



House Price Prediction Using Machine Learning And Deep Learning

Prof. Sayali Kokane¹, Anand Dhoka², Prajwal Bendale³, Mayur Bhise⁴, Nilesh Bhosale⁵

¹Professor, Information Technology, APCOER Pune, Maharashtra, India

^{2,3,4,5} B.E Student, Information Technology, APCOER Pune, Maharashtra, India

ABSTRACT

Now a days house sale is majorly seen in metro cities. The market demand for housing is always increasing every year due to increase in population and migrating to other cities for their financial purpose. Prediction of house resale price for long-term temporary basis is important especially for the people who stays who will stay the long time period but not permanent and the people who do not want to take any risk during the house construction. In this paper, the sale price prediction of the house is done using different classification algorithms like Logistic regression, Decision tree, Naive Bayes and Random forest is used and we use AdaBoost algorithm for boosting up the weak learners to strong learners. Several factors that are affecting the house resale price includes the physical attributes, location as well as several economic factors persuading at that time. Here we consider accuracy as the performance metrics for different datasets and these algorithms are applied and compared to discover the most appropriate method that can be used the reference for determining the sale price by the sellers. Index Terms— House sale Price, Prediction, Random Forest, AdaBoost, Naïve Bayes.

Keyword : House sale Price , Prediction ,Random Forest.

INTRODUCTION.

Investment is a business activity that most people are interested in this globalization era. There are several objects that are often used for investment, for example, gold, stocks and property. In particular, property investment has increased significantly since 2011, both on demand and property selling [1]. One of the increasing of property demand is because of high population in Indonesia. Indonesian Central Bureau of Statistics states that in East Java 50% of the population of East Java classified as a young population who have age approximately at 30 years old.

The result of this census indicates that the younger generation will need a house or buy a house in the future. Based on preliminary research conducted, there are two standards of house price which are valid in buying and selling transaction of a house that is house price based on the developer (market selling price) and price based on Value of Selling Tax Object (NJOP). According to Lim, et al the fundamental problem for a developer is to determine the selling price of a house . In determining the price of home, the developer must calculate carefully and determine the appropriate method because property prices always increase continuously and almost never fall in the long term or short .

This research aims to create a house price prediction model using regression and PSO to obtain optimal prediction results. PSO is used for selection of affect variables in house prediction, regression is used to determine the optimal coefficient in prediction. In this study, researchers wanted to know the performance of the developed model in time series data. Prediction house prices are expected to help people who plan to buy a house so they can know the price range in the future, then they can plan their finance well. In addition, house price predictions are also beneficial for property investors to know the trend of housing prices in a certain location also using map integration and layout shown.

LITERATURE SURVEY.

We have taken pune as city for case study. The property size in square feet, location, and its facilities are all key aspects affecting cost. 9 different attributes are used. The Multiple linear regression (Least Squares), Lasso/Ridge regression, SVM, and XG Boost are used for experimental work.

Luo suggests that to explain the factors that determine residential asset prices, most studies have concentrated on macroeconomic aspects. It looks at some micro characteristics, such as lot size and pool size, that can be utilised as features to estimate house price in this research. Random forest and support vector machine are two machine learning methods which are used to predict asset pricing. R squared is more than 0.9 in all regression models. Panjali and Vani state that forecasting the sale price of a house on a long-term is vital, especially for those who will be residing there for a considerable duration while selling it again later. It also applies for those who want no risks while the dwelling is being constructed. Authors utilize various classification methods such as Logistic regression, Decision tree, Naive Bayes, and Random Forest to work out the house's sale value. It also applies AdaBoost technique to assist weak learners to be strong ones. The physical characteristics, location, as well as numerous economic aspects persuading at the time decide the resale price of a house. Accuracy is used to measure performance for different datasets and unleash the optimal way for sellers while expecting the resale price.

Indicate that over the next decade, India's housing market is expected to increase at a rate of 30- percent. It is only second to the agriculture industry in terms of job creation. Pune makes it an excellent spot to invest in real estate. The inconsistency in housing valuation is a challenge for a house buyer. Estimated price must be a win-win midpoint for both the seller and the the buyer. This will confirm whether the price is underestimated or overestimated.

To do this, various features from the set of features are picked as input, while using algorithms such as Decision Tree and bagging techniques such as Forest. Wang et al.

State that studies that do not take into account all of the factors influencing property values, provide inaccurate forecast results. As a consequence, for house prediction, the authors propose a full circle joint self attention model. Authors employ satellite imagery to assess the environment around the residential area. Input information about public facilities such as gardens, academic institutions, and BRT stops are used to depict the amenities. The method leverages attention mechanisms that are commonly used in picture, voice, and translated content to identify important points that potential home buyers evaluate. When fed transaction data, the model can apply weights automatically. The proposed model differentiates itself from self attention models since it takes into account the interrelationship of two different parameters in order to learn the complex relationship.

COMPARISON OF RELATED WORK.

Nowadays there is a need of house pricing because constant population growth and intercity changes of the population they need a interface ,site ,page that helps them or tells the near to exact ratio,price of an apartment building flat house etc .that helps an user satisfying his/her problems. And pricing is shoen so that he/she knows where to take and where not to.

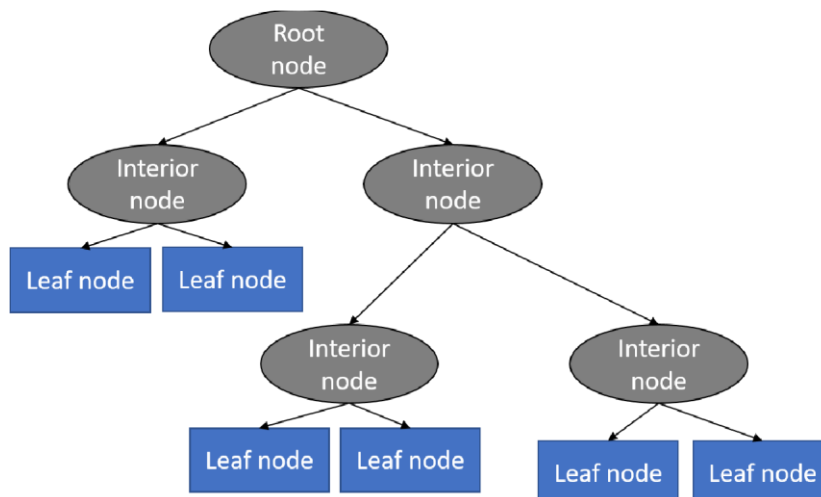
METHODOLOGIES.

The Ridge Regression

Is an L2-norm regularised regression technique that was introduced by Hoerl in 1962 . It is an estimation procedure to manage collinearity without removing variables from the regression model. In multiple linear regression, the multicollinearity is a common problem that leads least square estimation to be unbiased, and its variances are far from the correct value. Therefore, by adding a degree of bias to the regression model, Ridge Regression reduces the standard errors, and it shrinks the least square coefficients towards the origin of the parameter space [10]. Ridge formula is: $R = \text{Min}(\text{sum of squared residuals} + \alpha * \text{slope}^2)$ Where $\text{Min}(\text{sum of squared residuals})$ is the Least Squared Error, and $\alpha * \text{slope}^2$ is the penalty term that Ridge adds to the Least Squared Error. When Least Squared Error determines the values of parameters, it minimises the sum of squared residuals. However, when Ridge determines the values of parameters, it reduces the sum of squared residuals. It adds a penalty term, where α determines the severity of the penalty and the length of the slope. In addition, increasing the α makes the slope asymptotically close to zero. Like Lasso, α is determined by applying the Cross validation method. Therefore, Ridge helps to reduce variance by shrinking parameters and make the prediction less sensitive.

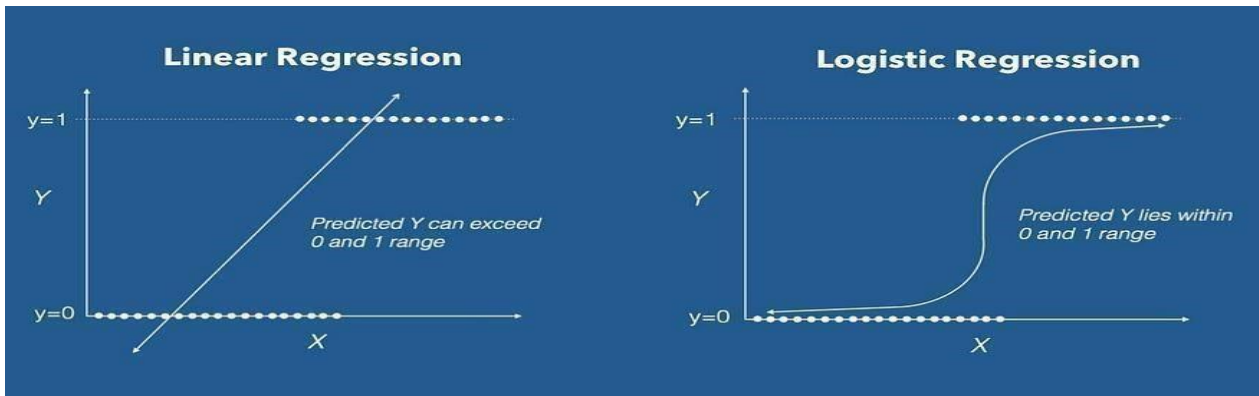
REGRESSION

In A Random Forest is an ensemble technique qualified for performing classification and regression tasks with the help of multiple decision trees and a method called Bootstrap Aggregation known as Bagging Decision Trees are used in classification and regression tasks, where the model (tree) is formed of nodes and branches. The tree starts with a root node, while the internal nodes correspond to an input attribute. The nodes that do not have children are called leaves, where each leaf performs the prediction of the output variable.

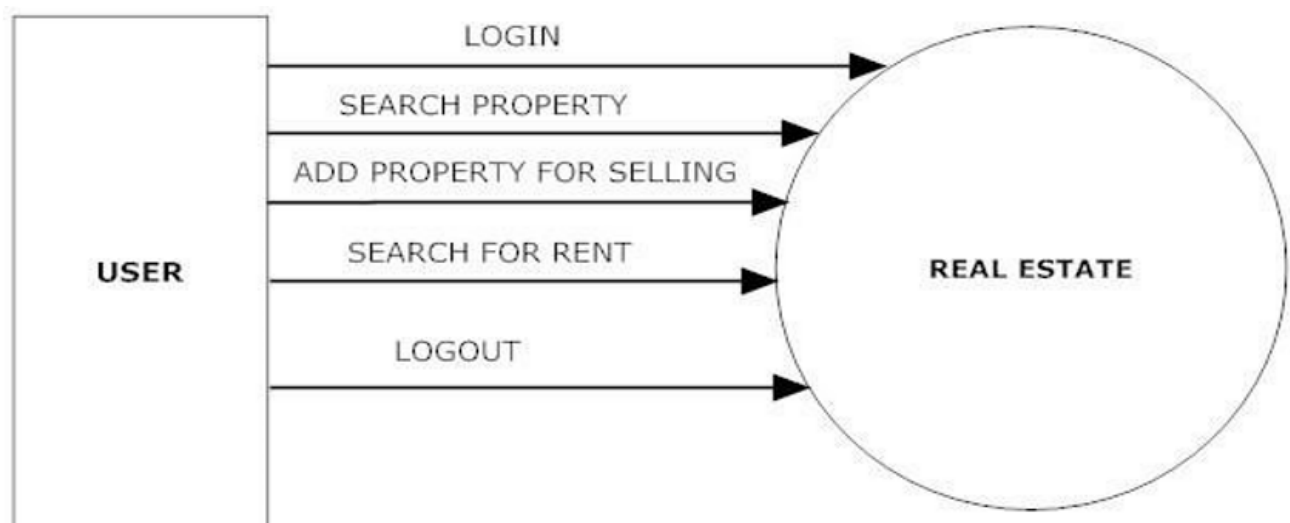


LOGISTICS REGRESSION

Logistic regression (or logit regression) is estimating the parameters of a logistic model (the coefficients in the linear combination). Formally, in binary logistic regression there is a single binary dependent variable, coded by an indicator variable, where the two values are labeled "0" and "1", while the independent variables can each be a binary variable (two classes, coded by an indicator variable) or a continuous variable (any real value). The corresponding probability of the value labeled "1" can vary between 0 (certainly the value "0") and 1 (certainly the value "1"), hence the labeling; the function that converts logodds to probability is the logistic function



SIMPLE SUBTYPE WORKING MODEL.



CONCLUSIONS

Efficient pricing of real estate customers for their priorities and budgets must therefore be predicted.

This project efficiently analyzes past industry trends and price ranges to predict future prices. Finally, Decision Tree using is more concentrated on generating rules, it works well for this dataset. Using rules, we can predict the value of the house will result in profit or loss.

REFERENCES

- R. Gupta, A. Kabundi, and S. M. Miller, "Forecasting the US real house price index: Structural and non-structural models with and without fundamentals," *Econ. Model.*, vol. 28, no. 4, pp. 2013–2021, Jul. 2011.
- J C. R. Madhuri, G. Anuradha, and M. V. Pujitha, "House price prediction using regression techniques: A comparative study," in *Proc. Int. Conf. Smart Struct. Syst. (ICSSS)*, Mar. 2019, pp. 1–5
- Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008