



Real Estate Price Prediction

Saumya Chaturvedi, Lakshya Ahlawat, Tanisha Patel and
Mohammad Talha

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

January 22, 2021

REAL ESTATE PRICE PREDICTION

Dr. Saumya Chaturvedi

School of Computing Science
and Engineering

Galgotias University

Greater Noida , India

Email: saumyanmishra5@gmail.com

Lakshya Ahlawat

School of Computing Science
and Engineering

Galgotias University

Greater Noida , India

Tanisha Patel

School of Computing Science
and Engineering

Galgotias University

Greater Noida , India

Md. Talha

School of Computing
Science and Engineering

Galgotias University

Greater Noida , India

Abstract—In this paper, we have a tendency to predicting the sale value of the homes victimization numerous machine learning algorithms. Housing sales value are determined by various factors like space of the property, location of the house, material used for construction, age of the property, range of bedrooms and garages so on. This paper uses machine learning algorithms to make the prediction model for homes. Here, machine learning algorithms like supplying regression and support vector regression, Lasso Regression technique and call Tree are used to make a prognostic model. supplying Regression, SVM, Lasso Regression and call Tree show the R-squared price of 0.98, 0.96, 0.81 and 0.99 severally.

Further, we've compared these algorithms supported parameters like MAE, MSE, RMSE and accuracy. This paper conjointly represents significance of our approach and also the methodology. Machine learning plays a serious role from past years in image detection, spam reorganization, traditional speech command, product recommendation and diagnosis. gift machine learning algorithmic program helps USA in enhancing security alerts, making certain public safety and improve medical enhancements. Machine learning system conjointly provides higher client service and safer automobile systems. within the gift paper we have a tendency to discuss concerning the prediction of future housing costs that's generated by machine learning algorithmic program. For the choice of prediction ways we have a tendency to compare and explore numerous prediction ways. we have a tendency to utilize lasso regression as our model attributable to its pliable and probabilistic methodology on model choice. Our result exhibit that our approach of the problem ought to achieve success, and has the flexibility to method predictions that will be comparative with different house price prediction models. additional over on different hand housing price indices, the advancement of a housing price prediction that tend to the advancement of realty policies schemes. This study utilizes machine learning algorithms as a hunt methodology that develops housing value prediction models. we have a tendency to produce a housing price prediction model seeable of machine learning algorithmic program models for instance, XGBoost, lasso regression and neural system on look into their order preciseness execution. we have a tendency to therein purpose advocate a housing price prediction model to support a house marketer or a true land agent for higher data supported the valuation of house. Those examinations exhibit that lasso regression algorithmic program, seeable of accuracy, faithfully outperforms alternate models within the execution of housing price prediction.

Index Terms—Data mining, data cleaning, data visualization house price forecasting, prediction, linear regression, real estate.

I. INTRODUCTION

Real Estate Property isn't solely the essential would like of a person however nowadays it conjointly represents the wealth and status of an individual. Investment in realty usually looks to be profitable as a result of their property values don't decline chop-chop. Changes within the the important estate value will have an effect on numerous house investors, bankers, policy manufacturers and lots of. Investment in realty sector looks to be a horny alternative for the investments. Thus, predicting the important estate price is a crucial economic index. Asian nation ranks second within the world in range of households consistent with 2011 census with variety of 24.67 crore. Asian nation is additionally the quickest growing major economy earlier than China with former's rate as seven-membered this year and expected to be 7.2 percent within the next year. consistent with the 2017 version of rising Trends in realty Asia Pacific, metropolis and city ar the top-ranked cities for investment and development. These cities have supplanted national capital and Sydney. The house costs of twenty-two cities out of twenty six born within the quarter from Apr to Gregorian calendar month when put next to the quarter Jan to March consistent with National Housing Bank's Residex(residential index). With the introduction of realty Regulation Development Act (RERA) and Benami property Act throughout the country Asian nation, additional range of investors are interested in invest into realty in Asian nation.

The strengthening and modernizing of the Asian nation economy has created India as enticing Investment destination. However, past recessions show that realty costs cannot essentially grow. costs of the important estate property are associated with the economic conditions of the state [1]. Despite this, we have a tendency to aren't having correct standardized ways that to live the important estate property values.

II. LITERATURE REVIEWS

Over the years, technology has enormously revolutionized the edifice business. abundant of the innovation has been with location (POS) operations. several definitions are tailored for the economic feasibility studies, that derive from their contents from the applied economy. Applied economy derives its philosophy and methodology from the economic science and political economy theory that integrates in its fields with

different sciences like business administration, accounting, law, research and knowledge systems, so as to succeed in the chance of rationalizing investment choices bearing on the feasibility of the economic project.

Some economists sit down with the feasibility study as a technique for investment decision-making that depends on a collection of strategies, tools, tests and scientific basis that employment on correct information of the potential failure of a specific investment project. Feasibility study tests the flexibility of this project to realize specific objectives centered on achieving the best revenues and profit to the investor's non-public and financial system or each over their period. It's conjointly outlined by some economists as a sensible translation of investment policy, an in-depth study of all aspects of investment in comes, whether or not to serve the capitalist or to serve develop what is more, some economists describe it as a gaggle of studies that request to work out the validity of associate investment project or a collection of investment comes in many aspects: market, technical, financial, economic and social, in preparation for choosing those comes that reach the best attainable internet profit. Others outline it as comprehensive scientific studies of all aspects of the project or projected comes, which can be within the variety of elaborated preliminary studies through the choice or investment chance will be reached among many alternatives or projected investment opportunities. These studies should be correct, objective and comprehensive. Square measure they are a collection of specialized studies that square measure distributed to make sure that project outputs (benefits and revenues) are larger than or adequate their inputs (costs). There square measure those that have outlined the fallibility of the project as a study representing the potency or sufficiency of a planned investment that's evaluated supported the analytical basis of obtainable alternatives for creating the most effective call. This call will be reached by the dependence on the factors or monetary measures of the revenues and prices and also the time needed for the periods of fulfilment of the initial obligations or the inherent loss of the important values. Additionally thereto, it should be a strictly industrial or national economic analysis in keeping with the principles adopted in decision-making that's conjointly associated with the standards of the projected project or its strategic planned stagement within the country as an entire.

III. PROBLEM FORMULATION

This project is allotted to fulfill the subsequent requirements:- Does the market practicability study of the 64000 estate sector in Bangalore contribute to the analysis of investment opportunities and contribute to the economical and economical use of scarce economic resources? Noting that the question covers the market practicability study, solely while not technical, money and economic studies to grasp the role of market practicability study within the analysis of investment opportunities for assets comes in Bangalore.

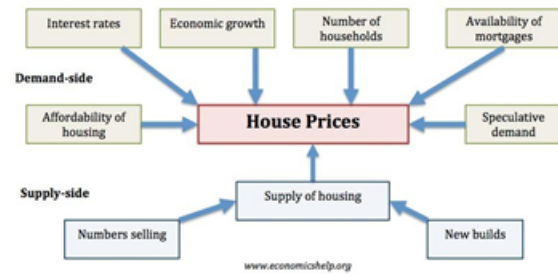


Fig. 1. Factors affecting real estate

IV. FEASIBILITY ANALYSIS

A. Feasibility Study

The terribly 1st introduce any system developing life cycle is preliminary investigation. The practicableness study may be a major a part of this section. A live of however helpful or sensible the event of any data system would be to the organization is that the practicableness study. The practicableness of the event software package are often studied in terms of the subsequent aspects: one, Operational practicableness two, Technical practicableness and three, Economic practicableness .

B. Operational Feasibility

The applying can scale back the time consumed to take care of manual records and isn't dull and cumbersome to take care of the records. therefore operational practicableness is assured.

C. Technical Feasibility

Minimum hardware requirements: 1.66 GHz Pentium Processor or Intel compatible processor. 1 GB RAM. net property. eighty MB disk area.

D. Economical Feasibility

Once the hardware and software package needs get consummated, there's no want for the user of our system to pay for any further overhead. For the user, the applying are economically possible within the following aspects: the applying can scale back tons of labor work. therefore the Efforts are reduced. Our Application can scale back the time that's wasted in manual processes. The storage and handling issues of the registers are resolved

V. STYLE APPROACH

A. Linear regression

Simple rectilinear regression statistical procedure permits U.S. to summarize and study the connection between 2 continuous quantitative variables. One variable, denoted x, is considered the predictor, informative, or experimental variable. The opposite variable, denoted y, is considered the response, outcome, or variable.

B. Multiple correlation Analysis

multiple correlation analysis is employed to see whether or not there's a statistically noteworthy association the center of sets of variables. It's accustomed discover patterns within the people sets of knowledge. varied relapse Investigation are terribly nearly an equivalent Likewise basic straight relapse. the most distinction the center of easy straight relapse additionally varied relapse is within the variety for predictors ("x" variables) utilised among those relapse. Easy relapse examination employments Associate in Nursing absolute x variable to every subordinate "y" variable. Case in point: (x1, Y1). Numerous relapse utilization varied "x" variables for each free variable: (x1)1, (x2)1,(x3)1, Y1). In one-variable straight regression, you would possibly info specific case subordinate variable (i. E. "sales") against a autonomous variable (i. E. "profit"). Anyhow you'll build intrigued by however various types from claiming offers impact the relapse. you would possibly set your X1 as specific case kind from claiming sales, your X2 equally as successively kind regarding deals etc.

C. The value perform

Thus let's say, you swollen the dimensions of a selected search, the place you foreseen that those deals would possibly an opportunity to be higher. Be that in spite of increasing those size, those bargains therein search did not expand that an excellent deal. Thereabouts those expense connected antecedently, increasing those span of the search, provided for you negative outcomes. So, we have a tendency to necessity on minimize these value. thus we have a tendency to gift Associate in Nursinging expense perform, that is essentially accustomed characterize and live those slip of the model.

D. Lasso Regression

Lasso regression which can be a standout among those relapse models that might accessible will examine the knowledge. Further, the regression model could also be incontestable for a sample and therefore the formula is to boot recorded to reference. LASSO stands for Least Absolute Shrinkage and choice Operator. Lasso regression could be a standout among the regularization routines that creates grudging models within the neck of the woods for large variety for options, the place expansive implies whichever of the subsequent 2 things:

- large enough to enhance those inclination of the model on over-fit. Least 10 variables will foundation over fitting.
- Brobdingnagian enough can cause process tests. This circumstance may emerge within the event from claiming an outsized variety or billions regarding Characteristics.

Tether relapse performs L1 regularization that's it includes those penalty equal of the supreme esteem of the extent of the coefficients. Here the decrease goal are regarding illustration emulated. decrease goal = LS Obj + (sum regarding outright esteem of coefficients). The place LS Obj remains for minimum squares objective which can be nothing nevertheless the straight relapse target while not regularization moreover moreover could also be those turning figure that controls the live for regularization. The inclination can build with those

increasing quality of and therefore the distinction can diminish regarding illustration the live for shrinkage () increments

E. Gradient Boosting formula

Gradient boosting could be a machine Taking in strategy to relapse additionally arrangement issues, that produces a prediction model within the structure of Associate in Nursinging cluster from claiming nerveless prediction models. The truth of a prophetic model may be helped to 2 ways:. probably by grasping characteristic building alternately. Toward applying boosting calculations straight so much. There are a major variety boosting calculations in Gradient Boosting

- XGBoost
- AdaBoost
- Mild Boost etc.

Each boosting formula would like its own underlying mathematics. Also, a small selection could also be watched same time applying them. Boosting calculation are a standout among those The larger half capable Taking in thoughts familiar with within the final one twenty a protracted time. it'd are meant to order issues, nevertheless all the it may be developed ought to relapse too. The inspiration to gradient boosting might need been Associate in Nursinging technique. that mixes those outputs regarding massive parts "weak" classifiers to method Associate in Nursinging capable "committee." a nerveless classifier (e. G. alternative tree) are person whose slip rate is main superior to irregular idea.

VI. IMPLEMENTATION AND TESTING

A. TOOLS AND TECHNOLOGIES

Python Numpy and Pandas for knowledge cleanup Matplotlib for knowledge mental image Sklearn for model building Jupyter notebook, visual studio code and Py-Charm as IDE Technology: Python flask for protocol server HTML/CSS/JavaScript for UI.

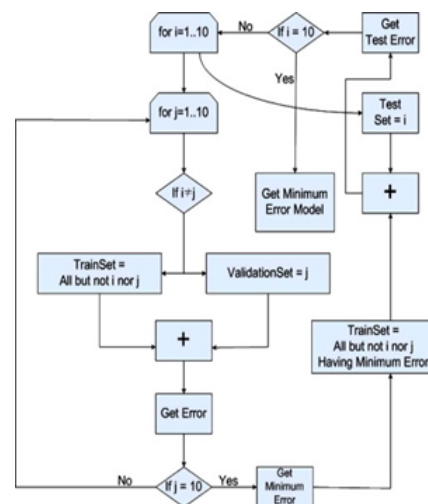


Fig. 2. Working Flowchart

B. Knowledge improvement

knowledge improvement suggests that the method of distinguishing the inaccurate, incomplete, inaccurate, orthogonal or missing a part of the information and so modifying, substitution or deleting them in keeping with the need. knowledge improvement is taken into account as a foundation part of the fundamental knowledge science.

C. Knowledge visualisation

knowledge visualization is that the graphical illustration of information —of knowledge— of knowledge and data. By exploitation visual components like charts, graphs, and maps, knowledge visualization tools give AN accessible thanks to see and perceive trends, outliers, and patterns in knowledge. Following is the graph after performing data cleaning:

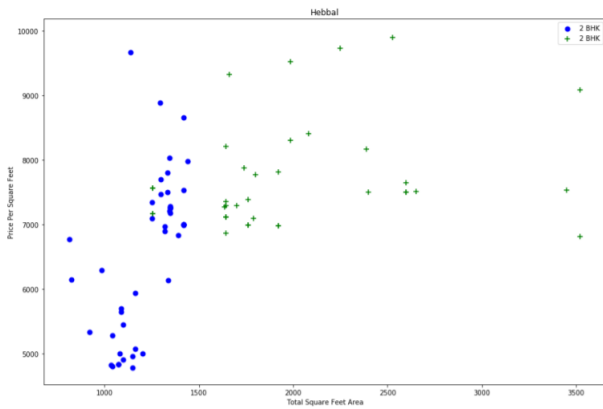


Fig. 3. Graph of Price per sqft. in Banglore of 2BHK and 3BHK flats

D. Model coaching

coaching a model merely suggests that learning (determining) sensible values for all the weights and therefore the bias from labeled examples. In supervised learning, a machine learning formula builds a model by examining several examples and trying to search out a model that minimizes loss; this method is named empirical risk step-down.

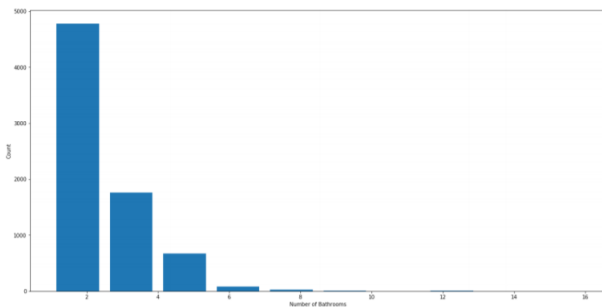


Fig. 4. Outlier Removal Using Bathrooms Feature

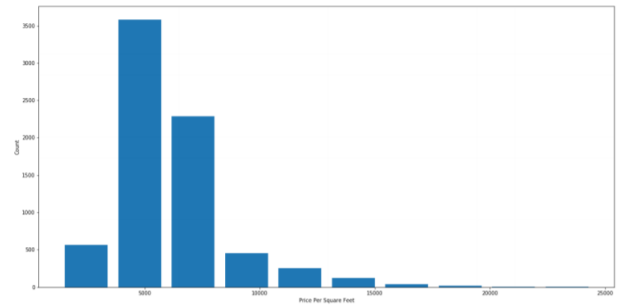


Fig. 5. Price per sqft. after Outlier Removal Using Bathrooms Feature

VII. CONCLUSION

In this analysis paper, we've used machine learning algorithms to predict the house costs. we've mentioned the step by step procedure to research the dataset and finding the correlation between the parameters. therefore square measure able to choose the parameters that aren't correlative to every alternative and are freelance in nature. These feature set were then given as Associate in Nursing input to four algorithms and a csv file was generated consisting of foreseen house costs. thence we tend to calculated the performance of every model mistreatment completely different performance metrics and compared them supported these metrics. For future work, we tend to advocate that performing on giant dataset would yield a higher and real image regarding the model. we've undertaken solely few Machine Learning algorithms that are literally classifiers however we'd like to coach several alternative classifiers and perceive their predicting behavior for continuous values too. By rising the error values this analysis work may be helpful for development of applications for numerous several cities.

VIII. ACKNOWLEDGMENT

We had to take the assistance and guidance of some esteemed person, who deserve our greatest gratitude, in carrying out our mission. Completing this mission gives us a great deal of satisfaction. We would like to express our gratitude to Dr. Saumya Chaturvedi, Project Mentor, University of Galgotia, who introduced us to the methodology of work and whose appreciation for the "underlying structures" had a lasting impact across various consultations and for providing us with strong guidance for assignment. In writing this task, we would also like to say a big thank to all those who have directly and indirectly guided us.

Many individuals, including the classmates and team members themselves, have made valuable suggestions for feedback on this proposal that motivated us to enhance our mission. We thank all the persons for their direct and indirect assistance in completing our assignment.

Finally, to our caring parents who took great care of us in these tough times. Your encouragement when the times got rough are much appreciated and duly noted. It was a great comfort and relief to know that you were willing to provide

management of our household activities while I completed my work. My heartfelt thanks! []

REFERENCES

- [1] M. Cain and C. Janssen, "Real estate price prediction under asymmetric loss," *Annals of the Institute of Statistical Mathematics*, vol. 47, no. 3, pp. 401–414, 1995.
- [2] J.-G. Liu, X.-L. Zhang, and W.-P. Wu, "Application of fuzzy neural network for real estate prediction," in *International Symposium on Neural Networks*. Springer, 2006, pp. 1187–1191.
- [3] A. S. Ravikumar, "Real estate price prediction using machine learning," Ph.D. dissertation, Dublin, National College of Ireland, 2017.
- [4] A. G. Sarip, M. B. Hafez, and M. N. Daud, "Application of fuzzy regression model for real estate price prediction," *Malaysian Journal of Computer Science*, vol. 29, no. 1, pp. 15–27, 2016.