Real Estate House Price Prediction System

Prof. Sushen Gulhane[1], Rushikesh Kotkar[2], Joshua Naik[3], Mahesh Shinde[4], Akash Pinate[5]
[1]Assistant Professor, Dept of Information Technology
[2, 3, 4, 5]Dept of Information Technology
[1, 2, 3, 4, 5]DYPCOE, Ambi, Talegaon, India

Abstract- *In recent years, Karnataka, one all told the hotspots for realty development, has seen an increase in demand from potential home buyers and investors, and is predicted to witness an additional boom within the arena by 2020. What are the things that a possible vendee considers before purchasing a house? things, the size of the property, vicinity to offices, schools, parks, restaurants, hospitals or the stereotypical white picket fence? What about the foremost important factor — the price? Now with the lingering impact of COVID-19, the enforcement of the 000 Estate (Regulation and Development) Act (RERA), and so the dearth of trust in property developers within the town, 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 within the half 2019, said a study published by property consultancy Knight Frank. Buying a home, especially in an exceedingly city like Bengaluru, could also be a difficult choice. While the foremost important factors are usually the identical for all metros, there are others to be considered for the region of India. With its help millennial crowd, vibrant culture, great climate and a slew of job opportunities, it's difficult to determine the worth of a house in Bengaluru. This paper reflects the difficulty towards solving the problems mentioned. during this paper, the authors have tried to create such a system which might successively, provides a awfully accurate prediction about the prices of the house within town of Bengaluru. The authors have tried to create a user friendly interface design which may enable the users to decide on their options as per their requirements and find the estimated price of the house per their needs.*

Keywords- Machine Learning, statistical procedure, 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 an honest deal or

Buyers are generally not tuned in to factors that influence the house prices. most the homes are described by the full area in square foot, the neighbourhood and also the amount of bedrooms. Sometimes houses are even priced at X rupees per area unit. This creates an illusion that house prices are dependent almost solely on the above factors. Most of the homes are bought though realty agents. People rarely buy directly from the seller, since there are many legal

terminology and paperwork's involved and folk are unaware of them. Hence assets agents are trusted with the communication between buyers and sellers additionally as laying down a legal contact for the transfer. This just creates a middle man and increase the worth of the house. Therefore the homes are overpriced and buyer should have an improved idea of the actual value of the homes.

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

II. LITERATURE SURVEY

In [1] [MLR]Multiple regression toward the mean is used which uses quite 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 employed during which Ridge regression regularizes the [rc] statistic by posing a interest on the dimensions. [LR]LASSO Regression is additionally same to Ridge but with a small amount 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 altogether the foremost efficient Regressions i.e., [GBR]Gradient Boosting Regression is utilized.

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

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

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

Given the existence of non-normality and nonlinearity within the information generating process of real house price returns over the quantity of 1831–2013, this text compares the pliability of assorted univariate copula models, relative to plain benchmarks (naive an autoregressive models) in forecasting real United States House of Representatives of Representatives price over the annual out-of-sample period of 1874–2013, supported 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 particularly 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 within the info generating process of real house price returns for the US economy for nearly two centuries of data.

III. PROBLEM STATEMENT

Machine learning has been used for years to supply 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 also can make customer service better and automobiles safer .When we started experimenting with machine learning, we wanted to return up with an application that might solve a real-world problem but wouldn't be too complicated to implement. We also wanted to practice working with regression algorithms. So I started trying to find an issue worth solving. Here's what we came up with. If you're visiting sell a house, you would like to understand what tag to place on that. And a computer algorithm can offer you an accurate estimate! With the given

features (categorical and continuous) build a model to predict the worth of homes in Bengaluru.

IV. PROPOSED SYSTEM SOLUTION

Nowadays, e-education and e-learning is extremely influenced. Everything is shifting from manual to automated systems. the target of this project is to predict the house prices so on minimize the problems faced by the customer. the present method is that the customer approaches a real land agent to manage his/her investments and suggest suitable estates for his investments. But this method is risky because the agent might predict wrong estates and thus leading to loss of the customer's investments. The manual method which is currently utilized within the market is out dated and has high risk. So on overcome this fault, there is a desire for an updated and automatic system. Processing algorithms additionally as Machine Learning algorithms are going to be accustomed help investors to require a foothold in an appropriate estate in line with their mentioned requirements. Also the new system are cost and time efficient .This will have simple operations. In our project, the proposed system works on regression Algorithm. In today's realty world, it's 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 regression Algorithm. The system makes use of such data within the foremost effective way. In this project XG booster helps to predict all Evaluation of houses.

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 assorted houses at different locations within the town of Bengaluru by various government authorities. This dataset is also an outsized collection of over 13321 records and 9 columns of house price data collected by various trusted sources. It consists of the following features: area, type, availability, loaction, size, society, total_ sqft, bath, balcony, price.

In these features, the price column is that the labelled attribute.

IV. METHODOLOGY

1.   Pre-Processing and Data Cleaning

Data preprocessing is an integral step in Machine Learning because the standard of data and also the useful information which can be derived from it directly affects the

ability of our model to learn; therefore, it's extremely important that we preprocess our data before feeding it into our model.

Feature Engineering Feature engineering is that the method of using domain knowledge of the information to create features that make machine learning algorithms work. If feature engineering is finished correctly, it increases the predictive power of machine learning algorithms by creating features from information 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.

2. Feature Engineering

Algoritms, like Lasso Regression, Decision Tree and easy regression using the GridSearchCV technique to go looking out the model with best accuracy, which we found that it's regression toward the mean

3. Dimensionality Reduction and Outlier Removal

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

Outliers badly affect mean and variance 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'll decrease normality. So we applied various logics like business logic, bathroom feature to induce eliminate the outliers.

4. Model Building and Accuracy

In our project, the model was implemented using the statistical regression Algorithm. All the specified libraries were imported andtraining of the model was done. We saw that in 5 iterations we get a score above 85% all the time. This was a extremely good accuracy score which we continued to use the algorithm.

VI. CONCLUSION

The framework makes ideal utilization of the easy regression Algorithm Which is XG BOOSTER.It makes

use of such information within the only 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 in every of the 000 future extensions is including home database of more urban areas which might give the client to investigate more domains and achieve an actual choice. More factors like subsidence that influence the house costs should be included. Top to bottom subtle elements of every property are added to present plentiful points of interest of a coveted domain. The authors were ready to create a system with over 85% accuracy and therefore the utilization of dataset was through with great efficiency which ultimately gave quite impressive results

VII. ACKNOWLEDGMENT

REFERENCES

[1] R Manjula, Shubham Jain, Sharad Srivastava andproperty [1] property prices in Montreal,‖ 2016.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.

[2] Babyak, M. A. What you see won't be what you get: a brief, nontechnical introduction to over fitting regressiontype models. Psychosomatic Medicine, 66(3), 411-421.

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

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

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

[6] Nihar Bhagat, Ankit Mohokar, Shreyash House Price Forecasting using processing.International Journal of Computer Applications 152(2):23-26, October 2016.

[7] Model,Azme Bin Khamis, Nur Khalidah Khalilah Binti Kamarudin, Comparative Study On Estimate House Price Using Statistical And Neural Network, International journal of scientific and technology, research volume 3,

ISSUE 12, December 2014,Page(s):126-131. Pranav Rajiv Kher, ―Real estate value prediction using multivariate regression models,‖ IOP Conference Series: Materials Science and Engineering, 2017.

[8] Eduard Hromada, ―Mapping of assets prices using data processing techniques,‖ Czech Technical University, European country, 2015

[9] Adyan Nur Alfiyatin and Ruth EmaFebrita, ―Modeling House Price Prediction using statistical method and Particle Swarm Optimization,‖ International Journal of Advanced computing and Applications, 2017

[10] Li Li and Kai-Hsuan Chu, ―Prediction of property Price Variation supported Economic Parameters,‖ Department of monetary Management, grad school, Nankai University, 2017.

[11] Nissan Pow, Emil Janulewicz and Liu Dave, ―Applied Machine Learning Project 4 Prediction of