ISSN No. 0976-5697

Volume 2, No. 6, Nov-Dec 2011

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Design of a Bdp Tool Using Data Mining Techniques

N.Sree Ram*
Associate Professor
Lakireddy Balireddy College of Engineering
L.B.Reddy Nagar,
Mylavaram, India
sriramnimmagadda@gmail.com

Dr.J.V.R.Murthy
Profssor
Department of CSE
JNTU Kakinada,
Kakinada, India
mjonnalagedda@gmail.com

Dr.K.Satya Prasad
Rector, JNTU Kakinada
Kakinada, India
rector@jntuk.edu.in

Abstract: Now a day's most of the financial organizations facing a major problem to recover the money from the borrowers, it becomes the frightening to banks in some situations. As a financial intermediary, one of its roles is to reduce lending risks. Bank lending is an art as well as a science. Success depends on techniques used, knowledge and on an aptitude to assess both credit-worthiness of a potential borrower and the merits of the proposition to be financed. In recent years, banks have increasingly used credit-scoring techniques to evaluate the loan applications they receive from consumers financial institutions always utilized the rules or principles built by the analysts to decide whom to give credit. In order to overcome these difficulties while recovering money the financial institutions and researchers have been developed various credit scoring models but they many not exactly fix in the situation like predicting the borrower attitude. Even though they are following rules and principles while lending money, they are unable to recover the loans from all the borrowers. In order to overcome these types of potential problems, as a precautionary measure, a software tool can be developed using Data Mining techniques aiming at giving qualitative and useful guidelines to the financial institutions while making the decision of money lending. This proposed work is aimed at designing a software tool to facilitate the effective money lending process by automating the prediction of customer attitude towards the money management and automation of decision making process.

Key words: money lending, customer attitude, software tool, automation of decision making, data mining techniques

I. INTRODUCTION

In earlier days financial organizations did not have an appropriate and efficient strategy to ascertain the customer's attitude and assess the customers whether he is capable to remit the borrowed money or not. The earlier process done through manually and it is very time consuming process and it was also inaccurate in some situations. Generally bankers lend the money based upon the security, surety and recommendations. Now the bankers use certain strategies to assess the customer attitude. Depending on that they took the decision whether he is eligible to take the loan or not. Several quantitative methods [1] have been developed for credit admission decision. The credit scoring models [1] are developed to categorize applicants as either accepted or rejected with respect to the applicants' characteristics. The objective of credit scoring models is to assign credit applicants to either a 'good credit' group that is likely to repay financial obligation or a 'bad credit [1] group whose application will be denied because of its high possibility of defaulting on the financial obligation.

The statistical methods, nonparametric statistical methods, and artificial intelligence approaches [5] have been proposed to support the credit decision [1]. Although the financial institutions use various methods to resolve the risks in money lending they are facing various difficulties in different situations. In order to solve such type of typical problems in money lending there is an emerging need to develop a software tool [2] which automates the decision making process in money lending. The proposed work is aimed at design of an automation toll which is very helpful in money lending [1][5] is known as Bad Debts Prediction (BDP) tool. The basic idea in designing of this tool is using the some data mining techniques [2] [7]. By using this tool there is a possibility to reduce the time in decision making and it can also possible to restrict the frauds in money lending. This tool provides simple user interface to the stake holder through this he can input the required parameters and can quickly get the decision. Designing of this tool is aimed at

a. To make efficient and qualitative decision while money lending
b. To speed up the decision making process
c. To reduce the risks in money lending
d. To prevent the fraud taken place in money lending process

II. BACKGROUND WORK

This BAD DEBT PREDICTION TOOL can easily asses the customer attitude depending upon the previous history [4], age, income as inputs and find the accuracy of customer[3] to remit the loan. It can make the make the decision whether the loan is safe or risky by using previous history of the customer, age and income of the customer as inputs. As assumptions the customer who applied for the loan must have an account in corresponding bank. Every

customer may have any number of loans based upon their integrity (60%). All these operations are done by CIBL (Credit Information Bureau Ltd). Different banks maintain different CIBL teams they alone can take the decision based upon field officer's presentation report. Every bank database tied up with all bank databases for previous history report of the customer.

A. CIBL:

The main task of CIBL in any financial institutions is to make the decision in money lending. Depending upon the customer details provided by the corresponding branches CIBL assess the attitude of customer by considering his history and it also makes the final decision that is whether granting loan or not by considering the customer previous history. The main aim of introducing this proposed work is to automate the functionality of the CIBL. While CIBL making decisions regarding money lending this proposed work can make the decisions as early as possible and as more accurately.

III. TOOL IMPLEMENTATION

The proposed work is aimed at design of an automation toll, which is very helpful in money lending process is known as Bad Debts Prediction (BDP) tool. The basic idea in designing of this tool [5] is using the some data mining techniques. By using this tool there is a possibility to reduce the time in decision making and it can also possible to restrict the frauds in money lending process. Before employing this tool in decision making process [6] there is need to construct training set from previous customer's history. So that CBIL can give its database as input to this tool and the tool can construct training set from the database. The CBIL officer needs to analysis of data to know which loan applications are safe and which are risky with the help of this tool. This tool works by employing one of the data mining techniques known as classification.

A. Classification:

Classification [7] is the process of giving known class labels to each and every tuple in the given data set. Data classification is a two steps process. In the first step a classifier is built describing a predetermined set of data classes. This is the learning step, where a classification algorithm builds the classifier by analyzing or "learning from" a training set made up of database tuples and their associated class labels. The individual tuples making up the training set are referred to as training tuples and are selected from database under analysis. This first step of classification can also be viewed as the learning of mapping or function y=f(x), that can predict the associated class label y of given tuple x. The following figure 1 illustrates the process of learning in classification. The outcome from the learning step is a classification model to which test data applied in the classification step.

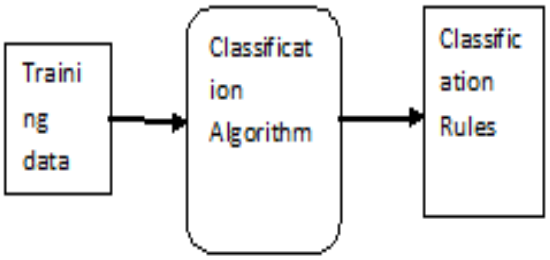

Figure 1: construction of classification model

In the second step test data are used to estimate the accuracy of the classification rules. If the accuracy is considered acceptable, the rules can be applied to the classification of new tuples. This process is known as classification. The following figure 2 illustrates the process of classification. There are different kinds of classification algorithms such as Decision tree based classification, basin classification, classification by back propagation etc.

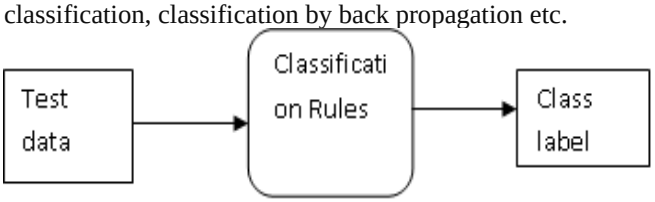

Figure 2: classification

By using any one of the classification algorithms the tool first construct a decision tree from the given training dataset by forming some classification rules. It is assumed that the training dataset is frequently updated in the CBIL database. Whenever CBIL receives a loan application from a branch first it retrieves the customer's data from loan application and considers it as test data. This test data collected from loan application is given as input to the BDP tool to predict the class label of the loan application. Finally BDP tool gives the decision that the granting loan to customer is safe or risky. This tool is employees the classification by decision tree induction.

B. Decision tree induction:

Decision tree induction is the learning of decision trees from class labeled training tuples. A decision tree is flow chart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test and each leaf node holds a class label. The top most nodes in a tree is a root node. The following figure 3 represents a typical decision tree.

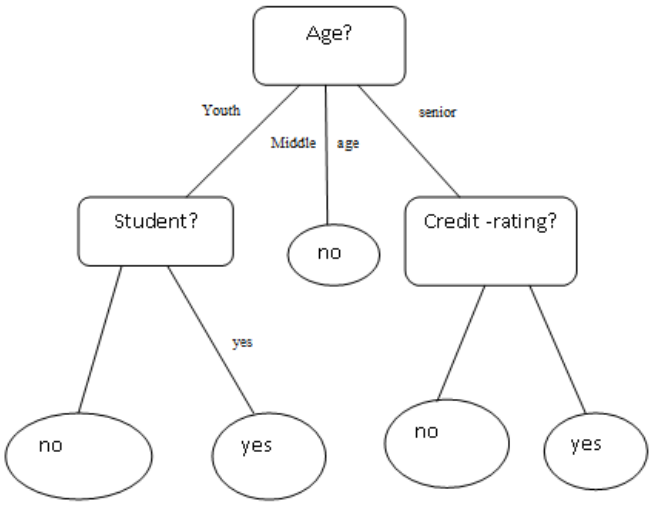

Figure 3: A decision tree

There are different algorithms to construct a decision tree. Some of the algorithms produce binary trees and some others produce non binary trees. The decision tree works as follows. Given a tuple X with unknown class label, the attribute values of the tuple test against the decision tree. A path is traced from root node to the leaf node which holds the class label of that tuple. These decision trees can easily convert to the classification rules. The construction of decision tree not requires any domain knowledge. The following is the algorithm to generate decision tree from the training tuples of the data partition D

C. *Algorithm:* Decision Tree Construction [7]:

Input to this algorithm

a. Data partition D, which is a set of training tuples and their associated class labels
b. Attribute-list the set of candidate attributes
c. Attribute-selection method, a procedure to determine the splitting criterion that "best" partition the data tuples into individual classes. This criterion consists of a splitting-attribute and possibly either a split point or splitting subset.

Output: A decision tree [7]

Method:

i. create a node N
ii. if tuples in D are all the same class C, then
iii. return N as leaf node labeled with the class C
iv. if attribute-list is empty then
v. return N as leaf node labeled with the majority class in D
vi. apply attribute-selection method(D, attribute-list) to find the "best" splitting criterion
vii. label node N with splitting-criterion
viii. if splitting-attribute is discrete-valued and multi way splits allowed then
ix. attribute-list← Attribute-list - splitting-attribute
x. for each outcome of J of splitting-criterion
xi. let DJ the set of data tuples in D satisfying outcome J
xii. if DJ is empty then
xiii. attach leaf labeled with majority class in D to node N
xiv. else attach the node returned by generate-decision-tree(DJ, attribute-list,) to node N end for
xv. return N

This algorithms employees an attribute selection measure is a heuristic for selecting the splitting criterion that best separates a given data partition D. There are a set of attribute selection measures available such as information gain, gain ration and gini index. In this tool the algorithm used information gain attribute selection measure.

D. *Information Gain:*

Let node N represents or holds the tuples of the data partition D. the attribute with the highest information gain is chosen as the splitting attribute for node N.

Info (D) =-∑mi=1 Pi log2 (Pi) [7]

Where Pi is the probability that an attribute in tuple in D belongs to class Ci and is estimated by

| CiD| / |D|

A log function base 2 is used because the information is encoded in bits. Info (D) is just the average amount of information needed to identify the class label of tuple in D. In addition to this there is need to calculate how much more information still need in order to arrive at an exact classification? This amount is measured by

Info A (D) = ∑j=1v |D j| / |D| * Info (D j) [7]

Now information gain is defined as the difference between original information required and new required

Gain (a) = Info (D) - Info A (D) [7]

Thus by using information gain attribute selection measure the BDP tool chose an appropriate attribute to split the tuples in data partition D.

The following figure 4 represents that the various entities and attributes involved in the BDP tool.

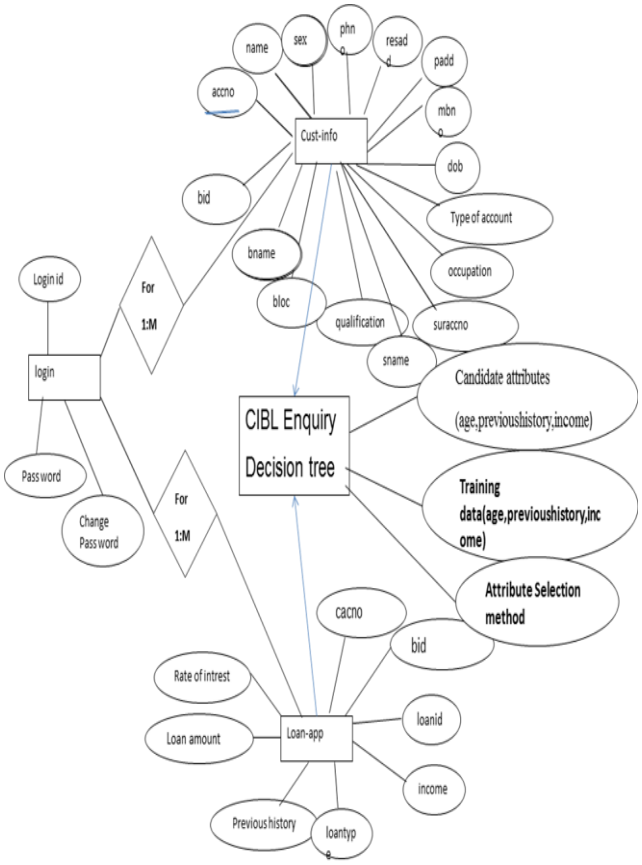

Figure 4:entites and attributes involved in tool

IV. EXPERIMENTAL RESULTS

We have carried out some experiments [8] in order to evaluate the performance and usefulness of this tool on

information in database of the CBIL for predicting class label of the loan application. Our objective is to classify the creditors in order to determine the riskiness in the money lending process [5]. The following are the some of the screenshots of the BDP tool

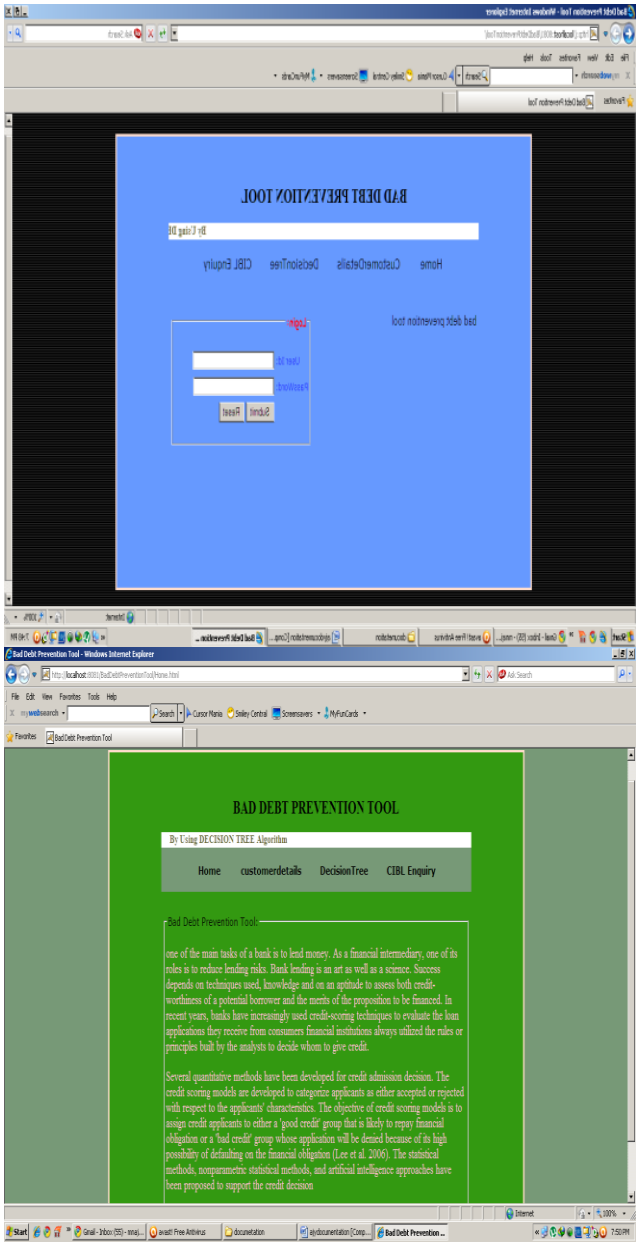

Figure 5: sample screen shots of BDP tool

V. CONCLUSION & FUTURE SCOPE

The main objective of the proposed work is to improve the efficiency and accuracy in decision making of money lending process. This toll is mainly useful to CBIL in financial institutions. The proposed work assumed that the customer's database of banks is frequently updated. It can also be possible to extend the scope of this work to a financial expert system by applying fuzzy clustering.

VI. REFERENCES

[1]. Adel Lah sasna, Raja Noor Ainon and The Ying Wah "Credit scoring methods with soft computing" International Journal of Information Technology vol. 7 April 2010 pp. 115-122.

[2]. Arun k pujari "Data Mining Techniques" computer science Univwersities Press

[3]. Cristóbal Romero, Sebastián Ventura, Pedro G. Espejo and César Hervás "Data Mining Algorithms to Classify Students" Computer Science Department, Córdoba University, Spain

[4]. Colgate, Mark and Lang, Bodo "Switching barriers in consumer market an investigation of consumer marketing vol 18. No 4 pp 332-347

[5]. Eric Rahmian, Seema singh, Thongchair Thanmachote , Rajiv Virmani " Bankruptcy Prediction by Neural Networks"

[6]. Gerben w. Dekker "Predicting student dropout" A case study April 2009 Depatment of Electrical Engineering Eindhoven Technology

[7]. Jiawei Han and Micheline Kamber "Data Mining concepts and Techniques" second edition Elesvier

[8]. Leslie Scism and Mark Maremont "Insures Test data Profiles to Idetify Risky clients"

[9]. M.J.Xavier, Sundaramurthy, PK viswanathan, G.Balasubramanian "Improving Prediction Accuracy of loan default" A case study in rural credit, Institute for Financial Management and Research IFMR