International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

House Price Prediction System for Bengaluru City

Prof.USHA C[1*], Madhurima ST[2*]
[1]Assistant Professor, Department of Master of Applications,
University B.D.T College of Engineering, Davanagere, Karnataka, India
[2]Student, Department of Master of Applications,
University B.D.T College of Engineering, Davanagere, Karnataka, India

Abstract- In recent years, Karnataka, one of the hotspots for real estate development, has seen an increase in demand from potential home buyers and investors, and is expected to witness a further boom in the sector by 2020. What are the things that a potential home buyer considers before purchasing a house? The location, the size of the property, vicinity to offices, schools, parks, restaurants, hospitals or the stereotypical white picket fence? What about the most important factor — the price? Now with the lingering impact of COVID-19, the enforcement of the Real Estate (Regulation and Development) Act (RERA), and the lack of trust in property developers in the city, housing units sold across India in 2019 dropped by 3 per cent. In fact, the property prices in Bengaluru fell by almost 5 per cent in the second half of 2019, said a study published by property consultancy Knight Frank. Buying a home, especially in a city like Bengaluru, is a tricky choice. While the major factors are usually the same for all metros, there are others to be considered for the Silicon Valley of India. With its help millennial crowd, vibrant culture, great climate and a slew of job opportunities, it is difficult to ascertain the price of a house in Bengaluru. This paper reflects the effort towards solving the problems mentioned. In this paper, the authors have tried to create such a system which will in turn, give a very accurate prediction about the prices of the house in the city of Bengaluru. The authors have tried to create a user friendly interface design which will enable the users to choose their options as per their requirements and get the estimated price of the house according to their needs.

Keywords- Machine Learning, Linear Regression, Prediction System, Accuracy, ML Model

I. INTRODUCTION

Buying a house is a stressful thing. One has to pay huge sums of money and invest many hours and even there is a persisting concern whether it’s a good deal or not. Buyers are generally not aware of factors that influence the house prices. Almost all the houses are described by the total area in square foot, the neighbourhood and the number of bedrooms. Sometimes houses are even priced at X rupees per square foot. This creates an illusion that house prices are dependent almost solely on the above factors. Most of the houses are bought though real estate agents. People rarely buy directly from the seller, since there are a lot of legal terminology and paperwork’s involved and people are unaware of them. Hence real estate agents are trusted with the communication between buyers and sellers as well as laying down a legal contact for the transfer. This just creates a middle man and increase the cost of the house. Therefore the houses are overpriced and buyer should have a better idea of the actual value of the houses.

This paper covers the measures that have been taken by the use of technology that is accessible and utilize them to create an unbiased system for predicting the house prices. The authors have utilized the previously gathered data from trusted resources and trained and designed a Machine Learning model in such a way that it provides the best possible predictions of house prices as an output to the user. Thus, this method in which the complete prediction is based upon the previously gathered data, the integrity and trustfulness of the system to the user is maintained.

II. LITERATURE SURVEY

In [1] [MLR]Multiple Linear Regression is used which uses more than one attributes for prediction.This article refers together with latest Forecast on Research predictions considering trends to further plan their economics.

In [2],[9] [RR]Ridge and[LR] LASSO Regressions are used in which Ridge regression regularizes the [rc] regression coefficient by posing a interest on the size. [LR]LASSO Regression is also same to Ridge but with a little difference, it uses the L1 penalty.

In [3] the [ER]Elastic Net Regression was used as a penalization method.

In [4] the classification algorithm Naive Bayes is used. In [5] one of the most efficient Regressions i.e.,

[GBR]Gradient Boosting Regression is used.

In [6],[10],[13] the Artificial Neural Network theory is used. Hedonic Pricing theory is used in [7],[11] which assumes the property value as the sum of its attribute values.

Linear Regression model is used in [8][12].In [14][16] Geographically Weighted Regression is used which allows local variations in rate. In [15][17] Bayesian Linear Regression is used.

In this paper we build on previous empirical studies by comparing the econometric Bayesian Vector Autoregressive (BVAR) and Bayesian Autoregressive (BAR) models, instead of Bayesian predictive regressions to avoid issues of endogeneity, with a novel forecasting methodology on one- year-ahead forecasting. We propose a methodology that combines Ensemble Empirical Mode Decomposition from the field of signal processing with the machine learning Support Vector Regression methodology for constructing forecasting models.

Given the existence of non-normality and nonlinearity in the data generating process of real house price returns over the period of 1831–2013, this article compares the ability of various univariate copula models, relative to standard benchmarks (naive an autoregressive models) in forecasting real US house price over the annual out-of-sample period of 1874–2013, based on an in-sample of 1831–1873. Overall, our results provide overwhelming evidence in favor of the copula models (Normal, Student's t, Clayton, Frank, Gumbel, Joe and Ali-Mikhail-Huq) relative to linear benchmarks, and especially for the Student's t-copula, which outperforms all other models both in terms of in-sample and out-of-sample predictability results. Our results highlight the importance of accounting for nonnormality and nonlinearity in the data generating process of real house price returns for the US economy for nearly two centuries of data.

I. PROBLEM STATEMENT

Machine learning has been used for years to offer image recognition, spam detection, natural speech comprehension, product recommendations, and medical diagnoses. Today, machine learning algorithms can help us enhance cyber security, ensure public safety, and improve medical outcomes. Machine learning systems can also make customer service better and automobiles safer .When we started experimenting with machine learning, we wanted to come up with an application that would solve a real-world problem but would not be too complicated to implement. We also wanted to practice working with regression algorithms.So I started looking for a problem worth solving. Here's what we came up with. If you're going to sell a house, you need to know what price tag to put on it. And a computer algorithm can give you an accurate estimate! With the given features (categorical and continuous) build a model to predict the price of houses in Bengaluru.

I. PROPOSED SOLUTION

Nowadays, e-education and e-learning is highly influenced. Everything is shifting from manual to automated systems. The objective of this project is to predict the house prices so as to minimize the problems faced by the customer. The present method is that the customer approaches a real estate agent to manage his/her investments and suggest suitable estates for his investments. But this method is risky as the agent might predict wrong estates and thus leading to loss of the customer's investments. The manual method which is currently used in the market is out dated and has high risk. So as to overcome this fault, there is a need for an updated and automated system. Data mining algorithms as well as Machine Learning algorithms can be used to help investors to invest in an appropriate estate according to their mentioned requirements. Also the new system will be cost and time efficient .This will have simple operations. In our project, the proposed system works on Linear Regression Algorithm. In today's real estate world, it has become tough to store such huge data and extract them for one's own requirement. Also, the extracted data should be useful. The system makes optimal use of the Linear Regression Algorithm. The system makes use of such data in the most efficient way.

In this paper, the linear regression algorithm helps to full fill customers by increasing the accuracy of estate choice and reducing the risk of investing in an estate. A lot of features that could be added to make the system more widely acceptable.

We have applied very efficient and logical feature extraction techniques so as to increase the accuracy. For example, we have done removal of outliers by using the business logic and bathroom feature. We know that if someone is going to buy a house of say 2000 sqft, then he should have at least 3 to 4 bedrooms. Any other cases in which there are only 2 rooms in 2000 sq ft, has been removed as an outlier. Another feature is the removal of cases where there are absurd numbers of bathrooms. By our logic, the total number of bathrooms should be at most 1 more than total number of bedrooms, i.e. total bath=total BHK+1. Therefore, we removed any other cases from the data frame where the cases were contradictory.

One of the major future scopes is adding estate database of more cities which will provide the user to explore more estates and reach an accurate decision.

DATASET

There's a lot of data involved in fully training the model. The dataset is kept under a same directory. All pre-preprocessing scripts will, by default, output the clean data to a new directory created in the datasets root directory.

The following dataset has been used:

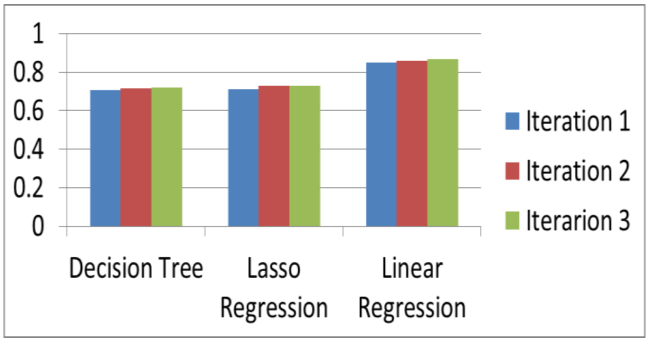
• Bengaluru_house_prices.csv
Link:http://www.kaggle.com/dataset/bengaluru_house_price/

This dataset was prepared as a record for the house prices of different houses at different locations in the city of Bengaluru by various government authorities. This dataset is a large collection of over 13321 records and 9 columns of house price data collected by various trusted sources.

II. METHODOLOGY

1. Pre-Processing and Data Cleaning

Data preprocessing is an integral step in Machine Learning as the quality of data and the useful information that can be derived from it directly affects the ability of our model to learn; therefore, it is extremely important that we preprocess our data before feeding it into our model.Feature Engineering



2. Feature Engineering

Feature engineering is the process of using domain knowledge of the data to create features that make machine learning algorithms work. If feature engineering is done correctly, it increases the predictive power of machine learning algorithms by creating features from raw data that help facilitate the machine learning process. Feature Engineering is an art .In our project, it includes exploring the total_sqft feature and also adds new feature price per square feet.

3. Dimensionality Reduction and Outlier Removal

Dimensionality reduction refers to techniques for reducing the number of input variables in training data. Fewer input dimensions often mean correspondingly fewer parameters or a simpler structure in the machine learning model, referred to as degrees of freedom. In our project, any location which had number of houses less than 10 has been marked as "others" so as to reduce the dimensions of the dataset.

Outliers badly affect mean and standard deviation of the dataset. These may statistically give erroneous results. It increases the error variance and reduces the power of statistical tests. If the outliers are non-randomly distributed, they can decrease normality. So we applied various logics such as business logic, bathroom feature to remove the outliers.

4. Model Building and Accuracy

In our project, the model was implemented using the Linear Regression Algorithm. All the necessary libraries were imported and training of the model was done. We saw that in 5 iterations we get a score above 85% all the time. This was a very good accuracy score and we continued to use the algorithm. Also we compared different algoritms, such as Lasso Regression, Decision Tree and Linear Regression using the GridSearchCV technique to find the model with best accuracy, which we found that it is Linear Regression

III. ACKNOWLEDGMENT

We would like to thank our mentors and all the respective people whose insights helped us to make this project possible. We would like to extend our gratitude to all staff of Department of Computer Science and Engineering for the help and support rendered to us. We have benefited a lot

from the feedback, suggestions given by them. We would also like to give a special mention to Prof. Abdul Razak M S and Dr. Nirmala C R from the department of Computer Science and Engineering in BIET, Davangere for their constant mentoring and guidance.

IV. CONCLUSION

The framework makes ideal utilization of the Linear Regression Algorithm. It makes use of such information in the most effective way. The direct relapse calculation satisfies customer by expanding the exactness of their decision and diminishing the danger of putting resources into a home. One of the real future extensions is including home database of more urban areas which will give the client to investigate more domains and achieve an exact choice. More factors like subsidence that influence the house costs should be included. Top to bottom subtle elements of each property will be added to give plentiful points of interest of a coveted domain. The authors were able to create a system with more than 85% accuracy and the utilization of dataset was done with great efficiency which ultimately gave quite impressive results.

REFERENCES

[1] R Manjula, Shubham Jain, Sharad Srivastava and Pranav Rajiv Kher, ―Real estate value prediction using multivariate regression models,‖ IOP Conference Series: Materials Science and Engineering, 2017.

[2] Eduard Hromada, ―Mapping of real estate prices using data mining techniques,‖ Czech Technical University, Czech Republic, 2015

[3] Adyan Nur Alfiyatin and Ruth Ema Febrita, ―Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization,‖ International Journal of Advanced Computer Science and Applications, 2017

[4] Li Li and Kai-Hsuan Chu, ―Prediction of Real Estate Price Variation Based on Economic Parameters,‖ Department of Financial Management, Business School, Nankai University, 2017.

[5] Nissan Pow, Emil Janulewicz and Liu Dave, ―Applied Machine Learning Project 4 Prediction of real estate property prices in Montreal,‖ 2016.

[6] Aminah Md Yusof and Syuhaida Ismail, Multiple Regressions in Analysing House Price Variations. IBIMA Publishing Communications of the IBIMA Vol. 2012 (2012), Article ID 383101, 9 pages DOI: 10.5171/2012.383101.

[7] Babyak, M. A. What you see may not be what you get: A brief, nontechnical introduction to over fitting regression-type models. Psychosomatic Medicine, 66(3), 411-421.

[8] Vasilios Plakandaras and Theophilos, Rangan Gupta*, Periklis Gogas
―Forecasting the U.S. Real House Price Index‖.

[9] Rangan Gupta ―Forecasting US real house price returns over 1831– 2013: evidence from copula models‖

[10] Valeria Fonti, Feature Selection using LASSO Research Paper in Business Analytics, VU Amsterdam, March 30, 2017.

Nihar Bhagat, Ankit Mohokar, Shreyash House Price