Prediction of Credit Risks in Lending Bank Loans

Mohit Lakhani[1], Bhavesh Dhotre[2], Saurabh Giri[3]

1,2,3Student, Dept. of IT Engineering, NMIMS College, Maharashtra, India

---***--

Abstract - *Looking at the current scenario there are huge risks involved for Banks to provide Loans. So as to reduce their capital loss; banks should assess and analyse credibility of the individual before sanctioning loan. In the absence of this process there are many chances that this loan may turn in to bad loan in near future. Due to the advanced technology associated with Data mining, data availability and computing power, most banks are renewing their business models and switching to Machine Learning methodology. Credit risk prediction is key to decision-making and transparency. In this review paper, we have discussed classifiers based on Machine and deep learning models on real data in predicting loan default probability. The most important features from various models are selected and then used in the modelling process to test the stability of binary classifiers by comparing their performance on separate data.*

Key Words: **Machine Learning, Loan Default Prediction, Credit Risk, Data Mining, Credit risk evaluation.**

1. INTRODUCTION

With rapid Economic Development, financial markets are maturing, the phenomenon of individuals as financiers to participate in financial markets has become common. The Lending of loans to individuals (personal loans) can be a profitable business if credit risk management and control is done in a precise manner. Individual loan risk management includes three aspects: risk assessment (i.e. assessment of the applicant's repayment ability and credit repayment to decide whether to grant the loan), repayment tracking, breach of contract, in which risk assessment is essential. At present, relevant information based on the applicant and also the relevant rules need to decide whether to grant loans.

Loans can be categorised in two ways they are: Open-ended loans are loans that you can have a loan of more and more. Credit Cards can be considered the example of open-ended loans. As we have a credit limit that we can buy with both of these two types of loans. At any instance when you purchase something, your available credit decreases. Since it is easy and provides you with money, we tend to use it more and more. Closed-ended loans, this type of loan is not called loans once they are repaid. While we make expenditure on closed ended loans, the balance of the loan became downward. As an option, if we want to lend more money, we'd have to make application for other loan. Types of closed-ended loans are auto loans, mortgage loans, and student loans.

Banks need Machine Learning Algorithms in order to predict precisely the extent to which an individual can ask for loan.

With Machine Learning Algorithms there are Data mining methods that can combined be used in financial sectors like customer segmentation and profitability, high risk loan applicants, predicting payment default, credit risk analysis, ranking of investments, fraudulent transactions, cash management and forecasting operations, most profitable Credit Card Customers and Cross Selling. There are many different types of loans you have to consider when you're looking to borrow money and it's important to know your options. Loan categorization is the process of evaluation loan collections and assigning loans to groups or grade based on the perceived danger and other related loans properties. The process of continual review and classification of loans enables monitoring the quality of the loan portfolios and to act to counter fall in the credit quality of the portfolios.

1.1 Data mining in banking

Due to huge growth in data the banking industry deals with, analysis and transformation of the data into useful knowledge has become a task beyond human ability. Data mining techniques can be adopted in solving business problems by finding patterns, associations and correlations which are hidden in the business information stored in the database. By using data mining techniques to analyses patterns and trends, bank executives can predict, with increased accuracy, how customers will react to adjustments in interest rates, which customers are most interested in new product offers, which customers will be at a higher risk for defaulting on a loan, and how to make customer relationships more profitable. Banks focus towards customer retention and fraud prevention. By analyzing the past data, data mining can help banks to predict credible customers. Thus, banks can prevent fraud and can also plan for launching different special offers to retain those customers who are credible.

2. METHODOLOGY

Based on reviewing several methodologies for prediction of credit risk through Machine Learning, we have mapped out most important and relevant factors that should be considered while predicting credit risk for loans. The variables can be categorized as below:

- **Dependent variable**: We differentiate between capable and incapable customers in by assigning 0 to identify capable customers and 1 to identify incapable customers.

- I**ndependent variables:**

1. **Category of Borrower:** database is divided into two groups, Such as Individual or group.

2. Borrower's relationship with the bank: Borrower's relationship with the bank in years.

3. Borrower's history: customer default history is known by assigning binary numbers to know whether customer is defaulter or not

4. Borrower's gender: databases are divided to female and male.

5. Amount of loan: funds provided by the bank to the customer.

6. Duration of loan: Loan duration in Months.

7. Loan Supervision: Loan supervision and monitor through Lender.

8. Type of loan: database is divided into three groups:

   1.Cash Credit   2. Agriculture Loan   3.SOD.

9. Loan recovery Status: Loan Recovery Status is divided into five:

   1. UC-Unclassified   2. SMA-Special Mention Account
3. SS-Sub Standard   4.DF- Doubtful Loan   5. BL-Bad Loan.

10. Interest rates: It determine amount of bank's profit.

Table -1: Variables Definitions and Measurement

Variable	Definition	Measurement
Dependent Variable		
CUST_S TATUS	Good and Bad Customer	Dummy (1,if Customer is good, 0 If Customer is bad)

Borrower Characteristics		
TYPE_BORR	Type of Borrower	Dummy (1, if Individual Borrower, 0 If Corporate Borrower)
RELAT_BORR	Years of Relationship with bank	Scale (Year)
HIST_BORR	Borrower Default History	Dummy (1, if Borrower has history of Default, 0 If Otherwise)
Loan Characteristics		
LOAN_SIZE	Loan Size	Scale (Lac)
LOAN_DURA	Loan Duration	Scale (Month)
SUPERV_LOAN	Loan Supervision and monitoring by lender	Dummy (1, if bank supervised loan, 0 If Otherwise)
RECOV_STATUS	Loan Recovery Status	Dummy (1, If Loan is Unclassified, 0.5 if loan is SMA, 0 if Loan is Sub-Standard, -0.5 if Loan is Doubtful, -1 If Loan is Bad Loan)

These variables can have a particular value like 0 or 1 depending upon the conditions which are being applied. The values which these variables would attain is mentioned above in Table 1. These variable values are then given as an input to calculate whether the customer is a good customer or a bad customer. For this assessment we can use a Network diagram like ANN (Artificial neural network) model.

It has three primary components: the input data layer, the hidden layer and the output layer.

The input layer receives external inputs that is the data is represented in the input layer.

The hidden layer contains two processes: the weighted summation functions and the transformation function. These functions are used to relate the values from the input data to the output measures.

In **the Output layer**, the output corresponds to the different problem classes. In our example there are two outputs, good or bad Customer.
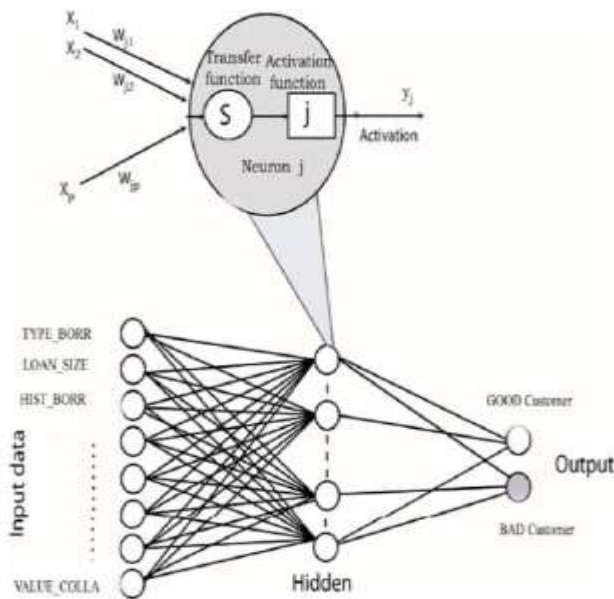

Fig. -1: ANN Model Representation

2.1 Evaluation of Bank Credit Risk Prediction Accuracy based on SVM and Decision Tree Models:

Support Vector Machine:

A support vector machine (SVM) is a special learning technique that analyzes data and isolates patterns applicable to both classification and regression technique. The classification technique is useful for choosing between two or more possible outcomes that depends on continuous predictor variables. Based on data, the SVM algorithm assigns the target data into any one of the given categories. The data is represented as points in space and categories and are represented in both linear and non-linear ways.

The SVM algorithm is used on simple data which focuses on speed and not predicting accurately. Decision Tree Models, are memory-intensive learners and so they are not suitable for large datasets hence we use **Cost-Sensitive Credit Risk Prediction** as it uses Two-Class Support Vector Machine (assigning numeric values and then converting them into binary values) trained on the replicated training data set which provides the best accuracy.

Boosted Decision Tree Model:

A boosted decision tree is a learning method in which the second tree corrects for the errors of the first tree, the third tree corrects for the errors of the first and second trees, and the procedure continues similarly correcting all the previous errors in each Subsequent Tree.

Cost-Sensitive Credit Risk Prediction Experiment:

The main aim of this experiment is to predict the cost-sensitive credit risk of a credit application using binary classification. The classification problem here is cost-sensitive because the cost of wrongly classifying the positive samples is five times the cost of wrongly classifying the negative samples.

3. FUTURE SCOPE

The purpose of our research review is to automate the banking process of selecting the loan applicants which are not risky for their bank or financial institution. In future, we can develop full-fledged early warning systems based on Machine Learning and that will help a bank or any other financial institution to reduce their losses and increase their profits. We are also trying to reduce the time required to do the predictions so the users can get results in real time, which will improve productivity of the users.

4. CONCLUSION

Machine Learning can help banks in predicting the future of loan and its status and depends on that they can act in initial days of loan. Using Machine Learning banks can reduce the number of bad loans and from incurring sever losses. Using above discussed methodology bank can easily identify the required information from huge amount of data sets and helps in successful loan prediction to reduce the number of bad loan problems. Data Mining and Machine Learning techniques are very useful to the banking sector for better targeting and acquiring new customers, most valuable customer retention, automatic credit approval and marketing.

REFERENCES

[1] Addo, Peter Martey, Dominique Guegan, and Bertrand Hassani. "Credit Risk Analysis Using Machine and Deep Learning Models." Risks 6.2 (2018): 38.

[2] Angelini, Eliana, Giacomo di Tollo, and Andrea Roli. 2008. A neural network approach for credit risk evaluation. The Quarterly Review of Economics and Finance 48: 733–55.

[3] Chen, Chongyang, and Robert Kieschnick. "Bank credit and corporate working capital management." Journal of Corporate Finance (2017).

[4]Fan,Jianqing, and RunzeLi. 2001. Variableselectionvianonconcavepenalizedlikelihoodanditsor acleproperties. Journal of the American Statistical Society 96: 1348–60.

[5] Hamid, Aboobyda Jafar, 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.

[6] Hasan, Iftekhar, and Sudipto Sarkar. "Banks' option to lend, interest rate sensitivity, and credit availability." Review of Derivatives Research 5.3 (2002): 213-250.

[7] Matoussi, Hamadi, and Aida Abdelmoula. "Using a neural network-based methodology for credit–risk evaluation of a Tunisian bank." Middle Eastern Finance and Economics 4 (2009): 117-140.

[8] Mitchell, Karlyn, and Douglas K. Pearce. "Lending technologies, lending specialization, and minority access to small-business loans." Small Business Economics 37.3 (2011): 277-304.

[9] Nguyen, Linh. "Credit risk control for loan products in commercial banks. Case: BIDV." (2017).

[10] Sudhamathy, G. "Credit Risk Analysis and Prediction Modelling of Bank Loans Using R."

[11] Sudhakar, M., and C. V. K. Reddy. "Two step credit risk assessment model for retail bank loan applications using Decision Tree data mining technique." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 5.3 (2016): 705-718.

[12] Zhang, Zenglian. "Research of default risk of commercial bank's personal loan based on rough sets and neural network." Intelligent Systems and Applications (ISA), 2011 3rd International Workshop on. IEEE, 2011.