House Price Prediction Based on Machine Learning

Aditya Wankhade¹, Umesh Shinde², Hrushikesh Wayal³, Prof.Neelam Gaikwad

Department of Computer Engineering
Genba Sopanrao Moze College of Engineering, Balewadi, Pune 45

---***--

ABSTRACT

House price forecasting is an important topic of real estate. The literature attempts to derive useful knowledge from historical data of property markets. Machine learning techniques are applied to analyse. Historical property transactions in Bangalore City to discover useful models for house buyers and sellers. Moreover, experiments demonstrate that the Multiple Linear Regression that is based on mean squared error measurement is a competitive approach.The project aims at building a model of housing prices to predict median house value in Bangalore City using the provided dataset this model should learn from the data and be able to predict the median housing price in any district given all the other metrics.

Keywords: - Multiple linear regression, Lasso Regression, Random Forest , Decision Tree, Machine Learning, HousePrice Prediction.

1. INTRODUCTION

Machine learning is a subfield of Artificial Intelligence (AI) that works with algorithms and technologies to extract useful information from data. Machine learning methods are appropriate in big data since attempting to manually process vast volumes of data would be impossible without the support of machines. Machine learning in computer science algorithms are used to solve problems in the real world today. However, some of them give better performance in certain circumstances. Thus, this thesis attempts to use regression algorithms and when it comes to predicting valuesof a given dataset. The performance will be measured upon predicting house prices since the prediction in many regression algorithms relies not only on a specific feature but on an unknown number of attributes that result in the value to be predicted. House prices depend on an individual house specification. Houses have a variant number of features that may not have the same cost due to its location. For instance, a big house may have a higher price if it is located in desirable rich area than being placed in a poor neighbourhood. The data used in the experiment will be handled by using a combination of pre-processing methods to improve the prediction accuracy. In addition, some factors will be added to the local dataset in order to study the relationship between these factors.

2. LITERATURE SURVEY

Machine learning is a form of artificial intelligence which compose available computers with the efficiency to be trained without being veraciously programmed. Machine learning interest on the extensions of computer programs which is capable enough to modify when unprotected to new-fangled data. Machine learning algorithms are broadly classified into three divisions, namely; Supervised learning, Unsupervised learning and Reinforcement learning. Supervised learning is a learning in which we teach or train the machine using data which is well labelled that means some data is already tagged with correct answer. After that, machine is provided with new set of examples so that supervised learning algorithm analyses the training data and produces a correct outcome from labelled data. Unsupervised learning is the training of machine using information that is neither classified nor labelled and allowing the algorithm to act on that information without guidance. Here the task of machine is to group unsorted information according to similarities, patterns and differences without any prior training of data. Unlike, supervised learning, no teacher is provided that means no training will be given to the machine. Therefore, machine is restricted to find the hidden structure in unlabelled data by our-self. Machine learning has many application's out of which one of the applications is prediction of real estate. The real estate market is one of the most competitive in terms of pricing and same tends to be vary significantly based on lots of factor, forecasting property price is an important modules in decision making for both the buyers and investors in attempts to solve problems algorithmically rather than purely mathematically. Therefore, it is based on creating algorithms that permit the machine to learn. However, Supervised is where the program gets trained on pre-determined set to be able to predict when a new data is given. Several Machine Learning.

Supporting budget allocation, finding property finding stratagems and determining suitable policies hence it becomes one of the prime fields to apply the concepts of machine learning to optimize and predict the prices with high accuracy. The study on land price trend is felt important to support the decisions in urban planning. The real estate system is an unstable stochastic process. Investors decisions are based on the market trends to reap maximum returns. Developers are interested to know the future trends for their decision making. To accurately estimate property prices and future trends, large amount of data that influences land price is required for analysis, modelling and forecasting. The factors that affect the land price have to be studied and their impact on price has also to be modelled. An analysis of the past data is to be considered. It is inferred that establishing a simple linear mathematical relationship.

Hence it became imperative to establish a non-linear model which can well fit the data characteristic to analyse and forecast future trends. As the real estate is fast developing sector, the analysis and forecast of land prices using mathematical modelling and other scientific techniques is an immediate urgent need for decision making by all those concerned. The increase in population as well as the industrial activity is attributed to various factors, the most prominent being the recent spurt in the knowledge sector viz. Information Technology (IT) and Information technology enabled services. Demand for land started of showing an upward trend and housing and the real estate activity started booming. All barren lands and paddy fields ceased their existence to pave way for multistore and highrise buildings.

Investments started pouring in Real estate Industry and there was no uniform pattern in the land price over the years. The need for predicting the trend in land prices was felt by all in the industry viz. the Government, the regulating bodies, lending institutions, the developers and the investors. Therefore, in this paper, we present various important features to use while predicting housing prices with good accuracy. We can use regression models, using various features to have lower Residual Sum of Squares error. While using features in a regression model some feature engineering is required for better prediction. Often a set of features multiple regressions or polynomial regression (applying a various set of powers in the features) is used for making better model fit. For these models are expected to be susceptible towards over fitting ridge regression is used to reduce it. So, it directs to the best application of regression models in addition to other techniques to optimize the result.

linear regression based on hedonic pricing. Previous research conducted by Gharehchopogh using linear regression approach get 0,929 errors with the actual price. In linear regression, determining coefficients generally using the leastsquare method, but it takes a long time to get the best formula.

3. OBJECTIVES OF SYSTEM

As a first project, we intended to make it as instructional as possible by tackling each stage of the machine learning process and attempting to comprehend it well. We have picked Bangalore Real Estate Prediction as a method, which is known as a "toy issue," identifying problems that are not of immediate scientific relevance but are helpful to demonstrate and practice. The objective was to forecast the price of a specific apartment based on market pricing while accounting for various "features" that would be established in the following sections.

4. EXISITING SYSTEM

There are several approaches that can be used to determine the price of the house, one of them is the prediction analysis. The first approach is a quantitative prediction. A quantitative approach is an approach that utilizes timeseriesdata. The time-series approach is to look for the relationship between current prices and prevailing prices. The secondapproach is to use .

5. IMPLEMENTATION DETAILS

1. Import Libraries
2. Load Dataset
3. Exploratory Data Analysis
4. Data Cleaning
5. Feature Engineering
6. Dimensionality Reductions
7. Outlier Removal using Business Logic
8. Outlier Removal using Standard Deviation & Mean
9. Data Visualization
10. Building a Model
11. Test the Model for few properties
12. Export the tested model to a pickle file

6. CONCLUSION

The goal is to achieve the system which will reduce the human effort to find a house having reasonable price. The proposed system. House Price Prediction model approximately try to achieve the same one. Proposed system focusedon predict the house price according to the area for that image processing and machine learning methods are used. The experimental results showed that this technique that are used while developing system will give accurate prediction of house price.

7. REFERENCES

[1] Model "BANGALORE HOUSE PRICE PREDICTION MODEL"

[2] Heroku "Documentation"

[3] Repository: "Web Application" https://github.com/msatmod/Bangalore-HousePrice-Prediction.

[4] Repository: "Web Application" https://github.com/Amey-Thakur/BANGALOREHOUSE-PRICE-PREDICTION.

[5] A. Varma, A. Sarma, S. Doshi and R. Nair, "House Price Prediction Using Machine Learning and Neural Networks,"2018 Second International Conference on

[6] Inventive Communication and Computational Technologies (ICICCT), 2018, pp. 1936-1939, doi: 10.1109/ICICCT.2018.8473231.

[7] Furia, Palak, and Anand Khandare. "Real Estate Price Prediction Using Machine Learning Algorithm." eConference on Data Science and Intelligent Computing. 2020.

[8] Middleton, Neil, and Richard Schneeman. Heroku: up and running: effortless application deployment and scaling. " O'Reilly Media, Inc.", 2013

[9] Aggarwal, Shalabh. Flask framework cookbook. Packt Publishing Ltd, 2014.

[10] Grinberg, Miguel. Flask web development: developing web applications with python. " O'Reilly Media, Inc.", 2018.

[11] The Danh Phan, "Housing Price Prediction using Machine Learnings Algorithms: The Case of Melbourne City, Australia", 2018.

[12] Sifei Lu, Zengxiang Li, Zheng Qin, Xulei Yang, Rick Siow Mong Goh Institute of High Performance Computing (IHPC), Agency for Science Technology and Research, Singapore, "A Hybrid Regression Technique for House PricesPrediction", 2018.

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

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

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

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

[17] Gaurav Kumar, Pradip Kumar Bhatiya, Fourth International Conference on Advanced Computing & CommunicationTechnologies, "A Detailed Review of Feature Extraction in Image Processing Systems", 2014.

[18] Guangliang Gao, Zhifeng Bao, Jie Cao, A. K. Qin, Timos Sellis, Zhiang Wu, "Location-Centered House PricePrediction: A Multi-Task Learning Approach", 2