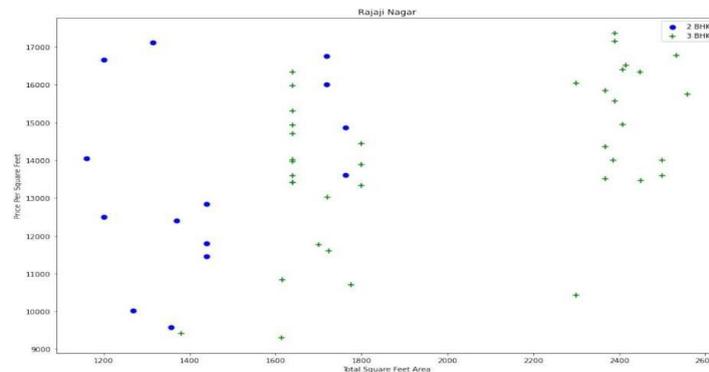


Fig. 3: Data set After pre processing

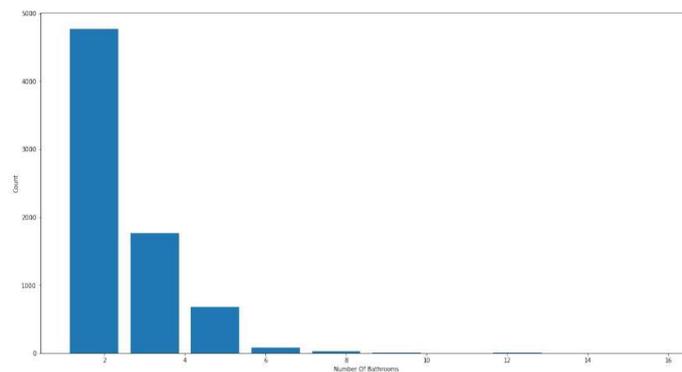


Fig. 4: Graph shows Count of bathrooms

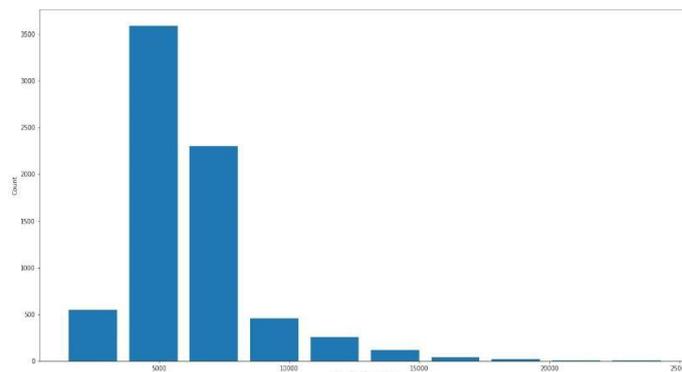


Fig. 5: Graph with count of price per sq. ft.

Table 1: Result Table

S.no	Model	Best_Score	Best_Params
0	Linear_regression	0.818354	{'normalize': True}
1	lasso	0.763861	{'alpha':0.1, 'selection': 'cyclic'}
2	ridge	0.819450	{'alpha':0.5}
3	Random_forest	0.759469	{'max_depth': 8, 'n_estimators': 500}
4	XG Boost	0.831172	{'n_estimators':300}

5. CONCLUSION

By using this automated system, we can easily estimate the price of houses of our preference. In this system Ensemble methods are used by which we can get accurate result. This system can also be extended further by including lands in real estate.

6. REFERENCES

- [1] J. Berry, Adair, W. McGreal, Hedonic modeling, housing submarkets and residential valuation, *Journal of Property Research*, 13 (1996) 67-83.
- [2] Muhammad Fahmi, Ragil Saputra, Mukhlishin and Adi Wibowo. "Predicting house sale price using fuzzy logic and K-Nearest Neighbor." *Informatics and Computational Sciences (ICICoS)*, 2017 1st International Conference on. IEEE, 2017.
- [3] Jinal Jain, Atharva chogle, priyanka khair, Akshata gaud. "House Price Forecasting using Data Mining Techniques" *International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 6, Issue 12, December 2017.*
- [4] B. Case, et al., *Journal of Real Estates Finance and Economic*, vol. 29 no. 2, pp. 167-191, 2004.
- [5] Gupta, R., Kabundi, A., & Miller, S. M. (2011). Forecasting the US real house price index: Structural and non-structural models with and without fundamentals. *Economic Modelling*, 28(4), 2013-2021.
- [6] Aasen M. R. (2011) Applying Altman's Z-Score to the Financial Crisis an Empirical Study of Financial Distress on Oslo Stock Exchange.
- [7] Mu, J., Wu, F. & Zhang, A., 2014. Housing Value Forecasting Based on Machine Learning Methods. *Abstract and Applied Analysis*, 2014(2014), p.7.