Decision support system for loan based using data mining

P.Yashwanth[1], Mr.G.Anjanbabu[2]
[1] *Student, Dept. of MCA, SVU college of cm & cs*
[2] *Professor of Dept. of MCA SVU college of cm & cs, Tirupati, A.P*

Abstract- **A Decision Support Systems (DSS) is a particular type of computerized information system that support business and organizational decision making activities. on the other hand, Data Mining (DM) expand the potentials for decision support by finding styles and connections hidden in the data and in this way enabling an inductive way to deal with data analysis. Data is analyzed through a mechanized process, known as Knowledge Discovery in data mining techniques. Data mining can be characterized as a process of browsing and analysis for large amounts of data with a particular focus on discovering significantly important patterns and rules. Data mining helps discovering knowledge from raw, not equipped data. Utilizing data mining techniques permits extracting knowledge from data mart, data warehouse and, specifically cases, even from operational databases. In this paper a methodology is introduced to integrate the DSS with DM for loans to the Real Estate developments fund (REDF) Customers. It causes to cooperative interaction of DSS, through getting more options to analysis, utilizing expert's data, and improving assessment process. So I will talk about the function of data mining to simplify decision support, the utilization of data mining methods in decision support systems, talking about applied approaches and present a data mining decision support system called DMDSS – (Data Mining Decision Support System). We also present some obtained results and plans for future development.**

I. INTRODUCTION

In straightforward words, information mining is characterized as a procedure used to separate usable information from a bigger arrangement of any crude information. It suggests breaking down information designs in vast clusters of information utilizing at least one programming. Information mining has applications in numerous fields, similar to science and research. As an utilization of information mining, organizations can take in more about their clients and grow more powerful procedures identified with different business capacities and thusly use assets in a more ideal and quick way. This encourages organizations be nearer to their target and settle on better choices. Information mining includes successful information gathering and warehousing and additionally PC handling. For sectioning the information and assessing the likelihood of future occasions, information mining utilizes refined scientific calculations. Information mining is otherwise called Knowledge Discovery in Data (KDD).

Affiliation Rule Mining is a promising procedure which has the intend to discover fascinating and helpful examples from the value-based database. Its principle application is in advertise bushel investigation helping in distinguishing examples of each one of those things that are bought together. Mining straightforward affiliation rules includes less unpredictability and considers just the nearness or nonattendance of a thing in an exchange. Quantitative affiliation mining indicates relationship with thing sets and their amounts. To discover such affiliation rules including amount, we segment everything into equispaced containers with each receptacle speaking to an amount extend.

II. RELATED WORK

Mining of regular thing sets is an imperative stage in affiliation minining which finds visit thing sets in exchange database. It is the center in numerous undertakings of information mining that attempt to discover intriguing example from datasets.

III. TECHNIQUES

A.Classification

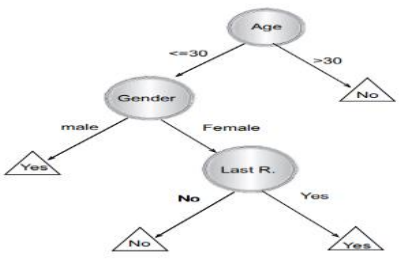
Classification is a task that happens as often as possible in everybody life. Basically it includes partitioning up objects so that each is assigned to one of a number of mutually exhaustive and exclusive classifications known as classes.

C.k-Nearest Neighbor

Nearest Neighbor[9] classification is fundamentally utilized when all attribute values are continuous, and sometimes may be modified .

B. Decision Tree

A choice tree is a classifier communicated as a recursive segment of the occurrence space. The choice tree comprises of hubs that frame an established tree, which means it is a coordinated tree with a hub called "root" that has no approaching edges. Every other hub have precisely one approaching edge. A hub with active edges is called an inside or test hub.All other nodes are called leaves (also known as terminal or decision nodes).[7] In a decision tree, each internal node splits the instance space into two or more sub-spaces according to a certain discrete function of the input attributes value[10]s. In the easiest and most continuous case, each test thinks about a solitary quality, with the end goal that the occurrence space is apportioned by the characteristic's esteem. On account of numeric qualities, the condition alludes to a range. Each leaf is doled out to one class speaking to the most suitable target esteem. Then again, the leaf may hold a likelihood vector demonstrating the likelihood of the objective quality having a specific esteem. Examples are grouped by exploring them from the foundation of the tree down to a leaf, as indicated by the result of the tests along the way. Figure 9.1 portrays a choice tree that reasons regardless of whether a potential client will react to an immediate mailing. Interior hubs are spoken to as circles, while leaves are signified as triangles. Note that this choice tree consolidates both ostensible and numeric properties. Given this classifier, the expert can foresee the reaction of a potential client (by arranging it down the tree), and comprehend the behavioral attributes of the whole potential clients populace with respect to coordinate mailing. Every hub is named with the quality it tests, and its branches are named with its comparing esteems.



IV. APRIORI ALGORITHM

General Process Association rule generation is usually split up into two separate steps:

1. First, minimum support is applied to find all frequent item sets in a database.
2. Second, these frequent item sets and the minimum confidence constraint are used to form rules.

As is regular in affiliation lead mining, given an arrangement of thing sets (for example, sets of retail exchanges, each posting singular things acquired), the calculation endeavors to discover subsets which are normal to no less than a base number C of the thing sets. Apriori utilizes a "base up" approach, where visit subsets are expanded one thing at once (a stage known as competitor age), and gatherings of hopefuls are tried against the information. The calculation ends when no further fruitful expansions are found. [12]Apriori utilizes broadness first inquiry and a tree structure to tally hopeful thing sets proficiently. It produces applicant thing sets of length k from thing sets of length $k - 1$. At that point it prunes the applicants which have an occasional sub design. As per the descending conclusion lemma, the competitor set contains all incessant k-length thing sets. From that point forward, it checks the exchange database to decide visit thing sets among the applicants.*A. AprioriA. Algorithm Pseudo code*

Procedure Apriori (T, minSupport) {//T is the database and minSupport is the minimum support
L1= {frequent items};
For (k= 2; Lk-1 !=∅; k++) {
Ck= candidates generated from Lk-1
//that is Cartesian product Lk-1 x Lk-1 and eliminating any k-1 size item set that is not
//frequent
For each transaction t in database do{
#increment the count of all candidates in Ck that are contained in t
Lk = candidates in Ck with minSupport
}//end for each
}//end for
Return Uk, Lk;

B. Sample usage of Apriori algorithm

A huge market tracks deals information by Stock-keeping unit (SKU) for everything, and in this way can recognize what things are regularly acquired together. Apriori is a modestly proficient approach to

construct a rundown of successive acquired thing sets from this information. Give the database of exchanges a chance to comprise of the sets {1,2,3,4}, {1,2,3,4,5}, {2,3,4}, {2,3,5}, {1,2,4}, {1,3,4}, {2,3,4,5}, {1,3,4,5}, {3,4,5}, {1,2,3,5}. Each number compares to an item, for example, "spread" or "water". The initial step of Apriori is to check up the frequencies, called the backings, of every part thing independently:

Step 1:

Item	Support
1	6
2	7
3	9
4	8
5	6

Step2:

Item	Support
{1,2}	4
{1,3}	5
{1,4}	5
{1,5}	3
{2,3}	6
{2,4}	5
{2,5}	4
{3,4}	7
{3,5}	6
{4,5}	4

The calculation closes here on the grounds that none of the 3-triples created at Step 3 have de wanted backings.

V. CONCLUSION

These techniques as classification, clustering and decision tree. Here we captured the relationships in the form of rules called as Association rules and here we used the apriori algorithm for the analyzation of the dataset and for producing the output.

REFERENCES

[1] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” AI Mag., vol. 17, pp. 37–54, 1996.

[2] J. Han and M. Kamber, Data Mining: Concepts and Techniques. San Francisco, CA: Morgan Kaufmann, 2001.

[3] W. Dwinnell, “Data visualization tips for data mining: pattern recognition provides data insight,” PC AI Mag., vol. 16.1, pp. 51–57, 2002.

[4] Assessing Loan Risks: A Data Mining Case Study Rob Gerritsen.http://xore.com/casestudies/dm_at_usda_(itpro).pdf

[5] Discovering Golden Nuggets: Data Mining in Financial Application Dongsong Zhang and Lina Zhou.

[6] Identifying Potential Default Loan Applicants - A Case Study of Consumer Credit Decision for Chinese Commercial Bank1 Qiwei Gan, Binjie Luo Southwestern University of Finance and Economics, Chengdu, Sichuan, China Zhangxi Lin Texas Tech University Lubbock, TX, USA (2008)

[7] S. Taylor, Modeling Financial Time Series. New York: Wiley, 1986.[5] A. N. Refenes, A. D. Zapranis, and Y. Bentz, “Modeling stock returns with neural networks,” presented at the Workshop on Neural Network Applications and Tools, London, U.K., 1993.

[8] Taylor, James (2012). Decision Management Systems: A Practical Guide to Using Business Rules and Predictive Analytics. Boston MA: Pearson Education.