

# Statistical data analysis on real estate using data mining

<sup>1</sup>Sanket Kurani, <sup>2</sup>Ramchandra Naik, <sup>3</sup>Vivek Bhaya, <sup>4</sup>Sharvari Patil

<sup>1,2,3</sup>Student, Dept. of Computer Engineering, Universal College, Maharashtra, India

<sup>4</sup>Assistant Professor, Dept. of Computer Engineering, Universal College, Maharashtra, India

Received: February 18, 2019

Accepted: March 28, 2019

**ABSTRACT:** : Real estate investments have become more popular last few decades. People who are investing in a new house are more conservative with their budget and market strategies. The existing system involves calculation of house prices without the necessary prediction about future market trends and price increase. The proposed system has two modes of operation the customer mode and the seller mode. The proposed real estate system gives the functionality for buyers, allowing them to search properties using features like amenities, pricing, car- pet area, address. The system enables the seller to display, update or delete the advertisements. The system will display predicted property value along with the searched property value. The future prices will be predicted by analyzing previous market trend and price ranges, and also upcoming developments future prices. The proposed system includes modules: Registration and Login, Parameter-based Filtering, Analyzing, Ranking of Advertisements and Future Value Prediction. Linear Regression is used for prediction. Since there is no involvement of brokers it reduces the risk involved during monetary transactions. The proposed system is a solution where the clients can view the future value of the real estate and invest accordingly without approaching an agent.

**Key Words:** Data Mining, Linear Regression, PricePrediction, Analysis.

## INTRODUCTION

Over the past thirty five years , a massive quantity of information has been accumulated on text mining for data Retrieval (IR). exploitation machine-driven text mining algorithms to get knowledge from language texts provides various challenges however conjointly supply distinctive potentialities . one in all the foremost natural sorts of storing data is within the sort of natural language texts .This can be simply taken by somebody's however it is still a good challenge for computers to derive which means from this data. However, computers do supply a very important advantage over human capabilities : computing power .this implies that computers will realize patterns, that area unit non-trivial recurrences, within knowledge quicker and additional correct than their human counterpart , however this may solely be done if the structure of the data is understood . language will contain implicit grammatical structure, however these structures area unit deeply complicated and vary across totally different languages. This paper brings along the most recent analysis on prediction markets to additional their utilization by economic forecasters .

Thus, there's a requirement to predict the economical house rating for real estate customers with reference to their budgets and priorities This paper with efficiency analyses previous market trends and price ranges, to predict future costs. this subject brings along the latest analysis on prediction markets to additional their utilization by economic forecasters. It provides an outline of prediction markets, and additionally the current markets that area unit helpful in understanding the market which helps in creating helpful predictions. Thus, there's a requirement to predict the economical house rating for realty customers with relation to their budgets and priorities . This project uses data mining algorithmic program to predict costs by analyzing current house costs, thereby foretelling the long run costs in step with the user's necessities.

Naïve Bayesian may be a applied math learning algorithmic program supported Bayes' rule to work out chance. It assumes conditional independence amongst the attributes .this is often used as a classification tool by initial dividing the information into freelance classes and scheming the likelihood distribution for every attribute of every category . For classification , the Naïve Bayesian finds the likelihood for the unknown in any given category and selects the category with the best likelihood.

## LITERATURE SURVEY

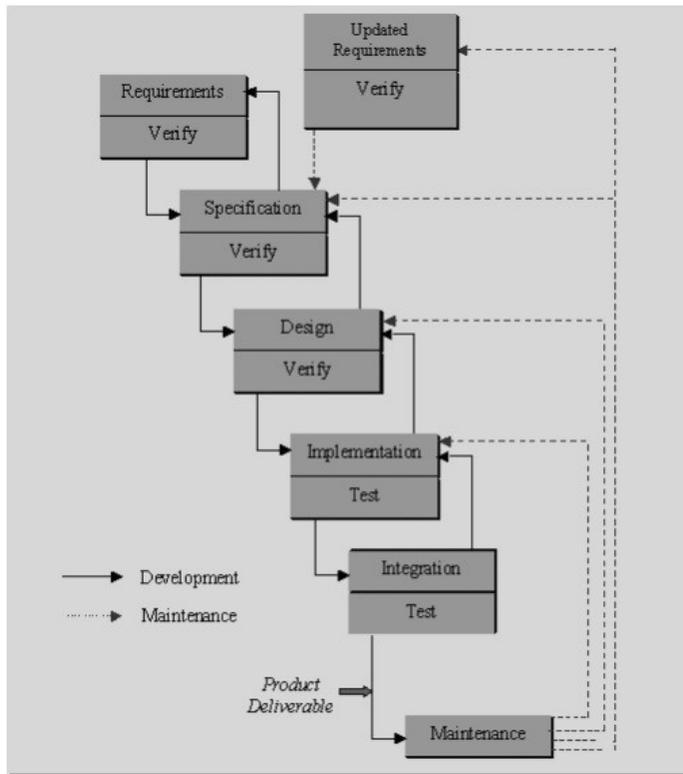
The following analysis articles area unit chosen for review, K. Y. Hindustani et al. in their work on intelligent street light-weight exploitation GSM, conjointly recommended the thought of sensible lamp posts within which it provides an outline of prediction markets, and conjointly the current markets that area

unit helpful in understanding the market which helps in creating helpful predictions. Thus, there's a desire to predict the economical house evaluation for land customers with regard to their budgets and priorities. This project uses data mining rule to predict costs by analyzing current house costs, thereby prediction the longer term costs in step with the user's necessities. Naïve theorem may be a applied math learning rule supported Bayes' rule to calculate probability. It assumes conditional independence amongst the attributes. this can be used as a classification

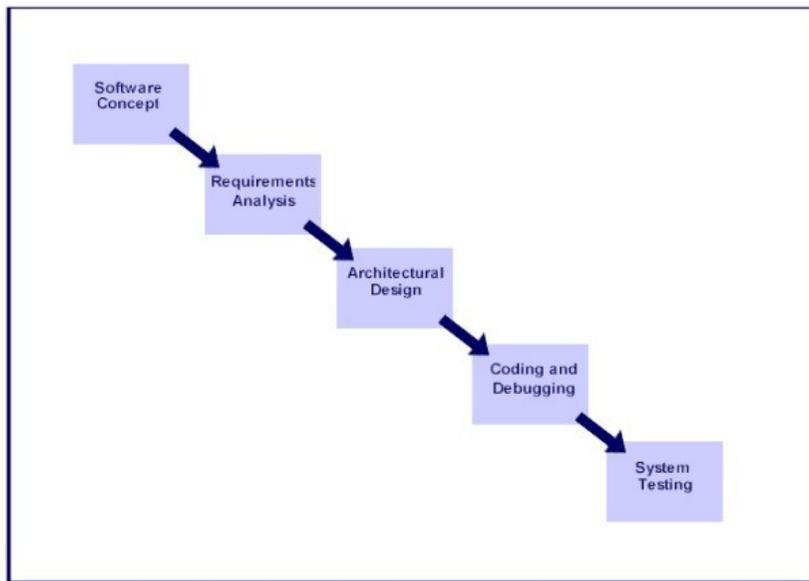
tool by 1st dividing the information into freelance classes and hard the chance distribution for every attribute of every category . For classification , the Naïve theorem finds the chance for the unknown in any given category and selects the category with the best chance. keeping in mind the normal and standard approaches of smart street lighting sys- tem: Real Estate value Prediction with Regression and Classification, CS 229 season 2016 Project Final Report In this project, house costs are fore- told given instructive variables that cowl several aspects of residential homes . As continuous house costs, they're going to be foretold with varied regression techniques as well as Lasso, Ridge, SVM regression, and Random Forest regression; as individual value ranges, they will be foretold with classification ways as well as Naive Bayes , supplying regression, SVM classification and Random Forest classification. They conjointly perform PCA to boost the prediction accuracy . The idea of project is to make a regression model and a classification model that area unit able to accurately estimate the value of the house given the options.Wang, Wen, Zhang,and Wang(2014) recommended land price prediction models supported particle swarm improvement (PSO) and support vector machine (SVM ). The experimental results indicated that the projected PSO-SVM based mostly land price prediction model has smart prediction performance compared to grid and genetic algorithms. Real Estate school Trends (2016) Properties online, Inc. has compiled vital applied math information for the \$64000 estate community. applied math sources include the 2015 National Association of REALTORS Profile of Home consumers & Sellers, the 2015 National Association of REALTORS Member Profile, the house agent technology survey report, The CA association of REALTORS vendee and Seller Sur- veys , WAV cluster agent responsiveness study, realestatesites.com and over three million web site traveller statistics from over fifteen thousand single property websites. K. Y. Rajput et al. in their work on intel- ligent street lightweight using GSM, additionally instructed the thought of goodlamp posts during which victimization machine learning algorithms for housing worth prediction. it's a well -known proven fact that housing worth valuation is one in every of most important commerce selections poignant a national property policy. In this study, they produce models victimization machine learn- ing algorithms like C4.5, liquidator (Repeated progressive Pruning to supply Error Reduction), Naïve theorem, and AdaBoost (Adaptive Boost-ing) to predict housing worth.

## PROPOSED SYSTEM

Nowadays , e-education and e-learning is extremely influenced . Everything is shifting from manual to automatic systems . the target of this project is to predict the house prices therefore on minimize the issues sweet-faced by the client. The present methodology is that the client approaches a true estate agent to manage his/her investments and recommend appropriate estates for his investments. however this methodology is risky because the agent might predict wrong estates and therefore resulting in loss of the customer's investments. The manual methodology that is presently used in the market is out dated and has high risk . So as to overcome this fault , there's a desire for Associate in Nursing updated and automated system. data processing algorithms will be wont to facilitate investors to speculate in associate in nursing acceptable estate consistent with their mentioned needs . additionally the new system are going to be value and time economical . this can have straightforward operations . The projected system works on classification algorithmic rule naïve Thomas Bayes. The administrator can add property details into the system .based on the main points the system can predict the hotels estimated worth .when user searches property the list of property will be presented the user along side the anticipated worth .the user will sell his property by adding his details onto the system, he can even seek for rent of the house via our projected system.



Schematic illustrating the Waterfall Model



The iterative Model

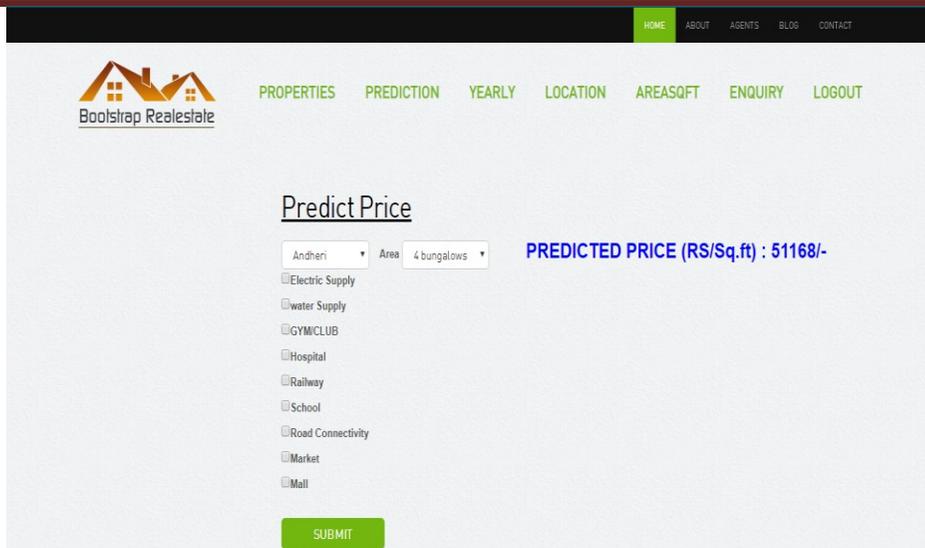


Fig.1. Prediction of properties based on filtering of amenities [1].

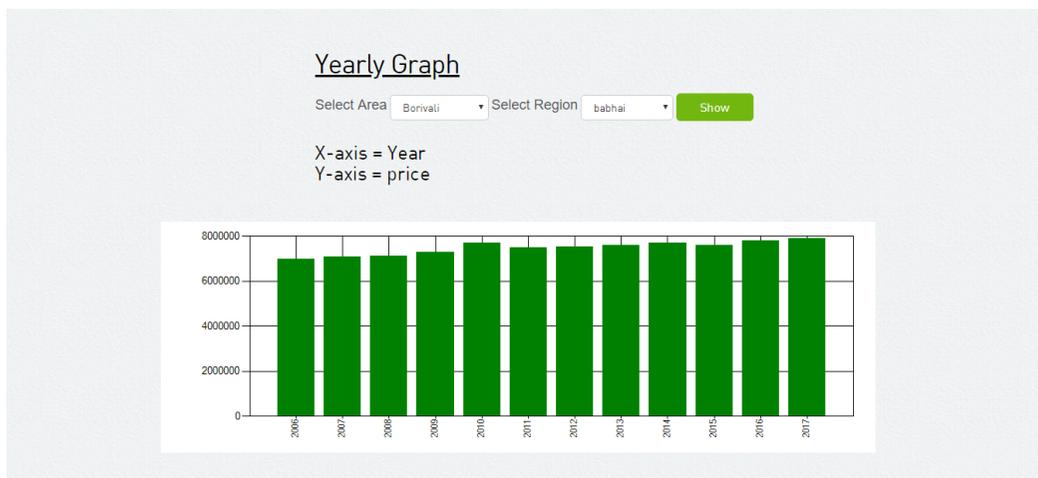


Fig 2.Graphical illustration of prices

**CONCLUSION AND FUTURE SCOPE**

In today’s realty world, it’s become robust to store such huge information and extract them for one’s own demand. Also, the extracted information ought to be helpful. The system makes best use of the Data mining formula. The system makes use of such information in the most economical manner. The information mining formula helps to fulfil customers by increasing the accuracy of estate alternative and reducing the danger of finance in associate estate. One of the foremost future scopes is adding estate info of additional cities which will offer the user to explore additional estates associated reach an accurate call. In-depth details of each property will be other to produce ample details of a desired estate. This will facilitate the system to run on a bigger level

**REFERENCES**

1. Vishal Raman,May 2014. Identifying Customer Interest inReal Estate Using Data Mining.
2. <http://www.99acres.com/property-rates-and-pricetrendsinnumbai>
3. Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, 2015. Introduction to Linear Regression Analysis
4. Gongzhu Hu, Jinping Wang, and Wenying FengMultivariate Regression Modeling for Home ValueEstimates with Evaluation using Maximum Information Coefficient
5. Wang, X., Wen, J., Zhang, Y., & Wang, Y. (2014). Real estate price forecasting based on SVM optimized by PSO. Optik-International Journal for Light and Electron Optics, 125(3), 1439–1443.
6. Real Estate Price Prediction with Regression and Classification, CS 229 Autumn 2016 Project Final Report