Business, Management and Economics Engineering
ISSN: 2669-2481 / eISSN: 2669-249X
2022 Volume 20 Issue 2

USING MACROECONOMIC INDICATORS TO PREDICT LOAN OUTCOME

Aman Upadhyay, Nishant Singh and Vinayak Shinde
Dept. of Computer Engineering, Shree L.R. Tiwari College of Engineering, Mumbai, India
Amanupadhyay32@gmail.com, nishant.0693@gmail.com, vdshinde@gmail.com

Abstract
The Reserve Bank is always concerned about India's loan default ratios, which are constantly on the higher sides. As much as 50% of certain banks' total loans have reached this level. The banking industry has experienced a corresponding increase in credit analytics, which aids the bank in deciding whether to authorise the loan or not, with the latest advancements in data mining. Our methodology may incorporate macroeconomic aspects into the decision-making process using the current credit analytics and, when combined with machine learning techniques, can greatly anticipate how the loan may perform. The lending and borrowing patterns in the country's economy are greatly influenced by variables such as crude oil prices, the dollar exchange rate, monsoon, fiscal deficit, inflation, industrial production, inflation rate, exports and imports, GDP growth, trade deficit, interest rates on government bonds, compensation of public employees, long-term unemployment rate, gross debt, government spending, and consumer spending. By training the proposed model the goal is to reach a practical level of accuracy using methods like KNN, Random Forest, and Logistic Regression. With this, the goal is to produce outcomes that are accurately defined in terms of objective values that are rational and well-suited to loan application criteria, such as "APPROVED" and "DEFAULT". This method can be used as a tool, to modify the present situation. This novel methodology has a lot of scope for improvements and iterations so that it is finally accepted into the real-world banking sector. This technique has a lot of room for refinement and iteration before it is fully adopted into the real-world banking industry.
Keywords: GDP, KNN Classifier, Loan Default Prediction, Machine Learning, Macro-Economic Parameters, Predicting Loan Outcome

1. INTRODUCTION

Data mining and machine learning play a crucial part in the current trend of dealing with enormous amounts of data to assist different companies achieve maximum profitability. Both techniques concentrate on extracting information from data and using different algorithms to look for patterns in it. Machine learning is becoming a key component of the financial sector.

Offering loans to its clients is one of the bank's primary duties. Banks should give these loans to clients who are able to repay them in order to reduce the risk of non-performing loans, but there is still a risk associated in determining who is deserving of receiving the credit. Predictive modelling is the ideal technique for understanding the various transaction patterns of the client and identifying it either as a fraud transaction or a genuine safe one because there are many frauds occurring in the banking industry [1]. Banks evaluate a customer's credit worthiness based on a number of characteristics, including their job, age, credit score, and pay.

However, macroeconomic circumstances have the potential to alter an individual's credit risk in the future. These issues must also be addressed since they have the potential to upset banks significantly.

In order to determine the relationship between various macro-economic characteristics and the Loan Default Rate of the nation at a specific moment, at first an analysis is conducted and employ a variety of machine learning methods. Secondly, several criteria are combined and ignored that have a little impact on the loan default rate using Principal Component Analysis. A machine learning system is then fed the data set, and it predicts whether the loan application will be "Approved" or "Default."

1.1 Highlights

► Microeconomic factors by themselves have no bearing on a person's loan result. However, the prediction is fairly accurate when macroeconomic factors are taken into account.

► When projecting the loan outcome, macroeconomic factors like GDP, crude oil prices, dollar exchange rate, etc. are taken into consideration.

► The accuracy of predictions using a machine learning model greatly improves the ability to forecast the result of a loan.

► KNN machine learning model gives the greatest loan prediction accuracy of 91.28%.

1.2 Problem Definition

The banking industry is currently especially affected by non-performing loans, especially after COVID-19. Hence the ideas of data mining and proper machine learning algorithms are of utmost relevance. Given this issue, the industry has the chance to use specific data analysis methods and utilise prediction algorithms to solve the non-performing loan issue. [1-2]. To put the entire loan disbursal procedure in perspective, let's say that it begins with a potential client submitting a formal loan application to a bank. The bank's credit staff enters the value into a credit score generator programme, which generates a value between 300 and 900, after collecting other personal information and information about the customers' income and expenses. Considering the current situation, a prediction model that takes into account the current business cycle by taking into account the macroeconomic conditions common in the country might help, if not replace, these judgement decisions. The chances of getting a loan are better the higher the score. While it is true that a person or an entity with a score between 800 and 900 may easily apply for a loan and the bank can be certain that it will be repaid in full, the real issue arises in the range of 600 where the credit team must make a decision on whether to approve the loan or not [3]. For the bank, such an application may go anyway. In these situations, the bank is either anticipating a significant profit or is directly faced with a non-performing loan. Macroeconomic considerations are not being taken into account, which may eventually result in loan defaults and non-performing assets [4-5]. It is undeniable that macroeconomic conditions always prevail and have a significant influence on all banking activity. Choosing the macroeconomic variables to take into account is the biggest problem because so many factors might have an impact on a certain loan occurrence. Crude Oil Prices, Dollar Exchange Rate, Foreign Exchange Reserves, Gross Fiscal Deficit, Index of Industrial Production, Inflation Rate, Monsoon, Export of Goods and Services, Growth in GDP, 10-Year Government Bond, Trade Deficit, Compensation of Employees, Long Term Employment Rate, Import of Goods, General Government Gross Debt, Final Consumption Expenditure by

Government, and Final Household Expenditure are some of the factors are taken into account [6]. This is a comprehensive list of the variables. Here, statistical methods like Principal Component Analysis may be applied to draw a connection between non-performing loans in a data set and the macroeconomic circumstances in place at the time this loan data set was created. Machine learning techniques might be used to forecast the result of loan applications after the association between macroeconomic circumstances and non-performing loan sets is identified. The bank may better assess the risk it bears when evaluating a loan application by using these outcomes, which might either be Approved or Default. This could be accomplished by utilising Machine Learning algorithms like Logistic regression [7], Support vector machines [8], K-nearest neighbours [9], or Decision Trees [9], which allow for the creation of an accurate prediction and the subsequent recommendation of changes to the credit score [10-11].

2. Literature Review

According to Laila Memdani et al. [12], the Hausman Taylor model, which uses the correlation formula, shows that the borrower's financial indicators and country-level macroeconomic indicators are more relevant than bank-specific factors. Using an econometric model, it is possible to determine the relationship between several macroeconomic parameters and the amount of non-performing loans held by commercial banks in Vietnam [13]. The findings show that these elements have a significant impact on non-performing loans. This paper's main drawback is that it only considers five of the 33 commercial banks in Vietnam that have appropriate data for the last ten years. Another significant drawback is the unreliable and erroneous data provided by secondary sources. In order to determine the relationship between economic growth and non-performing loans, Wang et al. [14] took into account statistics on GDP growth, interest rates, and inflation during a 12-year period (1998-2010). By using linear regression on its data set, it is possible to determine the relationship between these macroeconomic factors and non-performing loans.

The macroeconomic relationship between the real economy and the banking sector was reconfirmed by Upadhyay et al. [3]. Both the negative relationship between REER and NPL and the relationship between GDP and NPL were established. Macroeconomic variables in Sweden were compared by Nilsson et al. [6] who discovered that they may account for 75% of variations in the likelihood of loan default.

Logistic regression models were employed by Abid et al. [15] to determine Tunisia's default probability. This model successfully predicted a large proportion of the original group classifications. When comparing several models, Purohit et al. [16] concluded that logistic regression is one of the most effective systems for categorising loan applications. According to Xiao et al. [10], logistic regression produced extremely good results in terms of classification ratio.

A hybrid model based on KNN was created by Li et al. [17], who discovered that it improves the feature space by reducing duplicate features. For credit rating, Keramati et al. [18] used the Optimally Weighted Fuzzy K-Nearest Neighbour (OWFKNN) method. They discovered that, when compared to all other methods, this one performed the best at estimating the likelihood of default. Using the KNN approach, Aksakalli et al. [19] produced findings with a high classification rate.

When Ying [20] evaluated the accuracy of several classifying algorithms, she discovered that the Random Forest approach is superior to the others because it has a strong classification effect when the class is 0. Random Forest technique is preferable, according to Ghatasheh [21], because of its competitive classification accuracy and ease of use. Malekipirbazari et al. [19] .Computations revealed that Random Forest fared better at predicting successful outcomes than the other classifiers. This strategy is quite accurate in predicting the status of excellent borrowers, but at the expense of incorrectly classifying some good borrowers as bad.

Prepayment and default are found to have diverse patterns, with large portions of loans being prepaid and defaulted in online lending.To forecast both events, a multinomial logistic regression model is employed.A number of factors can reliably predict both prepayment and default.

Unsecure consumer loan data is provided out-of-sample validation [24].

Study conducted by Ankara Yıldırım Beyazıt University, Esenboğa Campus, Çubuk, Ankara,Turkey uncovered that, among other organisations, banks are seen as the most significant financial intermediaries, particularly in emerging nations. By balancing the accumulation of small and short-term savers' funds with the large and long-term funding demands of borrowers, banks carry out responsibilities linked to risk size, maturity, and transformation [25].

RESEARCH METHOD

The entire process of our investigation is described in this part. Different variety of approaches are used on the data set in this part.

3.1 Description of the Data set

Our local banks provided the data set for our study, which covered the years 2007 to 2017. To prevent violating consumers' privacy rights, some data set properties have been eliminated. The study used 100,000 loan applications that did not include loans with active status. The loan status, which might be Paid or in Default, is the target variable. Paid is categorised as 1, while Default is categorised as 0. The accuracy rate of projecting loan outcomes is impacted by taking into account each country's macroeconomic factors, yet the data set obtained from these banks did not contain India's macroeconomic statistics. Crude oil, the dollar exchange rate, the foreign exchange reserve, the gross fiscal deficit, the index of industrial production, exports of goods and services, the 10 year government bond, the trade deficit, employee compensation, the general government gross debt, the inflation rate, the monsoon, the growth in the gross domestic product, the long-term unemployment rate, and the final consumption amount were the 17 parameters are chosen to use after analysing and collecting data from various websites regarding the macroeconomic factors affecting the loan outcome. The table below provides an explanation of each parameter.

Table 1: Macroeconomic Factors and their descriptions

Parameter	Description	Unit
Crude Oil Prices	Cost of several barrels of crude oil	$/bbl

Dollar Exchange Rate	Native currency exchange rate	INR
Foreign Exchange Reserve	Assets held by the central bank	Billion USD
Gross Fiscal Deficit	Excess of total expenditure over revenue and capital receipts	% of GDP
Index of Industrial Production	Indicator that measures the growth rate of industries	% Growth
Inflation Rate	Increase in general price of goods and services	% CPI
Monsoon	Amount of rainfall collected annually	mm
Exports of Goods and Services	Value of goods and services exported to other nations	% Growth
GDP Growth	Rate of measure of how fast paced is the economy	% Growth
10 year Government Bond	Debt issued by the Government	%
Trade Deficit	Measure by which cost of country’s imports exceed the exports	Billion USD
Compensation of Employees	Total gross wages paid to the employees by employers in specific time period	Billion USD
Long Term Unemployment Rate	Percentage of people unemployed for more than 12 months	%
Imports of Goods and Services	Value of goods and services imported into the country from other nations	Billion USD
General Government Gross Debt	Total gross debt by the end of the year	% of GDP
Government Final Consumption Expenditure	Aggregate transaction amount for good and services incurred by the government	% of GDP
Household Final Consumption Expenditure	Nation’s transaction amount for goods and services incurred by residents	% of GDP

Pointless characteristics are removed from each customer's approximately 20 related parameters, such as ID, Policy Code, the number of accounts, the purpose of the loan, etc., and took into account crucial characteristics, such as DTI (Debt to Income Ratio), annual income, interest rate, loan amount, etc. The values in our data collection are changed to decimal form and removed the percent symbol for our convenience. In order to create our primary data collection, and finally integrated these two data sets. A subset of the primary data set is shown in Table 2.

Table 2: Example of Main Dataset

Loan Amount	Interest Rate	Annual Income	DTI	Crude Oil	Inflation Rate	Loan
25000	0.1527	2000	27.65	93.85	5.51	0
15000	0.2128	9000	8.72	80.12	9.7	1
21000	0.1296	18000	20	111.31	11.17	1
12500	0.1349	10000	13.45	80.12	9.7	1
18825	0.17	38000	23.18	108.02	6.49	0
17000	0.12	51500	12.26	37.84	6.32	0
21000	0.1296	18000	20	111.31	9.7	1

Overfitting was a problem because just 20% of the loans have failed and over 80% have been fully repaid. Only send 4,000 000 loan applications are used to our training model in order to prevent over fitting. Following the combination of the data set, the next step was to look for any missing values in our data and fill them by averaging the whole column. Our next step was to normalise every value in the data set since it is essential to do so before feeding such a sizable data set to the training model. The missing values were filled in by taking the mean of the whole column. After completing all three processes, data has been split into two sets: 80% for training and 20% for testing.

3.2 Feature Selection

In this part, the principal component analysis (PCA) approach is employed to determine the relationship between the loan default rate and macroeconomic parameters. The primary application of principle component analysis is the orthogonal transformation of a set of potentially correlated characteristics into a set of independent uncorrelated variables. Out of the available macroeconomic variables, the top 17 parameters are chosen, including the price of crude oil, the dollar exchange rate, the foreign exchange reserve, the gross fiscal deficit, the index of industrial production, the inflation rate, the monsoon, the growth of the gross domestic product (GDP), the 10-year government bond, the trade deficit, employee compensation, the long-term unemployment rate, the imports of goods and services, and the final consumption expenditure of the government. The principle component analysis is then used to translate these 17 parameters into two independent features. Additionally, the data set that the local bank has made accessible to us contains details on its clients, including 56 columns and features such loan amount, funded amount, term, interest rate, debt-to-income ratio (DTI), payment, yearly income, issue date, loan status, and description. Out of which, attributes like loan amount and financed amount have a strong correlation of 0.992, indicating a close relationship. To simplify computation, just one characteristic can be included. Some characteristics, such the description, grade, sub-grade, address, etc., are unrelated to the goal of our project and can be ignored as a result.

Table 3: Attributes of selected features

Variable	Attribute	Character
Loan Amount	The amount of loan applied by the borrower	Numerical
Term	The number of payments on the loan	Numerical
Interest Rate	The simple interest charged on the loan amount	Decimal
Annual Income	The annual income of the borrower	Numerical
DTI	A ratio of the borrower's monthly income that goes to paying the monthly debt to the total income	Decimal
Loan Status	Loan Default or Loan Paid successfully	Binary

3.3 Data Modelling

A number of methods are examined that will be used in our research, including SVM, KNN, Logistic Regression, Decision Trees, and Random Forest. In their article, Michelle et al. [22] go through how neural networks analyse and interpret hidden correlations that are utilised to predict the dynamics of variables in financial markets. However, it is not applied since it would have required a protracted training procedure to determine the most effective and ideal architecture. Additionally, SVM beats Logistic Regression when paired with other methods, however it takes more time to train the dataset. Given the size of our dataset, SVM takes a long time to forecast the result [8]. In this study, methods including K-Nearest Neighbours, Random Forest, and Logistic Regression were used. A brief description of these techniques and the logic behind employing them is discussed below:

As an estimator rather than a classifier, logistic regression is more useful. It operates on the assumption of maximum likelihood estimation and offers a forecast of the likelihood that event Y will occur given event X. Since it can handle close calls better, it works well with our model, which asks us to estimate whether a loan would default given a set of current macroeconomic circumstances [13]. Additionally, it offers us a very low mistake rate.

The KNN algorithm is crucial in projects where accuracy is crucial. In the paper, several k values are examined. Despite the lengthy computations required, particularly when working with a large dataset, the technique is anticipated to produce a more precise categorization [17]. KNN can quickly adapt to additional datasets as they are introduced in the future because it is a self-evolving approach.

The study uses a sizable dataset with a high degree of dimensionality. Random Forest algorithms significantly lower the danger of overfitting. Utilizing this algorithm in such circumstances is also quite simple since it can effectively manage the relevant factors and produce reliable results.

3. RESULTS AND ANALYSIS

Only the microeconomic factors are used of one particular individual while analysing the prediction performance for the three different model types (Logistic Regression, Random Forest, and K-Nearest Neighbour). You may find it in Table 4. The Random Forest Algorithm is the least accurate of the bunch, with a 57.91% accuracy rate. The logistic regression's accuracy is 62.865%. K-Nearest Neighbour predicted the best accurately (63.37%) on a sample of 4,00,000 loan applications, with the number of neighbours fixed at 5.

Table 4: Models and their accuracy considering only micro parameters

Method	Accuracy (Only Micro Economic Factors)
Logistic Regression	62.865%
Random Forest	57.91%
KNN	63.37%

Along with the provided microeconomic indicators, the forecast performance is also examined taking into account the national macroeconomic elements. In Table 5, it is displayed. Out of all the algorithms, the Random Forest Algorithm has the lowest accuracy at 86.31%. The accuracy of logistic regression is 89.11%. On a sample of 4,00,000 loan applications, K-Nearest Neighbour predicted the best accurately (91.28%), with the number of neighbours set at 5.

Table 5: Models and their accuracy considering macro-economic parameters as well

Method	Accuracy (Only Micro Economic Factors)
Logistic Regression	89.11%
Random Forest	86.31%
KNN	91.28%

The findings shown in Tables 4 and 5 demonstrate that when macroeconomic factors are taken into account, the accuracy of forecasting a given person's loan outcome increases noticeably. Only 63.37% of the supplied data set's microeconomic parameters can be accurately predicted using the KNN method. However, when macroeconomic indicators are included, the accuracy of predicting the loan outcome rises to 91.28%, or around a 28% improvement.

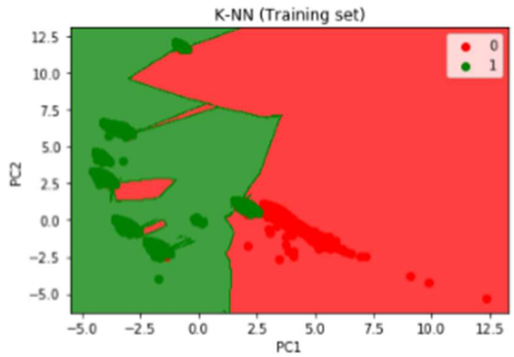

Figure 1: Training Data Set for KNN

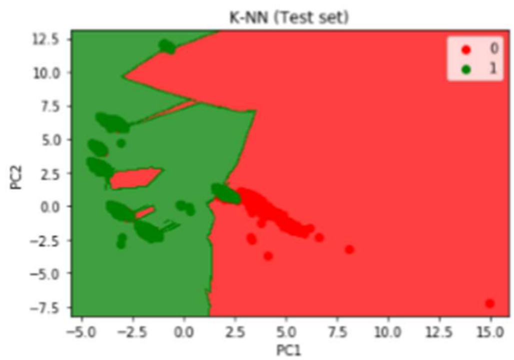

Figure 2: Testing Data Set for KNN

The above figure shows the training and testing data set of KNN algorithm with maximum accuracy of 91.28%. The accuracy of KNN algorithm varies with no. of neighbours (n) set in it. The accuracy with size of clusters ranging from 2 to 500 is calculated. This has been demonstrated in the graph below.

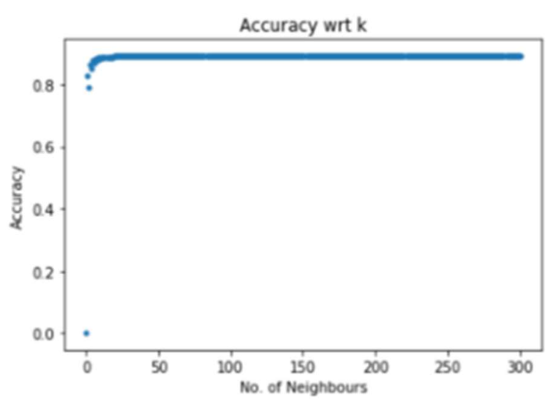

Figure 3: Accuracy vs. No. of Neighbours

The accuracy is the best when k value is 280 after analysing all the values can be concluded here. It demonstrates that the KNN algorithm's accuracy rises steadily up to a point before plateauing. Then, as n increases, it lowers. K=280 is the stationary point when accuracy is highest. The greatest accuracy at this moment is 91.28%, which is higher than that of any other method.

Interpreting the utility of the dataset requires careful consideration of the recall and precision numbers. The recall and precision functions are employed to forecast these values. The table below provides the values for both of them for the KNN model.

Table 6: KNN's Recall and Precision

Classes	Precision	Recall
Class 0	0.94	0.80
Class 1	0.81	0.95

The above KNN model's F1 score is 0.8866. The harmonic mean of the precision and recall values, also known as the F1 score, serves as a general indicator of a model's correctness.

4. CASE STUDY

This section examines particular situations when our model performs and makes more accurate predictions. This, as we've established, is a result of the fact that give the model more macroeconomic factor information. Although the accuracy percentage serves as a proof-of-concept, it is crucial to rationally support the conclusions. We'll talk about two situations here. At first, a situation where a default has occurred is looked upon, and then a situation where it hasn't is talked about. The model that just takes into account the micro-economic circumstances has been shown to be insufficient in both of these situations.

5.1 Case: Default

In this case study, a scenario is chosen in which the loan is deemed to be in default. The micro characteristics, however, are unable to accurately forecast the likelihood of a loan default. In this situation, the bank will lose money on the loan deal. Here, the process starts by examining the micro parameters. Even if the interest rate is on the higher side when compared to all other instances of the year, the loan to income ratio was sound, and the interest % appears to be covered. In the model that does not take into account macro characteristics, the DTI ratio, nevertheless, takes precedence over all other components.

Table 7: Micro Parameters for Case Study 1

Micro Parameter	Value
Annual Income	20,00,000
DTI	20.25%
Interest Rate	20.00%

However, when examining the macroeconomic circumstances of the time, the inflation index was among the highest. The monsoon did not significantly outperform previous years. The GDP also appeared to be at its lowest point. Finally, the price of crude oil, which was at a record high, appears to be by far the biggest role in this. In light of all these considerations, the model correctly concludes that it was not sensible for a business to take out a loan at a 20% interest rate.

Table 8: Macro Parameters for Case Study 1

Parameter	Value
Inflation Index	11%
Monsoon	1111.6 mm
GDP Growth Rate	6.6%
Crude Oil Prices	$108/bbl

5.2 Case: Fully Paid

In this study, an instance is chosen wherein the loan was paid in full but the model which considers only the micro economic parameters was unable to successfully predict. It classified it as default, in which the bank loses an opportunity at making profit. On the contrary, on the same chosen instance, the model which considers the macro-economic conditions can classify the instance correctly.

Firstly, a look at the micro parameters is made. In terms of credit ratings, the outlook looks bleak as the DTI ratio is quite high. The interest rate as well is very high, and the loan is for a longer term. These parameters are summarized below.

Table 9: Micro Parameters for Case Study 2

Parameter	Value
Annual Income	46,00,000
DTI	32.8%
Interest Rate	15.69%

Now, even though the micro parameters are not perfectly suited to grant a loan straightaway, a closer look at macroeconomic conditions helps us classify the loan application correctly. For the particular instance, Consumer spending and government spending was on the higher sides, indicating a more robust economy driven by demand. The 10-year government bond ratings were also better than other years, indicating trust in the market. Index of inflation were at its lowest as well as the crude oil prices hit rock bottom, making consumer spending easy and keeping the market buoyed. The GDP growth rates also touched 8%. So given the market conditions, the bank could afford a risky outlook and thus make higher profits off the loan transaction. Our model takes all this into consideration and gives a correct classification.

Table 10: Macro Parameter for Case Study 2

Parameter	Value
Inflation Index	3%
10 Yr Gov Bond	7.7%
GDP Growth Rate	8.125%
Crude Oil Prices	$37.84/bbl

5. CONCLUSION

In this study, a machine learning model that predicts loan outcomes based on macroeconomic characteristics is provided. The classifier's performance is estimated using real-world data, and the model is built on algorithms and methodologies. It is established that there is a connection between certain macroeconomic indicators and non-performing loans. Hue Sullivan's [7] and Stefan Lessmann's [11] studies have given researchers a clearer grasp of how machine learning algorithms are used in loan prediction. To determine which machine learning method was most appropriate for our research, three distinct approaches are tested. These methods include KNN, Random Forest, and Logistic Regression, with KNN providing the best accuracy (91.28%). In order to further improve the predictions' accuracy, a plan to gather more data based on numerous other parameters is planned. Banks can utilise the proposed model to take into account several macroeconomic parameters in addition to the commonly used microeconomic parameters. Our findings and outcomes in this area demonstrate the validity and applicability of the research done by Klejda Gabeshi [5], Nguyen Huu Quang, and Nguyen Xuan Nhi [13] to stabilise the economy and lower the loan default rate. The most fundamental factors that affect the prediction of loan outcomes are used to create a predictive model. This study has the effect of reducing the risk of non-performing loans in the banking industry.

REFERENCES

[1] Credit Evaluation Model of Loan Proposals for Indian Banks, Seema Purohit, Anjali Kulkarni 2017

[2] A Machine Learning Approach for Predicting Bank Credit Worthiness, Regina Turkson, Edward Baagyeere, Gideon Wenya 2016

[3] Upadhyay and Roy, Analysis of macroeconomic factors causing non- performing loans in India, International Journal of Business and Globalisation, January 2017.

[4] A Comparative Study of Data Mining Methods in Consumer Loans Credit Scoring Management, Wenbing Xiao, Qi Fei, Qian Zhao - 2006

[5] The Impact of Macroeconomic and Bank Specific Factors on Albanian Non-Performing Loans Klejda Gabeshi, European Journal of Sustainable Development Research

[6] Laurin, Martynenko and Nilsson, The Influence of Macroeconomic Factors on the Probability of Default - A Study of the Relationship between Default Probabilities and Macroeconomic Variables, August 2009.

[7] Machine learning for Credit Scoring: Improving Logistic Regression With Non Linear Decision Tree Effects, Hue Sullivan - 2018

[8] Support Vector Machines for Credit Scoring, Bc. Michal Haltuf-2014

[9] Research and Application of Credit Score Based on Decision Tree Model, Jiang Wei Li 2011

[10] Xiao, Zhaoand Fei, A Comparative Study Of Data Mining Methods In Consumer Loans Credit Scoring Management. J Syst Sci Syst Eng, Dec 2006.

[11] Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research, Stefan Lessmann, Hsin-Vonn Seow, Bart Baesens - 2015

[12] Factors Affecting Non-Performing Loans/Assets in the Public Sector Banks of India, Laila Memdani and Shilpam Dubey; April 2018; The Emperical Economic Letters, September 2017

[13] The Relationship between Macroeconomic Factors and the Level of Non-Performing Loans (NPLs) of Commercial Banks in Vietnam, Nguyen Huu Quang , Nguyen Xuan Nhi; March 2017; International Journal of Scientific Engineering Research, Volume 8, Issue 3; ISSN 2229-5518

[14] Muhammad Ishfaq Ahmad, Wang Guohui, Mudassar Hasan, Rizwan Ali , Muhammad Yasir Rafiq , Ramiz ur Rehman; Non-Performing Loans and Economic Growth; October 2016; Scholars Journal of Economics, Business and Management;

[15] Abid, Masmoudi and Zouari-Ghorbel, The Consumer Loans Payment Default Predictive Model: an Application of the Logistic Regression and the Discriminant Analysis in a Tunisian Commercial Bank, Springer Science + Business Media New York, May 2016.

[16] Purohit, Mahadevan and Kulkarni, Credit Evaluation Model of Loan Proposals for Indian Banks, International Journal of Modeling and Optimization, Vol. 2, No. 4, August 2012.

[17] Li,The Hybrid Credit Scoring Model based on KNN Classifier, Sixth International Conference on Fuzzy Systems and Knowledge discovery, 2009.

[18] Keramati, Yousefi and Omidvar, Default probability prediction of credit applicants using a new fuzzy KNN method with optimal weight, January 2014.

[19] Malekipirbazari and Aksakalli, Risk Assessment in Social Lending via Random Forests, February 2015.

[20] Ying, Research on bank credit default prediction based on data mining algorithm, The International Journal of Social Sciences and Humanities Invention 5(06): 4820-4823, 2018.

[21] Ghatasheh, Business Analytics using Random Forest Trees for Credit Risk Prediction: A Comparison Study, November 2014.

[22] An Artificial Neural Network Approach for Credit Risk Management, Vincenzo Pacelli, Michele Azzolini – 2010

[23] The Consumer Loans Payment Default Predictive Model: an Application of the Logistic Regression and the Discriminant Analysis in a Tunisian Commercial Bank, Lobna Abid Afif Masmoudi Sonia Zouari-Ghorbel-2016

[24] Predicting payments and and Default risks of unsecured consumer loans in online lending, Zhiyong Li, Ke Li, Xiao Yao, Quing Wen-2019

[25] Machine learning approach to drivers of bank lending: evidence from an emerging economy Ozgur *et al. Financ Innov (2021)*