International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-2 Issue-2, May 2013

Finding Interest of People in Purchasing Real Estate by Using Data Mining Techniques

Swati Singh, Gaurav Dubey

Abstract: ***Data mining is the extraction of hidden predictive information from large databases; it is a powerful technology with great potential to help organizations focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviours, helps organizations to make proactive knowledge-driven decisions. Hence by using data mining techniques we predict the interest of people in real estate and their pattern of purchasing them. The data has collected by moving the questionnaire among the people. We used two data mining techniques that classify the data based on certain attributes, are classification (Zeror classifier) and clustering (simple k means). And then based on their result several bar charts have been drawn.***

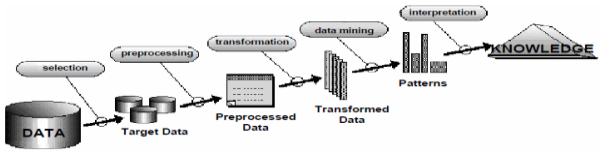
Keywords- Proactive Knowledge-Driven Decisions.

I. INTRODUCTION

To generate information it requires massive collection of data. The data can be simple numerical figures and text documents, to more complex information such as spatial data, multimedia data, and hypertext documents. To take complete advantage of data; the data retrieval is simply not enough, it requires a tool for automatic summarization of data, extraction of the essence of information stored, and the discovery of patterns in raw data. Data mining is the extraction of hidden predictive information from large databases; it is a powerful technology with great potential to help organizations focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, helps organizations to make proactive knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems.

II. KNOWLEDGE DISCOVERY AND DATA MINING

Data mining is also called knowledge discovery and data mining (KDD). It is the extraction of useful patterns from data sources, e.g .databases, texts, web, images, etc. Patterns must be valid, novel, potentially useful, understandable.



Most data mining goals fall under the following categories:

Revised Manuscript Received on 30 May 2013.
* Correspondence Author
Swati Singh, Computer science, Amity University, Noida, Uttar Pradesh, India.
Gaurav Dubey, Computer science, Amity University, Noida, Uttar Pradesh, India.

- *Pre-processing*

Preprocessing can provide accurate, concise data for data mining. Data preprocessing, includes data cleaning, user identification, user sessions identification, access path supplement and transaction identification.

- *Pattern discovery*

Pattern discovery mines effective, novel, potentially useful and ultimately understandable information and knowledge using mining algorithm. Its methods include statistical analysis, classification analysis, association rule discovery, sequential pattern discovery, clustering analysis, and dependency modeling.

- *Pattern Analysis*

Pattern Analysis is a final stage of the whole web usage mining. The goal of this process is to eliminate the irrelevant rules or patterns and to understand, visualize and to extract the interesting rules or patterns from the output of the pattern discovery process.

III. VARIOUS DATA MINING TECHNIQUES

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases.

- *Classification*

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit risk applications are particularly well suited to this type of analysis. This approach frequently employs decision tree or neural network-based classification algorithms.

Types of classification models:

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

- *Clustering*

Clustering can be said as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes.

Types of clustering methods

- Partitioning Methods
- Hierarchical Agglomerative (divisive) methods
- Density based methods
- Grid-based methods
- Model-based methods

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

- *Prediction*

Regression technique can be adapted for prediction. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict.

Types of regression methods

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

- *Association rule*

Association and correlation is usually to find frequent item set findings among large data sets.

Association Rule algorithms need to be able to generate rules with confidence values less than one.

Types of association rule

- Multilevel association rule
- Multidimensional association rule
- Quantitative association rule

- *Neural networks*

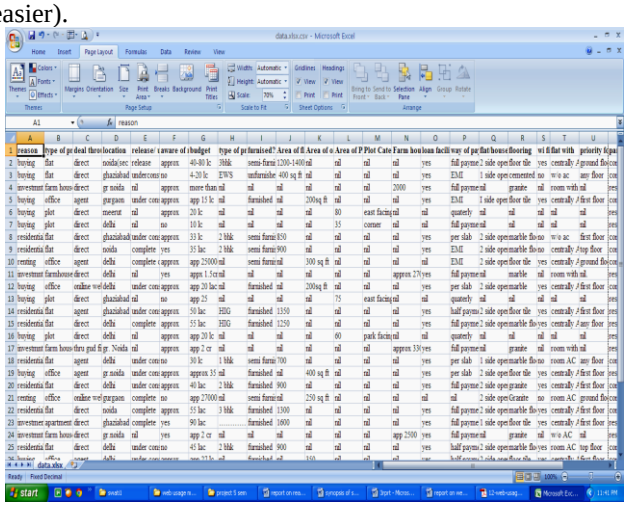
Neural network is a set of connected input/output units and each connection has a weight present with it. During the learning phase, network learns by adjusting weights so as to be able to predict the correct classlabels of the input tuples. Neural networks are best at identifying patterns or trends in data and well suited for prediction or forecasting needs.

Types of neural networks

- Back Propagation

IV. EXPERIMENT & RESULT

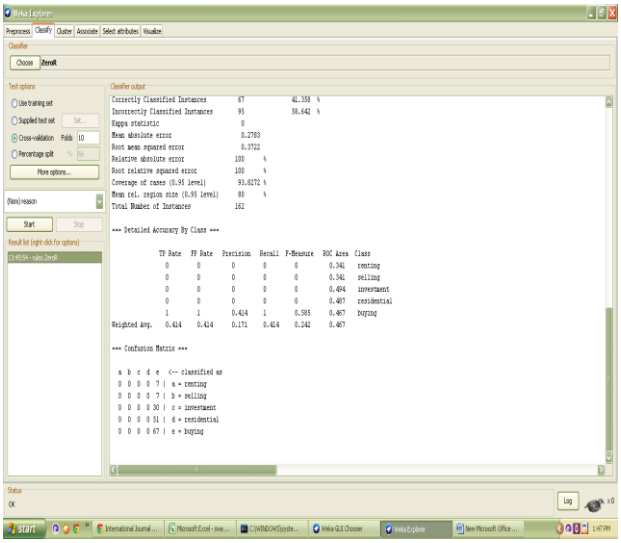
The very first task is to prepare a questionnaire, that contains about 25 multiple choice questions based *on real estate*, Now a survey will be done based on this questionnaire, as this step indicate data collection. Here the sample of questions distributed among 300 people and the information is gathered. This questionnaire contains questions on real estate that generally asked by a common man while he/she is interested in purchasing a property. The queries are basically related to for eg. Reason behind purchasing the property, their interest in type of property, their budget , the area they need for their property , their desired location, and many more. The collected data is saved in .xls file format, as shown in Fig (any other extension as .txt, .dat, will also work but finally we need a .csv file format and to change .xls in .csv is comparatively easier).



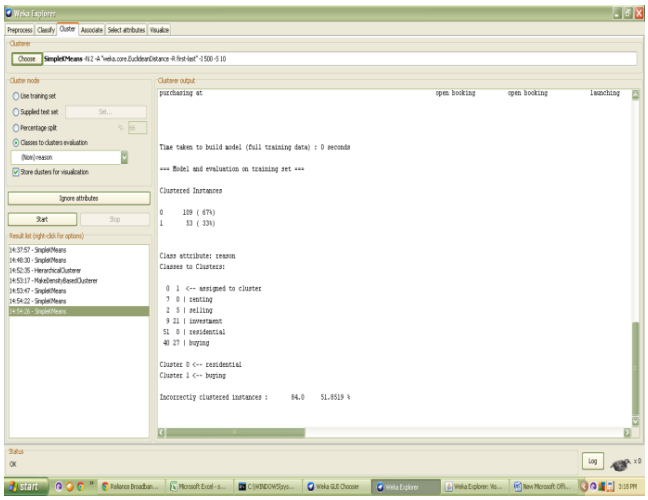
WEKA (version 3.7.5) is a full data mining suite which includes various preprocessing modules and data mining techniques. Therefore, for this research we are using WEKA package. The very first step is to import the data into WEKA. The data must be in .csv format or in .ARFF format (If it is in .csv format later on it can be save in .ARFF format).

As soon as the tool receive the input file, it will preprocess and filter the raw data.

Next task is to be achieve is KDD i.e Knowledge Data Discovery. Data mining is core of KDD or can say they are treated as synonyms. It is refer to the non trivial extraction of implicit, previously unknown, and potentially useful information from data in data bases.KDD is an iterative process that contains data mining as a step. Once the discovered knowledge is presented to the user , the evaluation measures can be enhanced , the mining can be further refined , new data can be selected or further transformed , or new data sources can be integrated in order to get different , more appropriate results. The kind of patterns that can be discovered depends upon the data mining tasks employed. There are two types of mining tasks descriptive data mining and predictive data mining.

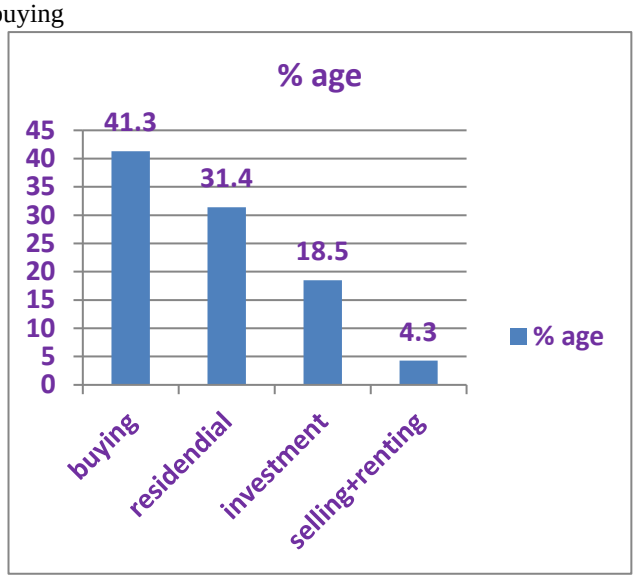


Classification of data using ZeroR classifier using class “reason”

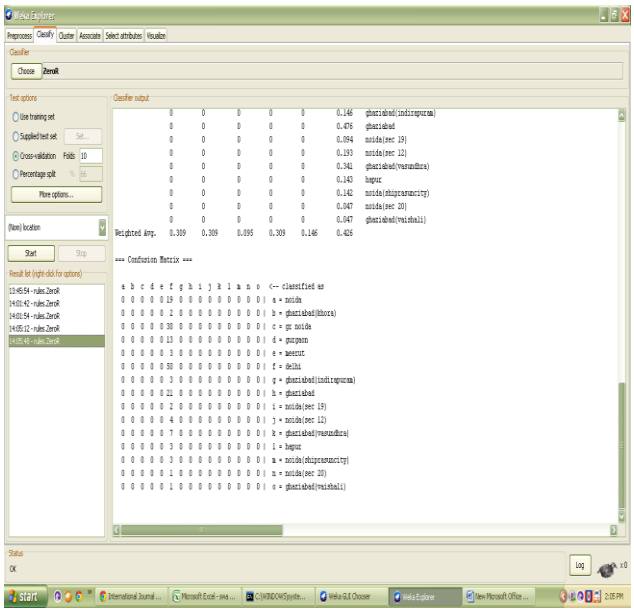


Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

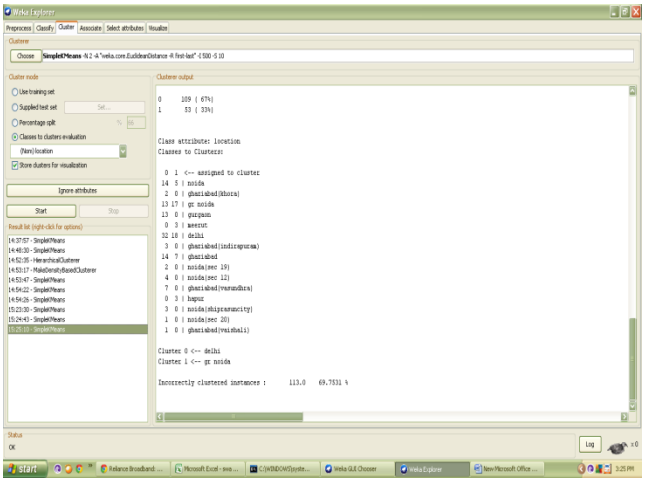
Clustering using Simple k means by using class "reason": that forms 2 clusters i.e, cluster 0:residential, cluster 1: buying



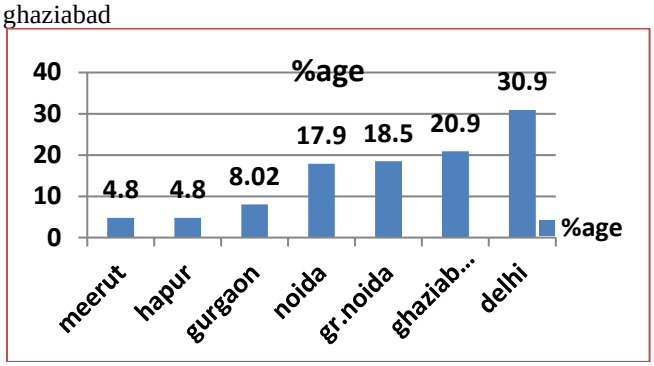
Bar chart based on class reason.



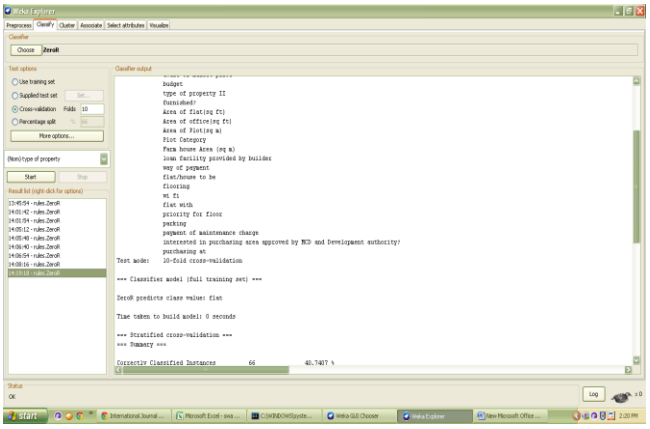
Classification of data using ZeroR classifier using class "locations"



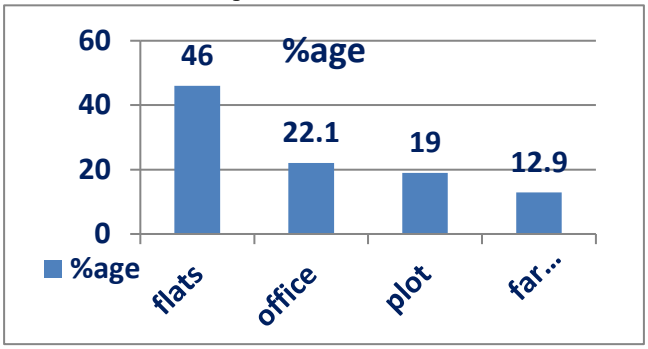
Clustering using Simple k means by using class "location": that forms 2 clusters i.e, cluster 0: delhi, cluster 1: ghaziabad



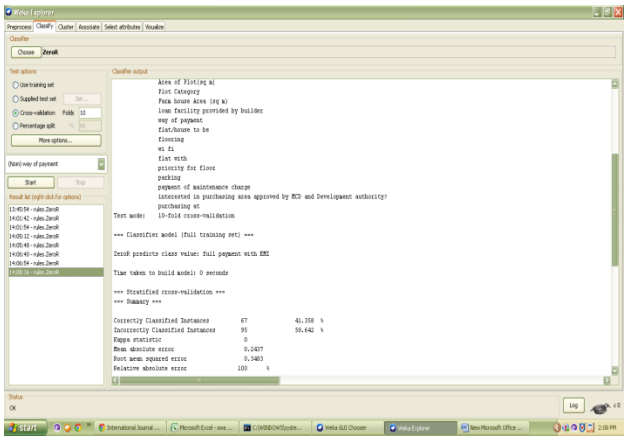
Bar chart based on class location.



ZeroR predicts class value "flat"

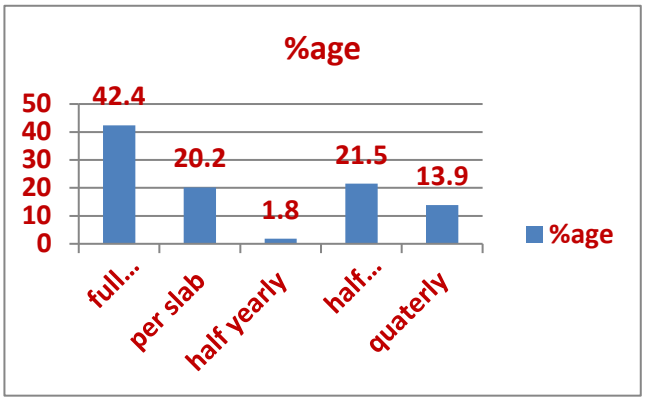


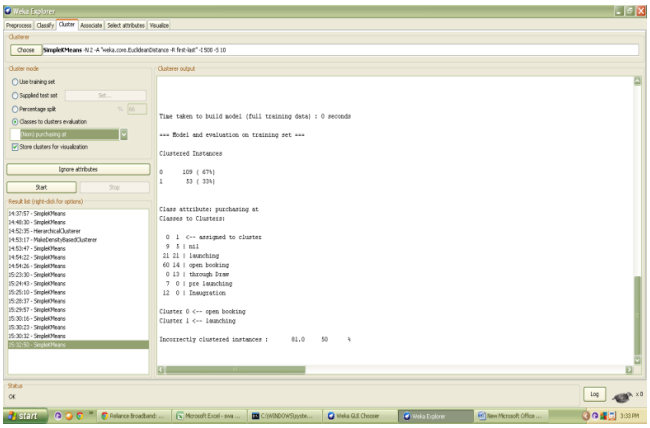
Bar chart based on type of property II



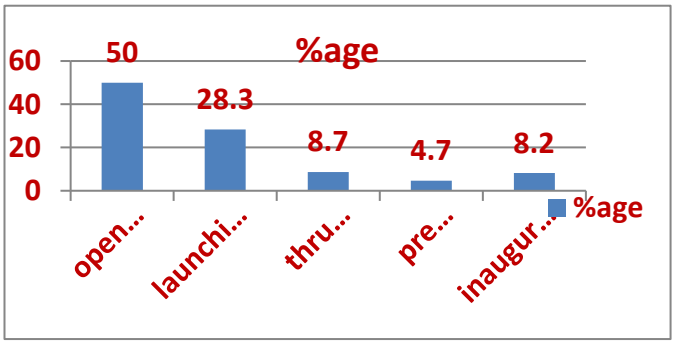
Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Zeror predicts class value “full payment with EMI”, when classification is done by using class “way of payment”





Clustering using Simple k means by using class “purchasing at ”: that forms 2 clusters i.e, cluster 0: open booking , cluster 1: launching



Bar chart based on class purchasing at.

V. CONCLUSION AND FUTURE WORK

We develop a model by using various existing data mining techniques that defines the pattern or interest of people during purchasing real estate.

WEKA (version 3.7.5) is a full data mining suite which includes various preprocessing modules and data mining techniques. Therefore, for this research I am using WEKA package.

I applied two data mining techniques i.e, classification (Zeror algorithm) and clustering (simple k means) on the collected data. Afterwards some bar charts has been drawn based on their result.

After classification & clustering, I got the conclusion that 55.3% of people have their budget between the range of 20 – 40 lacs for purchasing the property .And most of the people (41.3%) are interested in buying flats (apartments) of 2bhk in Delhi (30.9%) and then Ghaziabad (20.9%) & in Noida (17.9%). Approx 62.3% people want their deal directly to the second party & 29.3 % people shows their faith in agents (dealers). Those who are interested in acquiring residential property, 52.2% of people gives the preferences to first floor and 21.1 % have priority for ground floor. Generally Builder pre decides the schemes of payment for the purchaser and then the final choice is of them. By analyzing the data i got that about 50% of people are interested in purchasing at the booking point, and from them 42.4% opt the way of “full payment with EMI” alternative. Data says that, 18.5% of people purchase the property for investment purpose. Approx 67.7% spend their money in buying farmhouse and their choices for the location are Gr. Noida and Delhi and those who are spending their capital in apartment are 32.3 % and they show their interest in Ghaziabad and Delhi. According to the analysis, 22.1% of people have their interest in purchasing office and they want their deal must be done by agent (Dealer). The choices for the location are Delhi , Gurgaon , Gr. Noida but they give the preference to Gurgaon. As we collected the data through a questionnaire, and classified & clustered according to reason for purchasing property, desired locations, budget and in the type of property they are interested. But, there are certain issues exist in this model too. Therefore in the future, we plan to extend this work along the following dimensions:

- We did not consider all the questions that might cover entire coverage area of real estate market. So, finding such questions need further work.
- We will examine or set a knowledge query mechanism on the sequential patterns, page clusters or association rules that we got during KDD process.

REFERENCES

1. Jiawei Han and Micheline Kamber (2006), Data Mining Concepts and Techniques, published by Morgan Kauffman,2nd ed.
2. Agrawal R, Imilienski T, Swami A (1993). Mining association rules between sets of items in large databases, In Proceedings of the ACM SIGMOD international conference on management of data.
3. Basaltoa N, Bellottib R, De Carlob F, Facchib P, Pascazio S (2005). Clustering stock market companies via chaotic map synchronization, Physica A.
4. Berry MJA, Linoff GS (2000). Mastering data mining, New York: Wiley.
5. Boris K, Evgenii V (2005). Data Mining for Financial Applications, the Data Mining and Knowledge Discovery Handbook.
6. Data mining: Ford, C.W.; Chia-Chu Chiang; Hao Wu; Chilka, R.R.; Talburt,J.R.; Information Technology: Coding and Computing, 2005. ITCC 2005 International Conference Volume: Digital Object Identifier: 10.1109/ITCC.2005.270 Publication Year: 2005 , Page(s): 122 - 127 Vol. 1

AUTHOR PROFILE

[1]SWATI SINGH, pursuing M.tech in computer science (batch 2010-2013) from Amity university, Noida ,UP. India. B.Tech in Information Technology from UPTU . My research interest is in DATA MINING . A paper that is named as Mobile Adhoc network , had published in national conference on “Emerging Trends in IT & Computing Technology”. Another paper , named as “function of feature selection in data filtering” got published in international conference on “Emerging trends & Development in Science, Management & technology”.
2 year experience in Teaching.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)