

Prediction of Loan Risk using Naive Bayes and Support Vector Machine

Vimala S.¹, Sharmili K.C.²

PG Scholar, Department of Computer Science, Stella Maris College, Chennai, India. ¹

Assistant Professor, Department of Computer Science, Stella Maris College, Chennai, India. ²

Abstract: In recent years, banks and creditors face lots of challenges associated with the bank loan. In addition, banking transactions are growing rapidly and large data sets that represent customer behaviour and risks around debt burdens have increased. So, it became mandatory to know the risks related to bank loans. Each consumer knows what they want and what kind of business they would like to choose. Customer can simply find another bank if administrators are not ready to pay attention. Many data analysis technique exists that focus on loan risk. These techniques are oriented toward extraction of quantity and statistical data properties. They facilitate useful data explanations and help to gain better insights in the process behind data. While there can be traditional data analysis techniques lead to intuition knowledge, it is still made by human researchers. Banks and many investment companies are pioneers using data mining. The proposed work in data mining focused on using data from banking sector to predict the status of loans. Two classification algorithms Naive Bayes and Support Vector Machine are discussed here. The results have been proven that the speed and accuracy has been improved when combined these two techniques.

Keywords: Bank, Loan risk, Data mining, Prediction, Naive Bayes, Support Vector Machine.

I. INTRODUCTION

Data mining is a highly motivated and important area of research, with the intention of obtaining large amounts of set of data collected. In the present era, Data Mining is popular in the banking sector because there are efficient analysis methods for identifying unknown and useful information in banks data. Due to tremendous data available, the main focus is on database ranking and protection in order to make key decisions.

When you receive money from a bank or financial institution, you will be required to repay the principal and future interest. The key is the amount you borrowed, and the interest bills charged by bank. Loans are usually secured or unsecured. A secured loan involves pledging an asset such as a car, boat or house as link for the loan. If the borrower makes a mistake or

does not repay the loan, the lender takes possession of the asset. An unsecured loan option is preferred, but not common. If the borrower does not repay the insecure debt, the lender does not have the right to withdraw anything. There are different types of loan available [5].

- Conventional Loans
- Non-Conforming Loans
- Unsecured Loans
- Conforming Loans
- Secured Loans
- Open-ended Loans

When approving a loan for a customer, the bank first verifies their profile and documentation. Figure 1 will show the process of loan approval in bank.

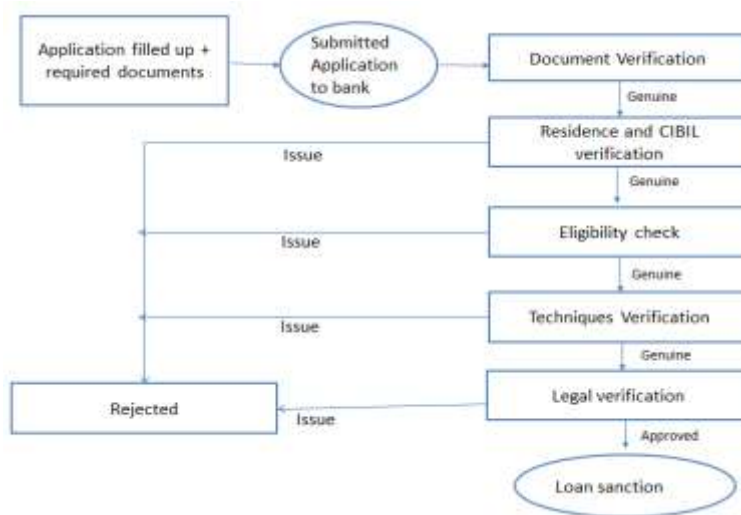


Fig. 1 Process for Loan Sanction

A. Credit Evaluation:

Any bank maintains a credit score for each customer the credit score is expressed as numbers based on an individual's credit file [2]. This process is known as a credit evaluation,

which takes time, but ending is either an approval or a rejection. The heart of bank is credit activity, which blow up under marking conditions. The two main reasons for the need for an expert support system are the absence of accurate methods of

measuring the lack of public credit risk system and credit risk in many banks [3].

B. Related Risk:

Many risks associated with bank loans, for bank and for those who get the loans. The analysis of the risk in bank loans should be understood as the meaning of the risk. credit risk, the risk that the loan won't be return back on time or at all; liquidity risk, the risk that too many deposits will be withdrawn too quickly, leaving the bank short on immediate cash; and interest rate risk, the risk that the interest rates priced on bank loans will be too low to earn the bank adequate money [5].

II. RELATED WORK

In [1] Cheng-Lung Huang a,*, Mu-Chen Chen b, Chieh-Jen Wang c, describe the "Credit scoring with a data mining approach based on support vector machines". This paper they used algorithm Support vector machine for credit score evolution. SVM based approach can be a credit rating model properly categorized regulated or rejected applications, reducing the risk of lenders and substantially translating into future savings.

In [2] KetakiChopde, Pratik Gosar, ParasKapadia, NiharikaMaheshwari, Pramila M. Chawan, "A Study of Classification Based Credit Risk Analysis Algorithm". Form this they discuss about the Credit score modeling is to categorize loan applicants into two classes: "Good Credit" class liability to lend financial obligation and "bad credit" classes. Accurate judgment of applicants' credit qualification allows financial institutions to increase the amount of credit provided when reducing potential losses. Then discuss about the various implementation methods of decision trees for credit risk analysis.

In [3] Sudhakar M, Dr. C. V. Krishna Reddy, Describe the "Credit Evaluation Model Of Loan Proposals For Banks Using Data Mining Techniques", Loan risk one of the biggest problem for the bank.is the technique of finding optimal solution for the problem The credit approval of financial resources is an important aspect of any financial assessment as it concerns the economy of a country. Before getting a borrower, he must be evaluated in some areas. Five sisters are involved in credit rating. Research findings suggest that SVM, End Tree and Logistic Recession are the best method of classifying loan applications. By analyzing the performance of these models in the standard database, missing data multilayer perceptron model and the logistic regression has been found to be good.

In [4] MrunalSurve, PoojaThitme, PriyaShinde, Swati Sonawane, SandipPandit, describe the "Data Mining Techniques To Analyses Risk Giving Loan(Bank)". The traditional lending system has a lot of manual work, which is a difficult task and can lead to some mistakes that may lead to risks. The main idea behind the program is to analyze the risk ratio when calculating the credit rating. Overall an innovative technique to analyses the risk while giving loan(bank) is presented here. The proposed system C4.5 is used to take a decision on the risk of lending to the customer. This reduces

human efforts and makes the banking system more effective. This results in improving customer retention.

In [5] AboobydaJafar Hamid and Tarig Mohammed Ahmed, describe the "Developing Prediction Model Of Loan Risk In Banks Using Data Mining", In this paper, This model has been created using a data banking system to foreclose loans. Three methods are used to create the proposed model: j48, Bayes Net and naive Bayes. By using the Weka application, the sample is processed and tested. The J48 algorithm is the best algorithm of the credit class after the implementation of the data mining techniques protocol J48, Bayes Net and naive Bayes. The Result is based on accuracy.

III. PROPOSED METHOD

For implementation of this paper we collect a dataset from UCI, they provide an original dataset germen credit. This dataset have 21 attributes with their respected class. We predict the dataset risk by using Naïve Bayes and Support vector machine.

A. Naïve Bayes

Naive Bayes is a simple technique based on the Bayes theorem, which is predicted for independent speculation. Because it is elegantly simple and robustness, it is widely used for classifying purpose. It uses the probability theory to classify data. Naive Bayes helps to develop the model that provides predictive abilities, providing a new way to understand data. All of these input attributes should be relatively independent. It is easy to create a Naive Bayes model and do not have complex refutation parameters [5].

B. Support Vector Machine

Support vector machines were created from statistical learning theory [1]. Support Vector Machine is a type of learning system algorithm that is used to make the classification more accurate. SVM have been widely applied to field's handwriting digit, character and text recognition, and more recently to satellite image classification. The essence of SVM The hyper plane is also known as the "Decision-range or Decision surface", and this high-plane tutor's model of data divides positive and negative [6].

C. Steps for NB and SVM

Step1: By using Naïve Bayes calculate the probabilities of each attribute's distinct values with the combination of class values.

Example: from dataset finds a probability of checking_status value <0 has class value as good so,
$$p(\text{checking_status}=<0|\text{class}=\text{good})=\frac{\text{total_number_of_co_mbination}(<0,\text{good})}{\text{total_number_of_good}}$$

Like that we have to find probabilities for each distinct value in the dataset expect class. From these got a model of dataset by applying Naive Bayes.

Step 2: Predict the model with original data set.

Step 3: Apply the prediction value to the Support Vector Machine.

Step 4: Again find the prediction value by what the model got from the previous step with original data.

Figure 2 show the combined prediction of Loan risk. To Processing the result of Naïve Bayes with the support vector machine we improved the speed and accuracy.

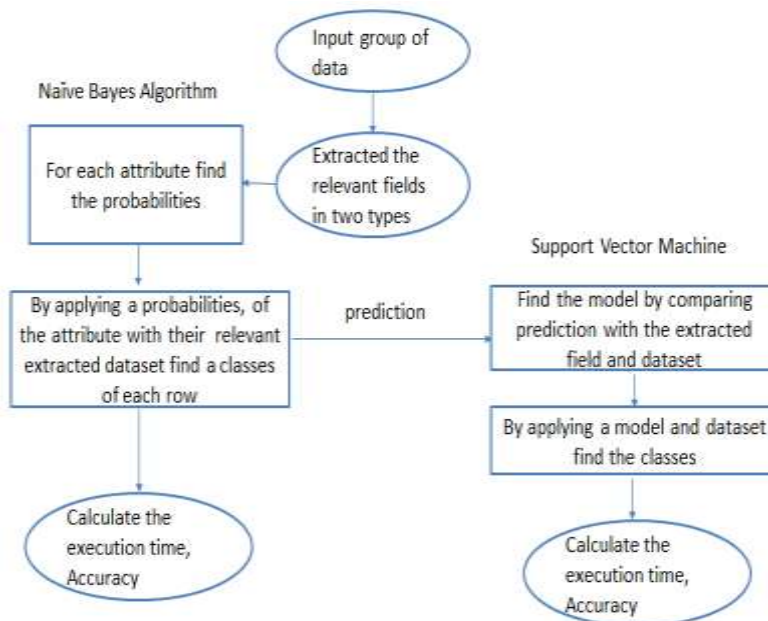


Fig. 2 Proposed Method Taxonomy

IV. EXPERIMENTAL RESULTS

In this paper we predict the Loan risk by using Naïve Bayes and SVM. Here we compare both the result of Naïve Bayes and

SVM separately and with a fusion of Naïve Bayes and Support vector Machine. First we compare the accuracy what we had form this experiment.

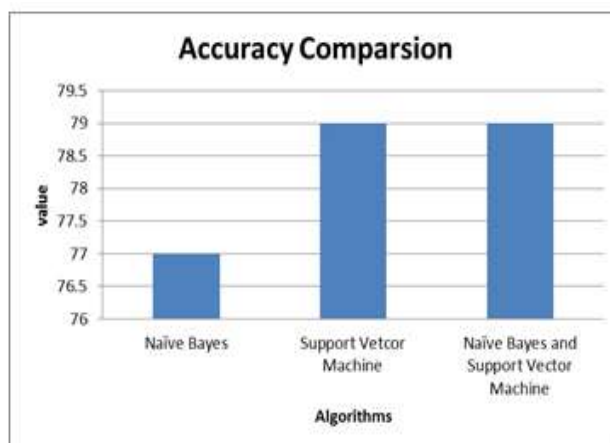


Fig. 3 Accuracy Comparison

From Figure: 3 we can analysis Naïve Bayes had a low accurate comparing to another two results. Result of SVM and NBSVM gave a same accurate result so they are in same level. Next we check the execution time of these algorithms.

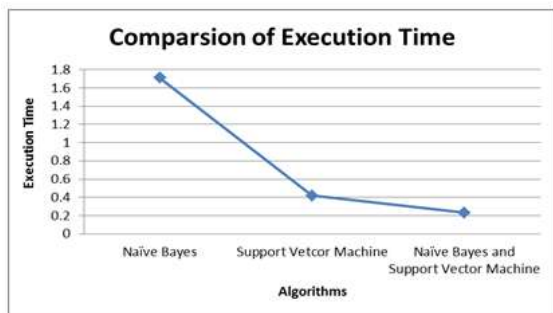


Fig. 4 Comparison of Execution time

By comparing execution time (Figure 4) of three methods, Naive Bayes had taken more time to execute the model comparing to others. The Combination of NB and SVM executive time is less than the SVM. Support Vector Machine had a time of (2* NBSVM) execution time. Figure 5 show the final experimental result of three algorithms accuracy and speed. Comparing both the execution speed and accuracy NBSVM gave a best result.

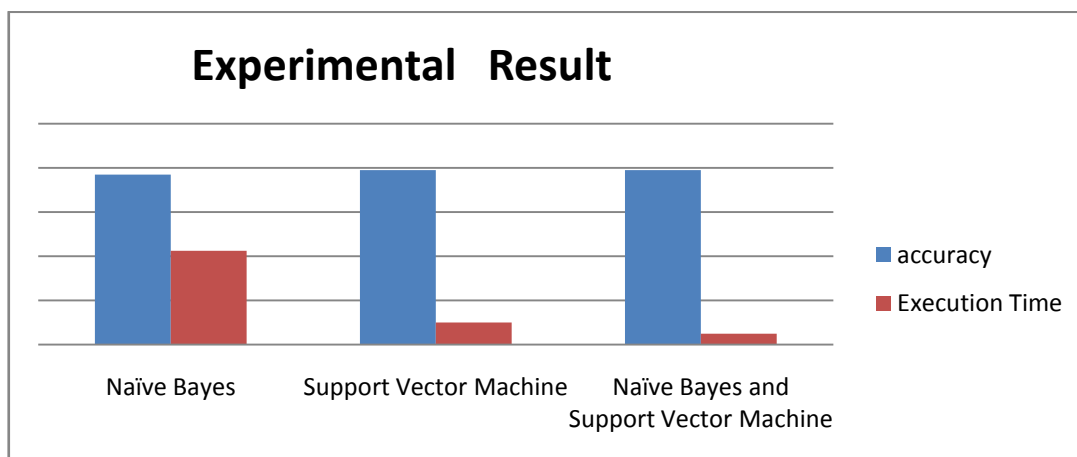


Fig. 5 Experiment Result

V. CONCLUSION

This paper has been discussed about Prediction of loan risk model By Combination of Naïve Bayes and Support Vector Machine. For a Prediction, data mining has different type of classification. From these classification techniques, Naïve Bayes elegantly Simple and robustness, it is widely used for classifying purposes for this reason. It uses the probability theory to classify data. Support Vector Machine is a type of learning system algorithm that is used to make the classification more accurate. To improving the accurate and speed of Naïve Bayes we used SVM. Analysis studies and our findings confirm that the correctness and efficiency of the proposed method has been improved for greater expansion in future.

REFERENCES

[1] Cheng-Lung Huang a*, Mu-Chen Chen b, Chieh-Jen Wang c, “Credit scoring with a data mining approach based on support vector machines”, ELSEVIER, C.-L. Huang et al. / Expert Systems with Applications 33 (2007) 847–856.

[2] KetakiChopde, Pratik Gosar, ParasKapadia, NiharikaMaheshwari, Pramila M. Chawan, “A Study of Classification Based Credit Risk Analysis Algorithm”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume- 1, Issue-4, April 2012.

[3] Sudhakar M, Dr. C. V. Krishna Reddy, “CREDIT EVALUATION MODEL OF LOAN PROPOSALS FOR

BANKS USING DATA MINING TECHNIQUES”, International Journal of Latest Research in Science and Technology ISSN (Online):2278-5299 Volume 3, Issue 4: Page No.126-131. July- August 2014.

[4] MrunalSurve, PoojaThitme, PriyaShinde, Swati Sonawane, SandipPandit, “DATA MINING TECHNIQUES TO ANALYSES RISK GIVING LOAN(BANK)”, IJARIII- ISSN(O)-2395-4396 Vol-2 Issue-1 2016.

[5] AboobydaJafar Hamid and Tarig Mohammed Ahmed, “DEVELOPING PREDICTION MODEL OF LOAN RISK IN BANKS USING DATA MINING”, Machine Learning and Applications: An International Journal (MLAIJ) Vol.3, No.1, March 2016.

[6] Dr.B.Srinivasan ,K.Pavya, “Diagnosis of Thyroid Disease Using Data Mining Techniques: A Study”,International Journal of Computer Science and Mobile Computing, International Research Journal of Engineering and Technology (IRJET) e- ISSN: 2395 -0056 Volume: 03 Issue: 11 | Nov -2016 www.irjet.net p-ISSN: 2395-0072.