RESEARCH ARTICLE

APPLICATION OF DATA MINING TECHNIQUES TO PREDICT CUSTOMER-LOANS PREFERENCES IN REAL ESTATE

POOJA PANT

Department of Computer Science and Engineering
Amity University, Noida, Uttar Pradesh, India

ABSTRACT

In today's highly informative world data mining plays crucial role in the success of organizations in a domain. Every day large volume of data is created and stored in data marts and data warehouses, therefore developing powerful data mining tools for analysis of such data and mining useful knowledge from it. This paper describes a specific case of real estate world dataset to demonstrate how each information step can be completed interactively to get useful and required knowledge in less time/efforts from its data warehouses with human efforts. In this paper, a survey of the research in the domain of real estate is presented with its findings. Applying various data mining techniques in real estate domain can be very useful in extracting.

INTRODUCTION

Data mining is the process of discovering meaningful information from data stored in various in sources or data in a data warehouse. In today's highly technological and informative world analysing data can be a game changer; it provides valuable information for decision-making. Currently, many data mining tools are available such as WEKA, Rapid Miner, Tanagra etc. These tools provide us with various techniques which enables better utilization of data these techniques includes clustering, classification, association rules, statistics etc. These techniques provide result oriented analysis of data.

In this paper we are considered various attributes of a customer who took a loan such as: Occupation, annual income and property price. This paper aims at analysing preferences of various occupation segments in the society for taking loans by using various data mining techniques from real estate data set.

Procedure

The data was collected from 302 customers in real estate projects which are going on in Delhi NCR region. This region has become a rapid developing ground for residential and commercial projects over the past decade. Data mining technologies allows organizations to analyse large quantity of data, compiles it and discover relationship and correlation in-between the data to improve decision making and help in future predictions of outcomes.

Data Mining can help in answering many vital questions related to the present scenario in different application domains. In our research we tried to find useful relations and information from real estate dataset. TANAGRA is a free, open-source and user-friendly data mining software and has been used in our experiment. Data visualization, statistics and supervised learning are the various components used in our experiment. From data visualization component we will use view dataset to view the dataset in use and scatter-

plot to visualize the relationship between 2 attributes. From statics we use uninvarient discrete stat and group characterization to characterize group by loan status and supervised learning. The dataset was created by survey; there are 25 attributes and 302 instances in the dataset. The missing values have been removed.

TABLE 1: Description of attributes of Real-estate Dataset

Attribute name	Description	Type
name	Name of participant	Discrete
Email_id	Email id of participant	Discrete
Timeconatct	Contact timing (date)of participant	Discrete
Age	Age of participant	Discrete
Gender	Gender of participant	Discrete
Address	Current Address of participant	Discrete
City	Current city of participant	Discrete
State	Current state of participant	Discrete
Occupation	Occupation of participant	Discrete
Occ_address	Occupation address of participant	Discrete
Pro_buying	Buying a property	Discrete
Pro_selling	Selling a property	Discrete
Pro_loaction	Property location	Discrete
Pro_persqft	Property size	Discrete
Pro_loan	Loan taken	Discrete
Pro_builder	Property builder	Discrete
Pro_status	Status of property	Discrete
mobile	Mobile number	Continuous
Pin code	Pin code of current address	Continuous
Family size	Number of family members	Continuous
Pro_city	Property city	Continuous
Pro_price	Property price	Continuous
Pro_size	Property size	Continuous
Pro_bedrooms	Number of bedrooms	Continuous
Annual_income	Annual income of participant	Continuous

The Table 2 shows the different components of TANAGRA which will be used for the analysis of the dataset. We use Tanagra for analyzing various basic statistical attributes. Later we find the impact of occupation attribute on the probability of loan taken. We finally provide description of customer queries preferences in respect of area, size, cost, location of the flat information.

TABLE 2: Components of TANAGRA Used in the Experiments

	Tab	Components
1	Feature selection	Define status
2	Data Visualisation	Scatter Plot
3	Statistics	Univarent Discrete stat
4	Statistics	Group Characterierization
5	Statistics	Univarent Continuous stat

Results and analysis

Data mining can help answer crucial question in the present application domain such as 'which segment in society take maximum loans ' or 'which property size is desired by people with different family income' The role of data mining is not to practice the outcomes but to fetch useful information and knowledge from the real-estate dataset

The number of people taking loan and the segment of people who have taken this survey can see in figure 1 and figure 2.around 73 % of the people take loans for buying a property. most of the people who participated in the survey were professionals like teachers, professors, doctors etc. followed by government employee and businessman.
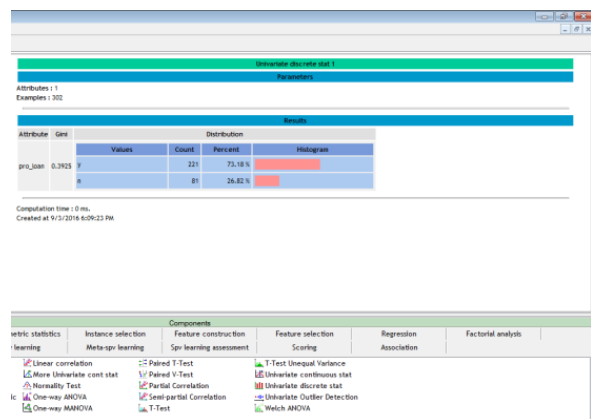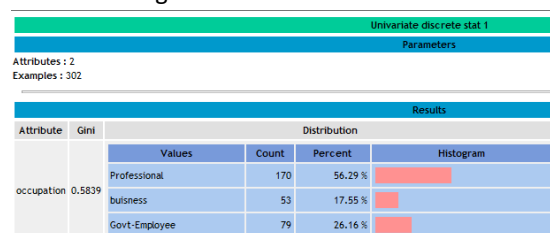

Figure 1: Loan Distribution

Figure 1: Occupation and loan distribution

Most of the loans taken for property ranging between 20-50 lakhs, loans are usually taken up for property with price less than 55lakhs irrespective of the annual income. Fig 3. Shows that most of the government employee and professional have an annual income less than Rs.12lac tend to buy property less than 40lac but still most of the business persons have annual income more than Rs.12lac buy property less than Rs.50lac .similarly in fig 4. We observe that the loans are not commonly taken by individuals for property less

than Rs. 20 lakhs irrespective of the annual income. The numbers of people opting for loans are more than the number of people not taking loans irrespective of their occupation.
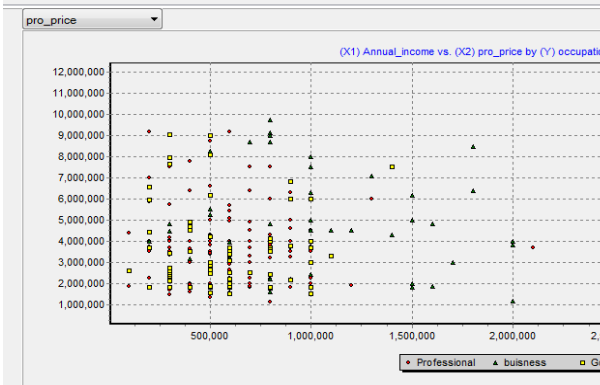

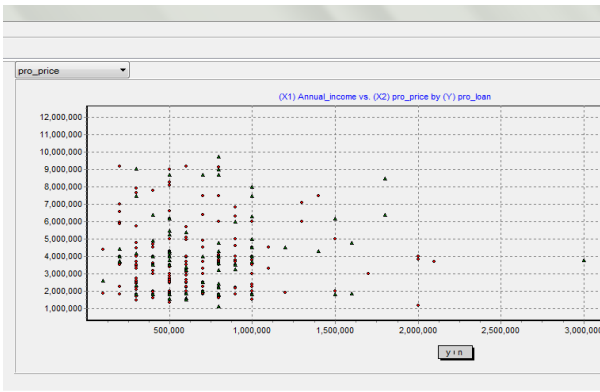Figure 3: Loan distribution in various sectors

Figure 4: Loan distribution

The analysis from Figure 5 and figure 6 indicates that a major portion of loans are taken by the people with professional occupations followed by government employee and businessman. By observing both the figures we can state that because the business segment has the highest income therefore they tend not to take loans as compared to other occupation segment.

Figure 5: loan distribution by group characterization

Figure 6: Annual Income by group characterization

Maximum property in the market are available but not furnished, these properties are not preferred to be taken on loans by buyers. Whereas property which are fully furnished and settled are the one bought by loans.

Figure 7: loan preference on property based on property status

Figure 8 shows that even with higher annual income the business occupation segment of people buy property which is under construction or ready to be used, this in fact can be considered as a form of investment.

Figure 8: loan preference on property based on occupation

Businessman buy property with the largest size and the property price whereas the professionals tend to buy property with lesser property size as compared to the business section

but the number of rooms in the property have higher priority than the cost of the property for the professionals. The average cost of the flat for professionals is higher as compared to government employees. On individually analysing each segment we observe that the businessman buy larger flats in terms of no. of bedrooms, property size and property prize. This is due to the fact that they have bigger incomes. Professionals tend to buy medium sized property with less property cost and which number of bedrooms as compared to the property bought by government employees.

Figure 9: Group characteristics of various attribute

CONCLUSION

This study was intended to provide some indication of the nature of the knowledge discovery process. After our research we came to the conclusion that people don't tend to buy high price property irrespective of their annual income. For property <20lac people don't take loans .Loans are mostly taken by professionals and govt. employee for small property size. Most of the government employee and professional have an annual income less than Rs.12lac tend to buy property less than 50lac but still most of the business persons have annual income more than Rs.12lac buy property less than Rs.60lac .We observe that the loans are not commonly taken by individuals for property less than Rs. 20lac irrespective of the annual income. Even if most of the loans are taken by professionals but the cost of the property for which loans have been taken is lower. Loans are mostly taken up by professionals for either available property or for properties under construction. But the cost of the property on which the loan is taken is lower as compared to the loans taken by business personals.

Extending researches in this area can provide banking scheme distribution over various investment and loan department. By including attributes like the number of kids in family or the max age of family members can help banking sector easily understand their loan targets.

References

[1]   Y. Ramamohan, K. Vasantharao, C. Kalyana Chakravarti, A.S.K.Ratnam" A Study of Data Mining Tools in Knowledge Discovery Process" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012.

[2]   Ramesh Gupta, Gaurav Dubey Sanjeev Choudhary "Developing a Rule Based Model for product analysis Using Frequent pattern analysis" International Journal of Advanced Research in Computer Science and Software Engineering ISSN: 2277 128X Volume 3, Issue 5, May 2013

[3]   http://tdan.com/

[4]   http://www.thearling.com/

[5]   http://www.tutorialspoint.com/

[6]   http://eric.univ-lyon2.fr/

AUTHORS BIOGRPHY

Ms Pooja Pant is presently pursuing M.Tech in the department of computer science and engineering from Amity University, Noida. She has done B.Tech degree course in Computer Science and Engineering from Ansal Institute of Technology, Gurgaon in 2015 under the affiliation of the Guru Gobind Singh Indraprastha University. Now she has focused her research interest in various aspects of Data Mining and Warehouse