IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 12, December 2017

House Price Forecasting using Data Mining Techniques

Atharva chogle[1], priyanka khaire[2], Akshata gaud[3], Jinal Jain[4]

Bachelor of Engineering, Dept. of Computer Engineering, RGIT, Mumbai, Maharastra, India[1-4]

Abstract: People looking to buy a new home tend to be more conservative with their budgets and market strategies. The existing system involves calculation of house prices without the necessary prediction about future market trends and price increase. Aim of this project was to develop a real estate web application using Microsoft ASP .NET and SQL 2008.The real estate system Give the functionality for buyers, allowing them to search for houses by features or address. It provides functionality for the seller, authorize them to log into the system and add new advertisements or delete existing ones. For this each user is provided a login account with login ID and password. Along with this, when the user will search for the property, original property value and predicted property value will be displayed. By analysing previous market trends and price ranges, and also upcoming developments future prices will be predicted. For the price prediction we will be using classification algorithm .The functioning of this project involves a website which accepts customer's specifications and then uses the application of data mining. This application will help customers to invest in an estate without approaching an agent. It also decreases the risk involved in the transaction. The property, original property value and predicted property value will be displayed. By analysing previous market trends and price ranges, and also upcoming developments future prices will be predicted. For the price prediction we will be using classification algorithm. The functioning of this project involves a website which accepts customer's specifications and then uses the application of data mining. This application will help customers to invest in an estate without approaching an agent. It also decreases the risk involved in the transaction.

Keywords: House prices; real estate price; classification algorithm; price prediction; data mining; market trends

I. Introduction

Over the past 35 years, a vast amount of knowledge has been accumulated on text mining for Information Retrieval (IR). Using automated text mining algorithms to discover knowledge from natural language texts provides numerous challenges but also offer unique possibilities. One of the most natural forms of storing information is in the form of natural language texts .This can be easily interpreted by a human but it is still a great challenge for computers to derive meaning from this data. However, computers do offer an important advantage over human capabilities: computing power. This means that computers can find patterns, which are non-trivial recurrences, within data faster and more accurate than their human counterpart, but this can only be done if the structure of the data is known. Natural language does contain implicit grammatical structure, but these structures are deeply complex and vary across different languages.

This project brings together the latest research on prediction markets to further their utilization by economic forecasters. Thus, there is a need to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. This project efficiently analyses previous market trends and price ranges, to predict future prices. This topic brings together the latest research on prediction markets to further their utilization by economic forecasters. It provides a description of prediction markets, and also the current markets which are useful in understanding the market which helps in making useful predictions. Thus, there is a need to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. This project uses data mining algorithm to predict prices by analysing current house prices, thereby forecasting the future prices according to the users requirements.

II. NAÏVE BAYES ALGORITHM

Nave Bayesian is a statistical learning algorithm based on Bayes rule to compute joint probability. It assumes conditional independence amongst the attributes. This is used as a classification tool by first dividing the data into independent classes and calculating the probability distribution for each attribute of each class. For classification, the Nave Bayesian finds the probability for the unknown in any given class and selects the class with the highest probability.

The general and standardized real estate characteristics are often listed separately from the asking price and general description because these characteristics are separately listed in a structured way, they can be easily compared across the whole range of potential houses. because every house also has its own unique characteristics, such as a particular view or type of sink, house sellers can provide a summary of all the important features of the house in the description.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 12, December 2017

All given real estate features can be considered by the potential buyers, but it is nearly impossible to provide an automated comparison on all variables due to the large diversity. This is also true in the other direction: house sellers have to make an estimation of the value based on its features in comparison to the current market price of similar houses. The diversity of features makes it challenging to estimate an adequate market price. Apart from providing a summary of the important features of the house, the house description is also a means of raising curiosity in the reader, or in other words to persuade the person. It is possible that there are certain word sequences in the natural language text that attracts potential buyers more than others. Therefore, there might be a relation between the language used in the description and the price of the property. For example, a description with the word highly can outperform one with the word very looking at price fluctuation: the difference between real estate asking- and selling price. This can mean that the word highly is commonly seen in descriptions that show an increase in real estate price while the word very generally leads to a decrease in price. In addition, we can also find words that are distinctive for a certain range in selling- or asking price, thus can be used for prediction tasks. Hence, we have determined three pricing indicators that will be meaningful to predict: selling price, asking price and price fluctuation.

III.AIM AND OBJECTIVE

The aim is to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. By analyzing previous market trends and price ranges, and also upcoming developments future prices will be predicted. The functioning involves a website which accepts customers specifications and then combines the application of Naive bayes algorithm of data mining. This application will help customers to invest in an estate without approaching an agent. It also decreases the risk involved in the transaction.

The current property buying or selling is hectic and expensive. As the customer has to roam places and has to pay commission to the Real estate agent. Also, the customer/buyer does not know whether the property is profitable in future or not. Hence, we design a website using data mining techniques to overcome the drawbacks of current system as everything is web based. We are implementing following :

1) Login page
2) Location based search
3) Future estimate of property

Approx. cost of property depending on no. of attributes considered.

IV. EXISTING SYSTEM

The present system is not dunce proof and has certain drawbacks. Being a manual system the possible limitations and loopholes in the present system is large. Some of them are:-

1. HUMAN resource: - The current system has too much manual work from filling a form to filing a document, delivering manifesto. This increases burden on workers but does not yield the results it should.
2. THORNY Job: - In current system if any modification is to be made it increases manual work and is error prone.
3. ERROR: - As the system is managed and maintained by workers errors are some of the possibilities.

V. PROBLEM STATEMENT AND SCOPE

The general and standardized real estate characteristics are often listed separately from the asking price and general description. Because these characteristics are separately listed in a structured way, they can be easily compared across the whole range of potential houses. Because every house also has its own unique characteristics, such as a particular view or type of sink, house sellers can provide a summary of all the important features of the house in the description.

All given real estate features can be considered by the potential buyers, but it is nearly impossible to provide an automated comparison on all variables due to the large diversity. This is also true in the other direction: house sellers have to make an estimation of the value based on its features in comparison to the current market price of similar houses. The diversity of features makes it challenging to estimate an adequate market price. Apart from providing a summary of the important features of the house, the house description is also a means of raising curiosity in the reader, or in other words to persuade the person. It is possible that there are certain word sequences in the natural language text that seduce potential buyers more than others. Therefore, there might be a relation between the language used in the description and the price of the property. This comparison does not focus primarily on the house characteristics, but on all words within the description.

For example, a description with the word highly can outperform one with the word very looking at price fluctuation: the difference between real estate asking- and selling price. This can mean that the word highly is commonly seen in descriptions that show an increase in real estate price while the word very generally leads to a decrease in price. In addition, we can also find words that are distinctive for a certain range in selling- or asking price, thus can be used for prediction tasks. Hence, we have determined three pricing indicators that will be meaningful to predict: selling price, asking price and price fluctuation.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 12, December 2017

VI. PROPOSED SYSTEM

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 customers' 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 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. The proposed system works on classification algorithm nave Bayes. The administrator will add property details into the system
based on the details the system will predict the hotels estimated price when user searches property the list of property will be displayed to the user along with the predicted price the user can sell his property by adding his details onto the SYSTEM, he can also look for rent of the home via our proposed system.

VII. METHODOLOGY

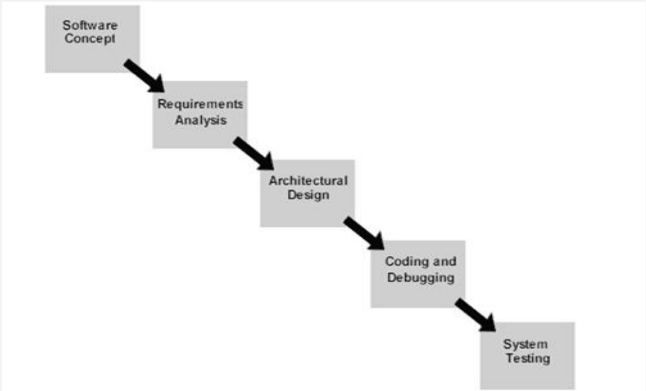

Figure.1: System development life CYCLE

The System Development Life Cycle is the process of developing information systems through investigation, analysis, design, implementation, and maintenance. The System Development Life Cycle (SDLC) is also known as Information Systems Development or Application Development.

Steps involved in the System Development Life CYCLE: Below are the steps involved in the System Development Life Cycle. Each phase within the overall cycle may be made up of several steps.

Step 1: Software Concept The first step is to identify a need for the new system. This will include determining whether a business problem or opportunity exists, conducting a feasibility study to determine if the proposed solution is cost effective, and developing a project plan. This process may involve end users who come up with an idea for improving their work. Ideally, the process occurs in tandem with a review of the organization's strategic plan to ensure that IT is being used to help the organization achieve its strategic objectives. Management may need to approve concept ideas before any money is budgeted for its development.

Step 2: Requirements Analysis Requirements analysis is the process of analysing the information needs of the end users, the organizational environment, and any system presently being used, developing the functional requirements of a system that can meet the needs of the users. Also, the requirements should be recorded in a document, email, user interface storyboard, executable prototype, or some other form. The requirements documentation should be referred to throughout the rest of the system development process to ensure the developing project aligns with user needs and requirements. Professionals must involve end users in this process to ensure that the new system will function adequately and meets their needs and expectations.

Step 3: Architectural Design
After the requirements have been determined, the necessary specifications for the hardware, software, people, and data resources, and the information products that will satisfy the functional requirements of the proposed system can be determined. The design will serve as a blueprint for the system and helps detect problems before these errors or problems are built into the final system. Professionals create the system design, but must review their work with the users to ensure the design meets users' needs.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 12, December 2017

Step 4: Coding and Debugging Coding and debugging is the act of creating the final system. This step is done by software developer.

Step 5: System Testing The system must be tested to evaluate its actual functionality in relation to expected or intended functionality. Some other issues to consider during this stage would be converting old data into the new system and training employees to use the new system. End users will be key in determining whether the developed system meets the intended requirements, and the extent to which the system is actually used.

Step 6: Maintenance Inevitably the system will need maintenance. Software will definitely undergo change once it is delivered to the customer. There are many reasons for the change. Change could happen because of some unexpected input values into the system. In addition, the changes in the system could directly affect the software operations. The software should be developed to accommodate changes that could happen during the post implementation period.

VIII. ANALYSIS

FEASIBILITY STUDY:- The very first phase in any system developing life cycle is preliminary investigation. The feasibility study is a major part of this phase. A measure of how beneficial or practical the development of any information system would be to the organization is the feasibility study.

The feasibility of the development software can be studied in terms of the following aspects:
1. Operational Feasibility
2. Technical Feasibility
3. Economic feasibility

OPERATIONAL FEASIBILITY: - The Application will reduce the time consumed to maintain manual records and is not tiresome and cumbersome to maintain the records. Hence operational feasibility is assured.
TECHNICAL FEASIBILITY: - Minimum hardware requirements: - 1.66 GHz Pentium Processor or Intel compatible processor. 1 GB RAM. Internet Connectivity. 80 MB hard disk space.
ECONOMICAL FEASIBILITY: - Once the hardware and software requirements get fulfilled, there is no need for the user of our system to spend for any additional overhead. For the user, the Application will be economically feasible in the following aspects: The Application will reduce a lot of labour work. Hence the Efforts will be reduced. Our Application will reduce the time that is wasted in manual processes. The storage and handling problems of the registers will be solved.

IX. SOFTWARE AND HARDWARE REQUIREMENT

HARDWARE REQUIREMENTS:
1 GB RAM.
200 GB HDD.
Intel 1.66 GHz Processor Pentium 4

SOFTWARE REQUIREMENTS:
Windows XP, Windows 7, 8,10
Visual Studio 2010
Microsoft SQL Server Windows Operating System

X. DESIGN DETAILS

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 customers' 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 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. The proposed system works on classification algorithm nave Bayes.
The administrator will add property details into the system based on the details the system will predict the hotels estimated price when user searches property the list of property will be displayed to the user along with the predicted price the user can sell his property by adding his details onto the SYSTEM, he can also look for rent of the home via our proposed system.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 12, December 2017

Activity diagram is another important diagram in UML to describe the dynamic aspects of the system. Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. The control flow is drawn from one operation to another. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all type of flow control by using different elements such as fork, join, etc.

Admin activity: - Admin can log in by providing his user name and password if it is incorrect, screen will show invalid message if it is correct log in successful message will be shown Admin can add his properties and start predictions over it.

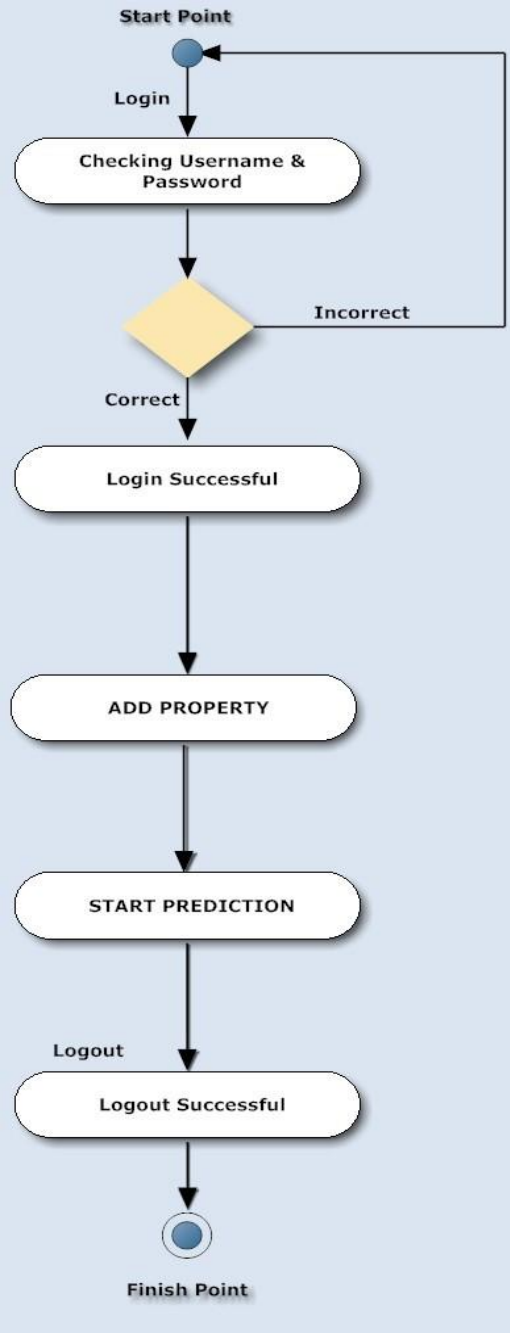

Figure 2: admin activity

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 12, December 2017

User activity:- User can Search property, add property for selling and can also look for rent by search method after logging in. Results are fetched and added from the main database.

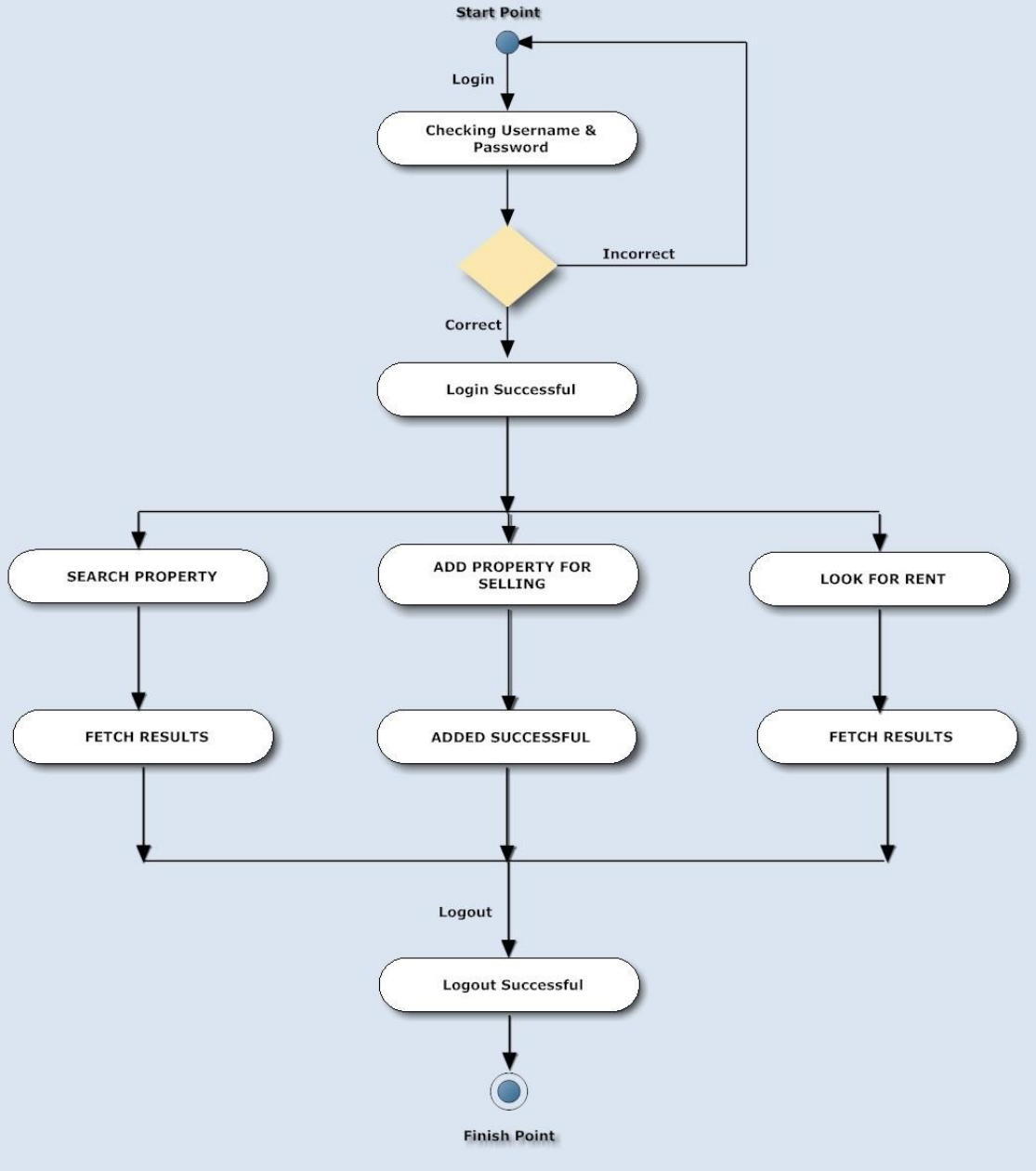

Figure 3: User activity

DOI 10.17148/IJARCCE.2017.61216

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 12, December 2017

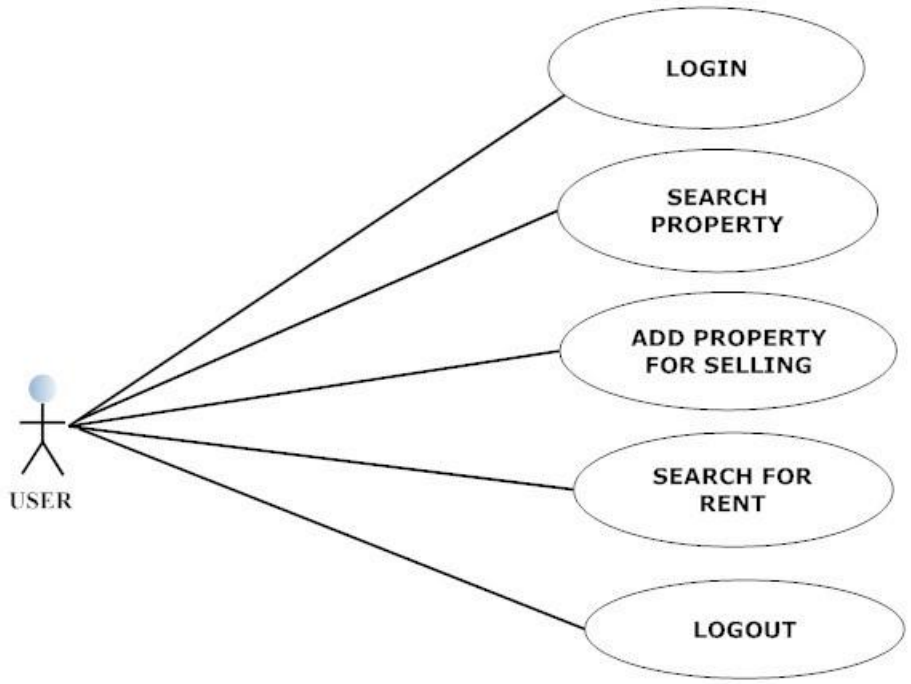

Figure 4: Use case diagram (user)

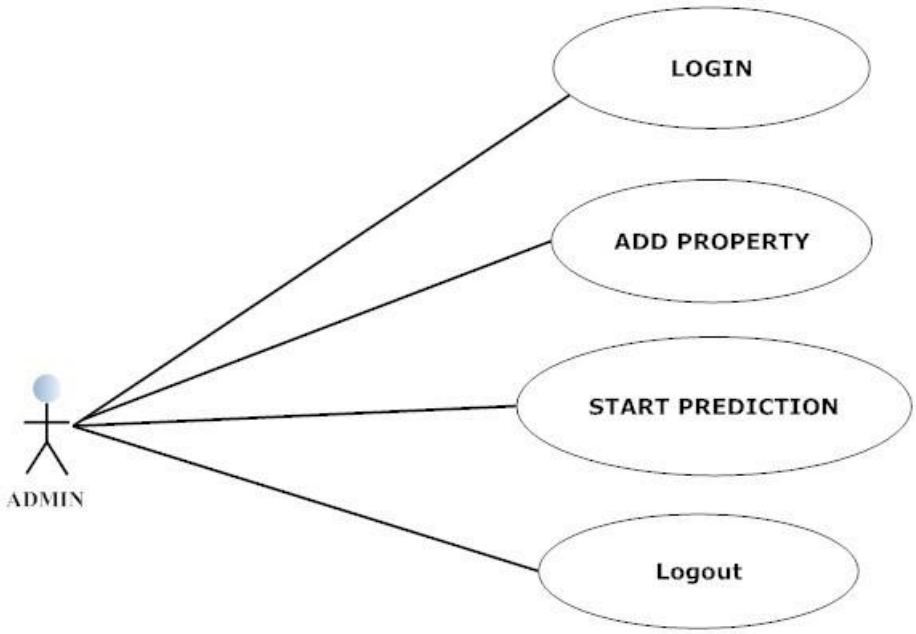

Figure 5: Use case diagram (admin)

Sequence:- Admin can log in in his account anytime by putting his credentials. If credentials match in database's data, system will grant log in. Vice versa is true for user as well. Admin and user both use some kind of web application to connect with database. Admin and user can both search for property and generate prediction codes from database. User can manage his purchase and selling information of system. For logging out, they both have to request a log out request to system. System grants the log out session and they can log out.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 12, December 2017

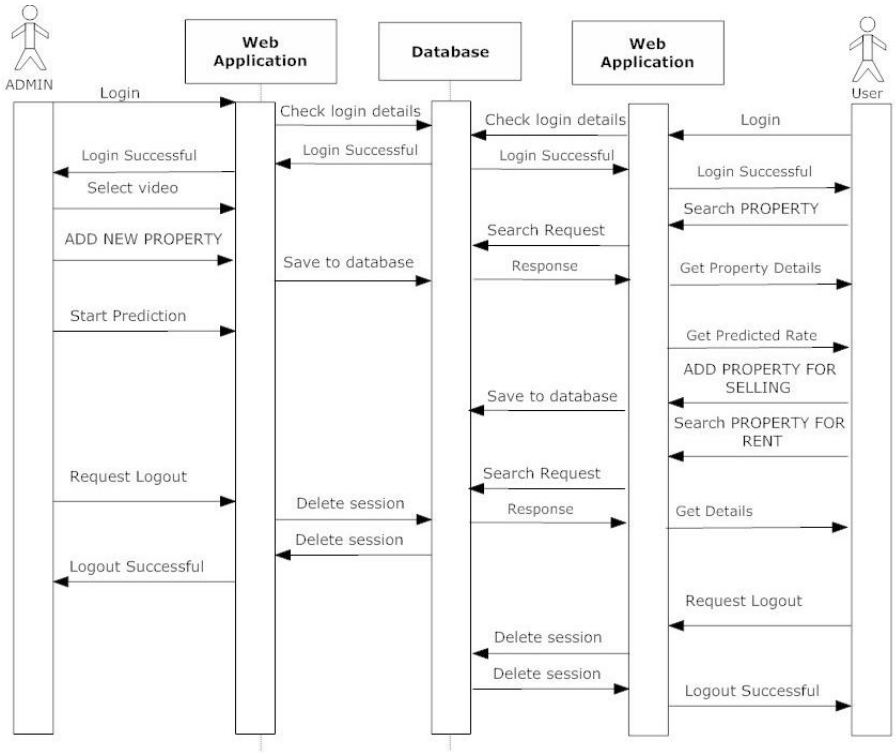

Figure 6: Sequence of action

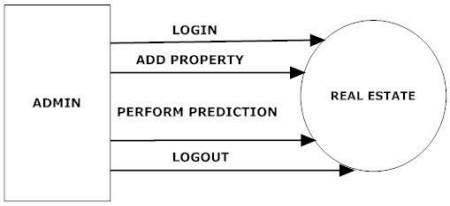

Figure 7: DFD level 0(admin)

DFD admin level1:- Admin have to first log in into the website. When he enters new property then that info will be stored into the database so that it will be easy to retrieve. when the predictions are made, that data also stored into the database.

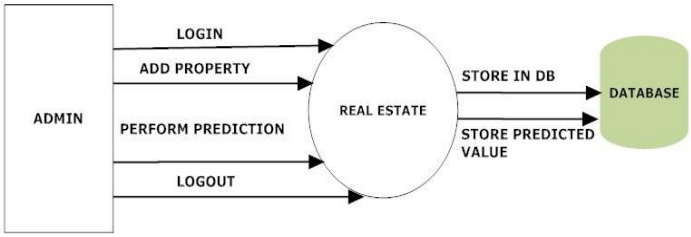

Figure 8: DFD level 1(admin)

Dfd user level0:- After user is successfully logged in, he can search properties according to his preferences.He can add his/her shortlisted properties. He can also search for respective rent for any provided property.

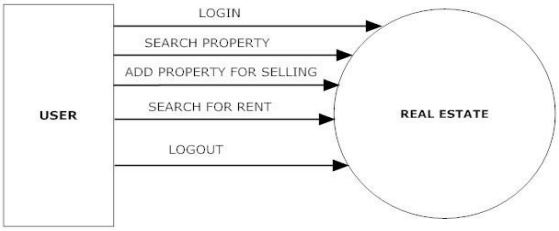

Figure 9: DFD level 0(user)

DOI 10.17148/IJARCCE.2017.61216

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 12, December 2017

DFD user level1:- User can log in in his account. When he search some property, system retrieves data from database and results are shown to user. When he adds some property it is stored by system in database. He can also search for rent of Houses. System will retrieve the requested data from the database and results will be shown to user.

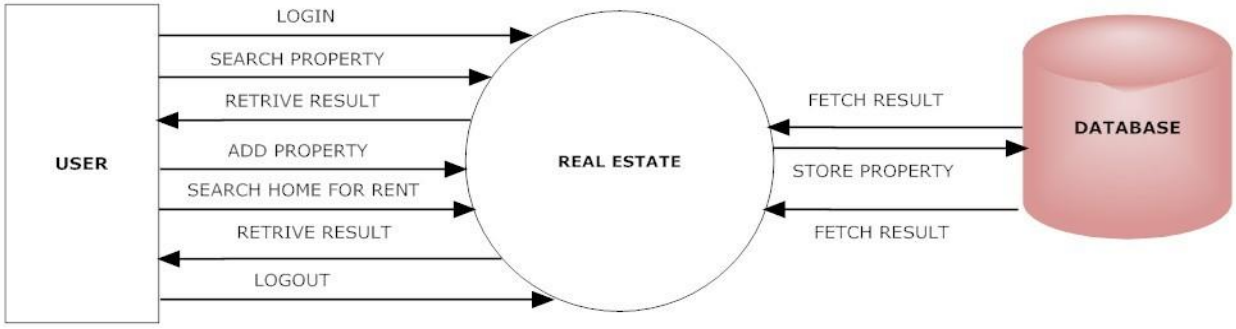

Figure 10: DFD level 1(user)

XI. IMPLEMENTATION PLAN

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 customers' 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 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. The proposed system works on classification algorithm nave Bayes. The administrator will add property details into the system based on the details the system will predict the hotels estimated price when user searches property the list of property will be displayed to the user along with the predicted price the user can sell his property by adding his details onto the SYSTEM, he can also look for rent of the home via our proposed system.

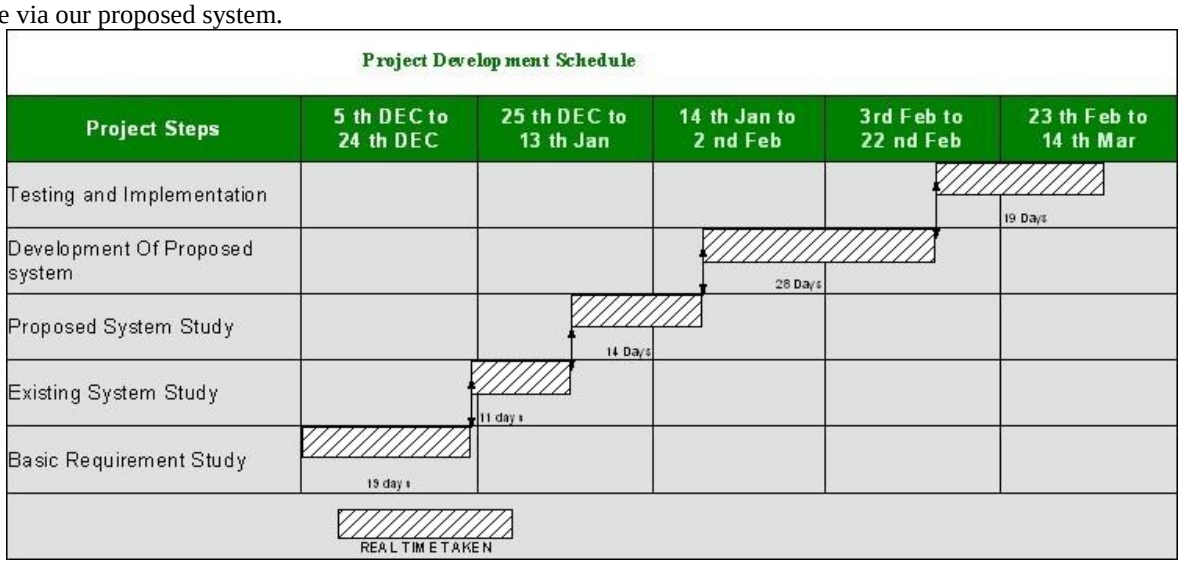

Figure 11: Gant chart

The Gantt chart: - It shows planned and actual progress for a number of tasks displayed against a horizontal time scale. It is effective and easy-to-read method of indicating the actual current status for each of set of tasks compared to planned progress for each activity of the set.

XII. CONCLUSION

In todays real estate world, it has become tough to store such huge data and extract them for ones own requirement. Also, the extracted data should be useful. The system makes optimal use of the Data mining Algorithm. The system makes use of such data in the most efficient way. The Data mining algorithm helps to fulfill customers by increasing the accuracy of estate choice and reducing the risk of investing in an estate. A lots of features that could be added to make the system more widely acceptable. 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. More factors like recession that affect the

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 12, December 2017

house prices shall be added. In-depth details of every property will be added to provide ample details of a desired estate. This will help the system to run on a larger level. There are quite a few things that can be polished or add in the future work. • Though, we were able to identify most of the residential areas. There may be some more places that have housing complexes or multi-storey apartments which are located in commercial areas. Such apartments were not included in this paper and can be counted in future to give a more accurate result. With more and more demand for housing in metropolitan cities, there is a definite increase in the number private builders that provide real estate with additional amenities to attract more customers. • There are several other models available that can be implemented for prediction. Data given as input to such model should be compatible with the tool used and the operators involved in the process. Also, more number of data sets can be used to increase the accuracy of the model. The main objective of using a different model should be to reduce the calculation time and carry out the whole process in ease.

XIII. FUTURE SCOPE

There are quite a few things that can be polished or add in the future work. • Though, we were able to identify most of the residential areas. There may be some more places that have housing complexes or multi-storey apartments which are located in commercial areas. Such apartments were not included in this paper and can be counted in future to give a more accurate result. With more and more demand for housing in metropolitan cities, there is a definite increase in the number private builders that provide real estate with additional amenities to attract more customers. • There are several other models available that can be implemented for prediction. Data given as input to such model should be compatible with the tool used and the operators involved in the process. Also, more number of data sets can be used to increase the accuracy of the model. The main objective of using a different model should be to reduce the calculation time and carry out the whole process in ease.

XIV. ACKNOWLEDGEMENT

This research was supported by our project guide **Mrs. Preeti Satao** (Assistant professor), We thank her for providing us an opportunity to do the project research on real estate financial Management using Data mining at Rajiv Gandhi Institute of Technology (R.G.I.T.), Andheri, Mumbai and by giving us the support and guidance.

REFERENCES

[1] Real Estate Price Prediction with Regression and Classification, CS 229 Autumn 2016 Project Final Report
[2] Gongzhu Hu, Jinping Wang, and Wenying Feng Multivariate Regression Modelling for Home Value Estimates with Evaluation using Maximum Information Coefficient
[3] Byeonghwa Park , Jae Kwon Bae (2015). Using machine learning algorithms for housing price prediction , Volume 42, Pages 2928-2934
[4] Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, 2015. Introduction to Linear Regression Analysis
[5] Iain Pardoe, 2008, Modelling Home Prices Using Realtor Data
[6] Aaron Ng, 2015, Machine Learning for a London Housing Price Prediction Mobile Application
[7] Wang, X., Wen, J., Zhang, Y.Wang, Y. (2014). Real estate price forecasting based on SVM optimized by *P*SO. Optik-International Journal for Light and Electron Optics, 125(3), 14391443.
[8] Vishal Raman, May 2014. Identifying Customer Interest in Real Estate Using Data Mining.
[9] http://www.99acres.com/property-rates-and-pricetrendsin-mumbai
[10] Real Estate Price Prediction with Regression and Classification, CS 229 Autumn 2016 Project Final Report In this project, house prices will be predicted given explanatory variables that cover many aspects of residential houses. As continuous house prices, they will be predicted with various regression techniques including Lasso, Ridge, SVM regression, and Random Forest regression; as individual price ranges, they will be predicted with classification methods including Naive Bayes, logistic regression, SVM classification, and Random Forest classification. They also perform PCA to improve the prediction accuracy. The goal of this project is to create a regression model and a classification model that are able to accurately estimate the price of the house given the features.
[11] Suggested real estate price forecasting models based on particle swarm optimization (PSO) and support vector machine (SVM). The experimental results indicated that the proposed PSOSVM based real estate price forecasting model has good forecasting performance compared to grid and genetic algorithms.
[12] Real Estate Tech Trends (2016) Properties Online, Inc. has compiled important statistical information for the real estate community. Statistical sources include the 2015 National Association of REALTORS Profile of Home Buyers Sellers, the 2015 National Association of REALTORS Member Profile, The Realtor Technology Survey Report, The California Association of REALTORS Buyer and Seller Surveys, WAV Group Agent Responsiveness Study, RealEstateSites.com and over 3 million website visitor statistics from over 15 thousand single property websites.
[13] Using machine learning algorithms for housing price prediction, Byeonghwa Park , Jae Kwon Bae, 2015 It is a well-known fact that housing price valuation is one of most important trading decisions affecting a national real estate policy. In this study, they create models using machine learning algorithms such as C4.5, RIPPER (Repeated Incremental Pruning to Produce Error Reduction), Nave Bayesian, and AdaBoost (Adaptive Boosting) to predict housing price.