



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

Improving the Efficiency of Loan Analyzer Using AI

Madhuri Mahajan¹, Pooja Barhate², Jitendra Bunde³, Hemant Zope⁴, Ass. Pro. S. D. Rajput⁵

U.G. Student, Department of Computer Engineering, SSBT's COET Bambhori, Jalgaon, India¹⁻⁴

Department of Computer Engineering, SSBT's COET Bambhori, Jalgaon, India⁵

ABSTRACT: The objective of this paper is to calculate the percentage risk and to determine whether the applicant can get a loan, to reduce the efforts during the process of loan, and it also determine the suitable amount of loan that can be borrowed by the applicant. Loan analysis is a big problem for the banks while processing their work as well as for customers. It's a time taking process and also it is very risky. Using artificial intelligence, we can calculate the percentage risk and overcome this problem. By using Naive Bayes we can determine whether the applicant is suitable for borrowing the loan or not. Also we are increasing its accuracy of prediction.

KEYWORDS: Training, Artificial Intelligence, Prediction, Naive Bayes.

I.INTRODUCTION

The main work of the banks is to take money from the people at a particular rate of interest and then lending it to others at a rate which is higher than the rate taken from the people. This way banks function in a healthy economy. But real problem comes in the scene when people borrow money from the bank but in the end are not able to repay the loans leading to problem for the banks.

This has been a very major cause of concern especially when the economy of the world is moving at a very fast pace. Bad Loans reduces the efficiency of banks in any financial setup and is very bad for the economy. The country is more economically developed if the banks are better. Banks are the main components of an economy. Banks face problems when people don't return loan within due date. This likelihood of non-repayment of loans at the payable date is termed as Credit Risk.

The credit risk has long been an important and widely studied topic in banking. For lots of commercial banks, the credit risk remains the most important and difficult risk to manage and evaluate. In the last years the advances in information technology have lowered the costs of acquiring, managing and analyzing data, in an effort to build more robust and efficient techniques for credit risk management.

The advantage of Credit Rating would be such that if the customer does not possess the sufficient credit rating, he will not be granted the loan. Likewise if the customer does possess the required credit rating he will be granted the loan.

Borrowers usually have the better information for the projects to be financed, but lenders usually don't have sufficient information about the projects. If a bank has good customers it is more likely to increase its efficiency and allow the banks to lend more but here can be a contradictory case of banks itself going bankrupts when they face bad loans and bad customers. Also customer itself can determine whether he/she will grant a loan or not.

Nowadays many banks approves loan after a long process of verification and validation but still there is no surety whether the applicant is suitable for borrowing the loan or not. Through this system we can predict whether that particular applicant is safe or not. Loan prediction is very helpful for employees of banks as well as for applicant also. The main purpose of this system is to provide quick, immediate and easy to predict the loan status for applicant. It save the valuable time of applicant rather than applying to the banks because it takes lots of time in bank to process that application.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

II. RELATED WORK

Data mining is the process of analyzing data from different perspectives and extracting useful knowledge from it. It is the core of knowledge discovery process. The various steps involved in extracting knowledge from raw data as depicted in figure-2. Different data mining techniques include classification, clustering, association rule mining and prediction etc. [3]. 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 [4]. Fraud detection and credit risk applications are particularly well solved by classification technique. In classification, a training set is used to build the model as the classifier which can classify the data items into its appropriate classes. A test set is used to validate the model.[7]

Due to huge growth in data the banking industry deals with, analysis and transformation of the data into useful knowledge has become important task beyond human ability [5]. Data mining techniques can be used in solving business problems by finding patterns, associations and correlations which are hidden in the business information stored in the data bases [4].

By using data mining techniques to analyze patterns and trends, bank executives can predict, with increased accuracy, how customers will react to adjustments in interest rates, which customers are likely to accept new product offers, which customers will be at a higher risk for defaulting on a loan [2]. Globalization had led the banks focus towards customer retention and fraud prevention. Data mining is used to do so. By analyzing the past data, data mining can help us to predict credible customers. Thus they can plan for launching different special offers to retain those customers who are credible. Certain areas that effectively utilize data mining in banking industry are marketing, risk management and customer relationship management.

The Figure -1 shows the decision tree induction. A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

A decision tree is a flowchart-like structure in which each internal node represents a “test” on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label. The paths from root to leaf represent classification rules.

Tree based learning algorithms are considered to be one of the best and mostly used supervised learning methods. Tree based methods empower predictive models with high accuracy, stability and ease of interpretation. Unlike linear models, they map non-linear relationships quite well.

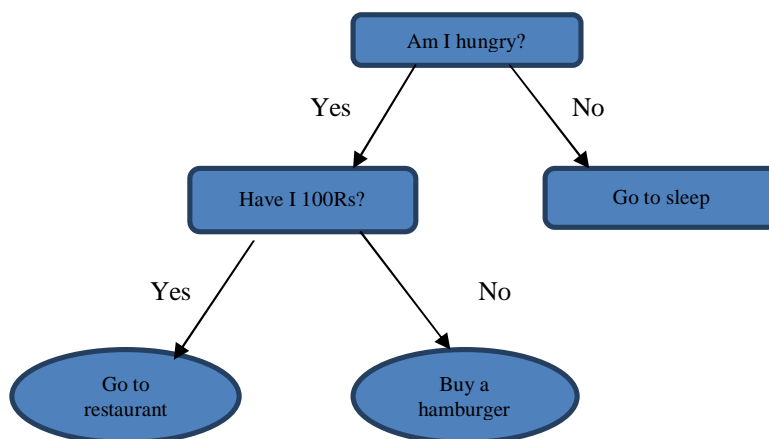


Fig. 1: Decision Tree Induction



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

In our model we have used Naive Bayes algorithm. As we know simple decision trees tend to over fit the training data more so that other techniques which mean you generally have to do tree pruning and tune the pruning procedures. You didn't have any upfront design cost, but you'll pay that back on tuning the trees performance. Simple decision trees divide the data into squares so building clusters around things means it has to split a lot to encompass clusters of data. Splitting a lot leads to complex trees and raises probability you are overfitting. Big trees get pruned back so while you can build a cluster around some feature in the data it might not survive the pruning process [1].

Naive Bayes works effectively over large dataset. It gives more accuracy than decision tree for large dataset.

III.METHODOLOGY

The proposed system focuses on predicting the feasibility of customers for granting loan by analyzing their behavior. The input to the model is the customer behavior collected. On the output from the classifier, decision on whether to approve or reject the customer request can be made. In this process it is required to train the data using Naive bayes algorithm and then compare user data with trained data to predict the nature of loan using Decision tree algorithm. The steps involve in Loan Analysis are represented as below.

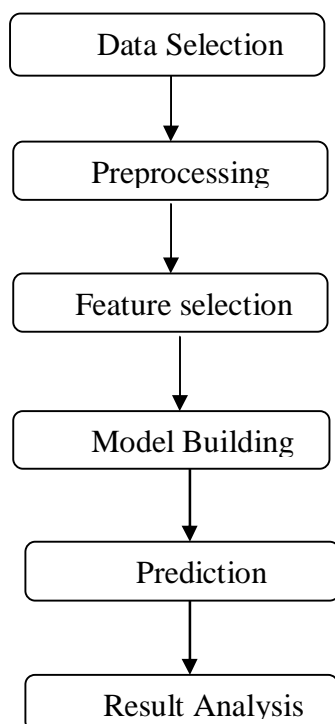


Fig.2 Major steps in Loan Analysis



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

IV. SYSTEM ARCHITECTURE

For deciding whether the applicant will get loan or not, there are some factors for deciding that. These factors are basically data which is provided by the applicant itself.

Table-1: Loan Deciding Factors

VARIABLE	DESCRIPTION
Loan_ID	Unique Loan_ID
Gender	Male/Female
Age	Applicant age
Marrital_Status	Applicant Married(Y/N)
Dependents	No of dependents
Education	Graduate/Under Graduate
Self_employed	Self_employed(Y/N)
Applicant_Income	Applicant_Income
Co_Applicant_Income	Co_Applicant_Income
Loan_Amount	Loan_Amount in Lacks
Loan_Amount_Term	Term of Loan in years
Credit_History	Meets guidelines
Property_Area	urban/Semi Urban/Rural
Loan_Status	Loan Approved(Y/N)

TheFigure-3 shows the architecture of our model. It contains the Data collecting, data preprocessing and cleaning and classification of the attributes.The Table-1 consists of 14 such factors from which our model predicts whether the applicant is able to take the loan or not.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirce.com

Vol. 7, Issue 5, May 2019

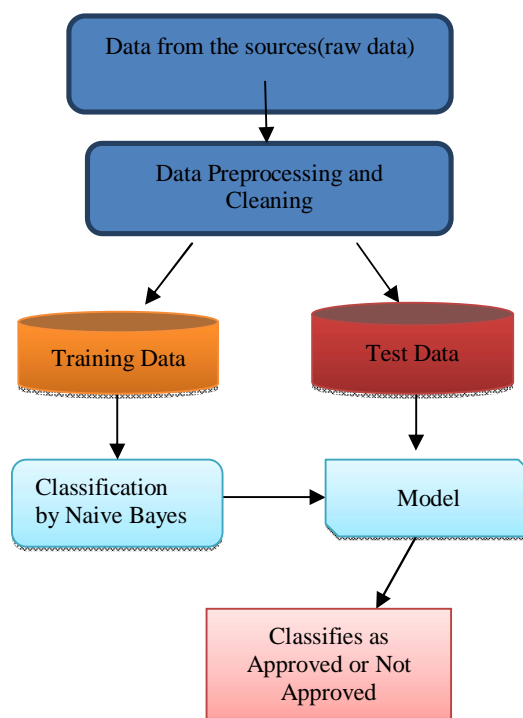


Fig. 3 Architecture of proposed model

V. DATA CLEANING AND PREPROCESSING

Data cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data. Data cleaning may be performed interactively with data wrangling tools, or as batch processing through scripting [6].

The dataset has many missing and imputed data which is replaced in this step. Data preprocessing is one of the critical step in data mining process which deals with preparation and transformation from the initial data set to the final data set. Data preprocessing is the most time consuming phase of a data mining process. Data cleaning of loan data removed several attributes that has no significance about the behaviour of a customer. Data integration, data reduction and data transformation are also to be applicable for loan data. For easy analysis, the data is reduced to some minimum amount of records.

For the above purpose we had used the Pandas (Software Library) and NumPy libraries. In particular, it offers data structures and operations for manipulating numerical tables and time series. Pandas is a Software Library written for the Python Programming Language for creating dataframe object for data manipulation with integrated indexing and provide tools for reading and writing data into different file formats(CSV files). NumPy is a Python Programming Library used for data alignment and integrated handling of missing data.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

VI. NAIVE BAYES ALGORITHM

Naive Bayes is a simple but surprisingly powerful algorithm from predictive Analysis. It is a classification technique based on base theorem with an assumption of independence amount predictors. It comprises of two parts which is Naïve and Bayes. In simple terms Naïve Bayes classifiers assumes that the presence of a particular feature in a class is unrelated to the any other feature. Even if this features depend on each other or open the existence of any other features all of this properties independently contribute to the probability whether a fruit is an apple or an orange or a banana. So that is why it is known as Naïve. Naïve Bayes model is easy to build and particularly useful for very large datasets. In probability theory and statistics base theorem which is ordinarily known as base law or base rule describes the probability of an event based on the prior knowledge of the condition that might be related to the event. Now base theorem is a way to figure out a condition probability. The condition probability is a probability of an event happening given that it has some relationship to one or more other events. For example your probability of getting a parking space is corrected to the time of day you park, where you park and what conventions are you going on that time. Base theorem gives us actual probability of an event given information about the test.

If you look at the definition of a base theorem we can see that Given a hypothesis H and evidence E, Bayes' theorem states that the relationship between the probability of the hypothesis before getting the evidence P(H) and the probability of the hypothesis after getting the evidence P(H|E) is

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

Posterior Probability= P(H|E)

Prior Probability= P(H)

Likelihood= P(E|H)

Evidence=P(E)

Steps of Algorithm:

Step 1: Start

Step 2: Create a frequency table for all the features against the different classes.

Step 3: Draw the likelihood table for the features against the classes.

Step 4: Calculate the conditional probabilities for all the classes.

Step 5: Calculate maximum probability.

Step 6: Stop.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

VII.RESULT

The following figure Fig.4 shows the accuracy of the naive bayes algorithm. It gives 88.61% accuracy which is better than the decision tree induction.

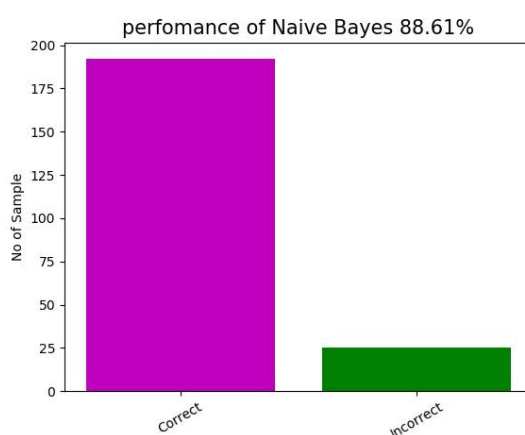


Fig. 4 Performance Graph of Naive Bayes

VIII. CONCLUSION

It is necessary to evaluate feasibility of customer before approving loan. The system is useful for bank as well as customers. It helps the bank management to enhance their decision making and reduce their losses by taking right decision and also filtering loan application at preliminary stage in order to reduce the work overhead. It is also useful for customer to check their feasibility by avoiding long bank process. We had used Naive Bayes algorithm to improve the efficiency of the system which accurately predict whether loan will get approved to customer or not.

REFERENCES

1. Sivasree M S and Rekha Sunny T., "Loan Credibility Prediction System Based on Decision Tree Algorithm", *International Journal of Engineering Research and Technology*, Volume 3, No. 9, Sep-2015
2. Dr. Madan Lal Bhasin, "Data Mining: A Competitive Tool in the Banking and Retail Industries", *The Chartered Accountant*, October 2006
3. Frawley, W. J., Piatetsky-Shapiro, G., and Matheus, C. J., "Knowledge discovery in databases: An overview.", *AI Magazine*, Volume13, No. 3 Page-57.
4. Bharati M. Ramageri, "Data Mining Techniques and Applications", *Indian Journal of Computer Science and Engineering* Volume 1, No. 4
5. P.Sundari, Dr.K.Thangadurai "An Empirical Study on Data Mining Applications", *Global Journal of Computer Science and Technology*, Volume 10, No. 5, Page 23, July 2010
6. S. Kotsiantis, D. Kanellopoulos, P. Pintelas, "Data Pre-processing for Supervised Learning", *International Journal of Computer Science*, Volume 1, No. 2, Paper page 111–117, 2006.
7. Vivek Bhambri "Application of Data Mining in Banking Sector", *International Journal of Computer Science and Technology*, Volume 2, No.2, June 2011