

# Performance Evaluation of Traditional Classifiers on Prediction of Credit Recovery



Mohammad Rajib Pradhan, Sima Akter, and Ahmed Al Marouf

**Abstract** In the era of Big Data, machine learning is an emerging technique to analyze the large volume of data and is used to make critical business decisions. It is broadly used in different area such as medical, telecom, social media, banking data analysis and so on to learn the data and perform predictive analysis as well as building recommendation system. With the progression of technology, data availability and computing power, most of the banks and financial institutions are adapting their business model with technological development. Credit risk analysis is a cardinal field for banking and financial institutions, and there are numerous credit risk technique exists to predict the creditworthiness of the customer and loan default probability. In this study, we explore credit defaulter dataset of Bangladeshi bank and conduct several traditional machine learning classifier to predict the delinquent clients who possessed the highest probability of short-term credit recovery. Furthermore, we perform feature engineering to identify the important features for credit recovery prediction. We then apply our final features on different machine learning classifier and compare the predictive accuracy with the other classifier. We observe that random forest classifier gives 90% accuracy in credit recovery prediction. Finally, we propose a noble strategy to identify the potential customer for recovering the credit amount by using supervised machine learning techniques.

**Keywords** Credit recovery · Classification · Confusion matrix · Feature extraction · Machine learning algorithms · Neural networks · Predictive models

---

M. R. Pradhan (✉)

Research & Development, Sonali Intellect Limited, Dhaka, Bangladesh

e-mail: [mrpradhan.rajib@gmail.com](mailto:mrpradhan.rajib@gmail.com)

S. Akter · A. Al Marouf

DIU HCI Research Lab (HCI RL), Department of Computer Science and Engineering, Daffodil International University (DIU), Dhaka, Bangladesh

e-mail: [sima15-5812@diu.edu.bd](mailto:sima15-5812@diu.edu.bd)

A. Al Marouf

e-mail: [marouf.cse@diu.edu.bd](mailto:marouf.cse@diu.edu.bd)

© Springer Nature Singapore Pte Ltd. 2020

T. Sengodan et al. (eds.), *Advances in Electrical and Computer Technologies*,

Lecture Notes in Electrical Engineering 672,

[https://doi.org/10.1007/978-981-15-5558-9\\_48](https://doi.org/10.1007/978-981-15-5558-9_48)

# 1 Introduction

The process of deciding to accept or reject a client’s credit by banks commonly executed via judgmental techniques, credit scoring models and/or machine learning techniques. In this process, banks examine the creditworthiness of their customers, and if the result of this process is at a satisfactory level, they sanction the loan application as a creditworthy applicant. However, due to the different types of the problem such as lack of central integration, judgmental decision error, lack of valuable information, violation of borrower commitments and political influence few of loans are get classified and it’s become a Non-performing Loans (NPL). The average percentage of classified loans is higher in Bangladesh compared to other countries.

The banking system is a vital part of economic development of each country, and it is clearer that the poor banking system is an obstacle for economic development of the country. The classified loan is an acute problem for the banks of Bangladesh. This country suffered high levels of NPL which reached 10.8% in March 2018 [1]. Likewise, others South Asian country is also experiencing the same problem. The position of few South Asian countries is illustrated in Fig. 1. Data of all countries were collected from the World Bank Web site [2] excepting Bangladesh. Data for Bangladesh were collected from Survey Report of Study on Credit Risk arising by Bangladesh Bank [3] and most popular Bangladeshi Newspaper [1] The Daily Star. Now, Bangladesh is on course to become a developing country, and of course, different types of initiative should be made to reduce the amount of non-performing loan to strengthen the financial institutions, which is the backbone of an economy. The government has taken different types of initiative to reduce the rate of the classified loan, but it is still at high rate and gradually increasing the bad debt amount.

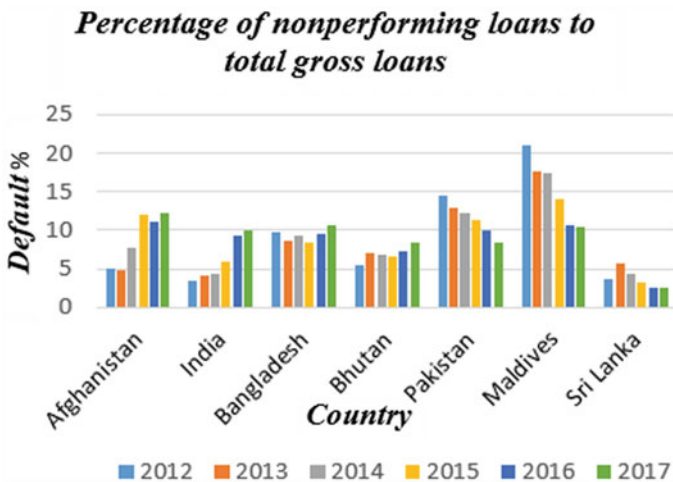


Fig. 1 NPL position of South Asian country

There is plenty of research has been conducted [4–7] to determining the credit-worthiness of a customer, predicting the classified loan, partitioning the credit groups of good or bad payers. However, once the loan has classified, it could be paying back due amount and become a good credit in near future, and there has little research on identifying customers with the most potential to return to normal situation again. This is why we conduct this study to identify the potential customer for recovering the credit amount by using supervised machine learning techniques.

To work with this problem, we have collected loan defaulter dataset from a banking data warehouse (DW) server. Our dataset is consisting of categorical and numerical feature. We have performed data analysis to understand the nature of data and feature extraction and identification process. We used a statistical feature extraction process to identify the most important feature and creating new features. After identifying various feature from loan defaulter dataset, we applied supervised machine learning algorithm to determine the probability of credit recovery for a given bank customer. Finally, we have performed the choice of the parameters for each model and check the role of variables to avoid bias. After model performance evaluation, we have found our satisfactory result and finalize the best performing model.

The rest part of this paper is organized as follows: Sect. 2 is used for related work discussion. In Sect. 3, we have presented the methodology of our work. Data description and feature section process have been discussed in Sect. 4. The experimental results are outlined in Sect. 5, and finally, we conclude this paper in Sect. 6 and indicate some feature works.

## 2 Related Works

Over the decades, a numerous number of studies have been conducted to predict the creditworthiness of bank customers. Machine learning approach used by Pandey et al. [8] to evaluate the credit risk of a customer in their dataset. They survey different machine learning technique for credit risk analysis in German and Australian datasets and reported ELM classifier gives better accuracies. By using weighted-selected attribute bagging method, Li et al. [9] analyze the attributes of customers to assess credit risk. They used two credit datasets for experimental analysis and reported outstanding performance in terms of prediction accuracy and stability and compare classification ability of different models. Li et al. [10] proposed a credit risk assessment algorithm using deep neural networks with clustering and merging technique to assess the default risk of borrowers. They divided the class sample into several subgroups by k-means clustering algorithm, and then, subgroups are merged with minority class sample to produce balanced subgroups, and finally, balanced subgroups are classified using deep neural networks. All of this study shows that there is adequate research has been done to predict the loan defaulter.

The perspective of loan default problems outlined by Chowdhury and Dhar [11] in a study of Commercial Banking sector in Bangladesh. They elucidate contributory factors and loan default problems of both the state-owned commercial banks (SCBs)

and private commercial banks (PCBs) of Bangladesh. Their analysis revealed that state-owned commercial banks are more likely affected by NPL far more adversely than private commercial banks due to some adversary contributory factors. Their emphasized contributory factors are poor credit recovery policy, political interference, lack of managerial efficiency and lack of adapting modern technological changes. Adhikary [12] stated a large amount of NPLs is liable for bank failure as well as economic slowdown. The cause of NPLs attributed to the lack of supervision, effective monitoring and credit recovery strategies.

Banik and Das [13] conducted simple and multiple regression analysis to find the impact of a classified loan on bad debts, and as a result of analysis, they identified classified loans have a significant impact on bad debts in both state-owned commercial banks and first-generation private commercial banks. They also emphasize to develop specific tools and technique to distinguish the willful defaulters from the genuine ones.

Now, it is clearer that besides the good credit risk assessment approach, a good credit recovery approach can help the financial institution to recover bad credit amount from their customers to increase profit and reduce loss. But adequate number of research is not present on classified loans though it is a crucial issue at present. We employed different machine learning technique in this paper to predict the good payer for earlier credit recovery so that amount of NPL could be reduced.

### **3 Methodology**

We have used a supervised machine learning approach to build our credit recovery prediction system, and in order to develop it, we have collected the raw data from the bank DW. Then, we have performed data analysis, feature extraction, data preprocessing, feature scaling, build a model using machine learning classification algorithm and finally generate the performance report. The overall process is outlined in the following sub-sections.

#### ***3.1 Data Collection***

The best predictive results require relevant input data for the business specific modeling that is why we collected data by selecting only relevant inputs and using loan recovery domain knowledge. During the data collection, we have considered few features as categorical and few as numerical. It has information about customers and bank financial data. The data used are collected from Oracle database using Structured Query language (SQL) and Procedural Language for SQL (PL-SQL). Various ETL jobs have been developed to aggregate data into single table using SQL and PL-SQL language. Finally, we export table data into comma-separated values (CSV) file from Oracle database table.

### 3.2 *Data Analysis and Transformation*

Data analysis and transformation is a vital part of machine learning process. It is highly interactive and iterative process [14]. Typically, this process includes data visualization, analyzing missing values, resolving inconsistencies, identifying or removing outliers, data cleaning and transformation, feature engineering and construction. Proper data preparation is required for data visualization.

For building perception relationships among the features and ensuring machine learning models work optimally. The first step in exploratory data analysis is to read the data and then explore the variables. It is important to get a sense of how many variables and cases there are, the data types of the variables and range of values they take on. Initially, from the source data, we have 4600 observations and 47 features.

### 3.3 *Feature Extraction*

Feature extraction is the process of creating new features out of an existing dataset to help a machine learning model to learn a prediction problem [15]. In most cases, machine learning is not done with the raw features. Features are transformed or combined to form new features in forms which are more predictive. This process is known as feature engineering. In many cases, good feature engineering is more important than the details of the machine learning model used. It is often the case that good features can make even poor machine learning models work well, whereas, given poor features even the best machine learning model will produce poor results. We have explored new feature by categorical variable transformation, variable aggregation and statistical feature extraction process.

### 3.4 *Preparing Input and Output*

To build the predictive model, we grouped all features into a matrix  $X$  denoted  $X \in \mathbf{R}^{n \times m}$ , where  $n$  refers to the number of rows, and  $m$  refers to the number of columns, respectively, and a vector  $y$  denoted  $y \in \mathbf{R}^{n \times 1}$ . Matrix  $X$  represents our input data where  $m= 4600$  and  $n= 31$  and vector  $y$  represent output data which contains all outputs, either 1 for good payer or 0 for not a good payer. In order to build classifier model, the dataset is spited into two partitions or sets—training and testing. The training datasets were used to build the classifier model, and the testing datasets were used to evaluate the model performance to make sure that the model performance is well for never-before-seen data.

Before feeding the data to the classifiers, we have performed feature scaling for our dataset. Feature scaling is a crucial step in data preprocessing pipeline of machine learning. The main purpose of feature scaling is bringing different features

onto the same scale [16]. It is required because the majority of machine learning and optimization algorithms behave much better if features are on the same scale. There are two common approaches to bring different features onto the same scale: normalization and standardization. We used standardization feature scaling technique for our datasets. Using standardization, we center the feature columns at mean 0 with standard deviation 1 so that the feature columns take the form of a normal distribution.

The procedure of standardization can be expressed by the following equation.

$$x_{\text{std}}^{(i)} = \frac{x(i) - \mu_x}{\sigma_x}$$

Here,  $\mu_x$  is the sample mean of a particular feature column and  $\sigma_x$  the corresponding standard deviation, respectively.

### 3.5 Modeling

The learning process of machine learning algorithm is partitioned into supervised and unsupervised learning. Supervised learning is used to predict output from a given input, and unsupervised learning is used to segmenting or clustering entities in different groups [17]. Our dataset consists of both input and output feature for which we used the supervised method to make an accurate prediction for never-before-seen data. To generate the best output for credit recovery prediction, we have studied the efficiency and scalability of each classification model by varying several parameters. The logistic regression, Naive Bayes, KNN classification, decision tree, random forests, support vector machine, multilayer perceptron, neural networks, linear discriminant analysis and two boosting algorithms AdaBoost and XGboost have used in this study.

Hyper-parameters are parameters which determine the characteristics of a model [18]. The ultimate objective of the learning algorithm is to find a function that minimizes the loss of the classification model. The learning algorithm produces a function through the optimization of a training criterion with respect to a set of parameters also known as hyper-parameter. For example, with logistic regression, one has to select a regularization penalty  $C$  to reduce the complexity of a model by penalizing large individual weights. Therefore, it is important to find the best optimal hyper-parameter values. Though default value of hyper-parameter is a good start to build a model, they may not produce the optimal model. Grid search is the most widely used strategy for hyper-parameter optimization [19]. In this study, we used Grid-SearchCV() implemented in scikit-learns to produce the best model. It requires set of values for each hyper-parameter that should be tuned and evaluates a model trained on each element of the Cartesian product of the sets.

**Table 1** Confusion matrix of binary classification

Actual class	Predicted class	
	Positive class	Negative class
Positive class	TP	FP
Negative class	FN	TN

### 3.6 Evaluation

In this research study, we employed several performance criteria to evaluate the model performance and assist in model selection process. The most commonly accepted evaluation measures are accuracy, precision, recall,  $F$ -score and score of ROC area (AUC). Many of them come from a confusion matrix which is a specific matrix that shows the relationship between true class and predicted class. Table 1 presents the confusion matrix.

**Accuracy (ACC)** One of the most common metrics is accuracy, which gives the ratio of correctly classified samples to misclassified samples. Accuracy does not take class distribution into account, which makes it poor measure for evaluating performance on imbalanced data.

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

**Precision (PR)** It indicates how many values, out of all the predicted positive values, are actually positive. It is formulated as following.

$$\text{Precision} = TP/(TP + FP) \quad (2)$$

**Recall (RE)** It indicates how many positive values, out of all the positive values, have been correctly predicted. The formula to calculate the recall value is in Eq. (3).

$$\text{Recall} = TP/(TP + FN) \quad (3)$$

**$F1$ -Score**  $F1$  score is the harmonic mean of precision and recall. It lies between 0 and 1. Higher the value, better the model. Equation (4) shows the formula for  $F1$ -score.

$$F1\text{-score} = (2 * \text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall}) \quad (4)$$

**AUC** The area under the curve (AUC) refers to the area under the receiver operating characteristics (ROC) curve. The overall performance of classifier model is measured by the AUC. The higher the AUC the lower the increase in false positive rate required to achieve a required true positive rate. For an ideal classifier, the AUC is 1.0. A true positive rate is achieved with a 0 false positive rate. This behavior means that AUC

is useful for comparing classifiers. The classifier with higher AUC is generally the better one.

## 4 Data Description and Feature Selection

The dataset used in this research was obtained by the extraction of information from a Bangladeshi Bank DW, and due to data confidentiality, the whole dataset cannot be published. As we have collected data from DW, the data preparation activity was reduced to aggregating data into single database table containing 47 features with combination of numerical and categorical variables. However, not all the 47 features were adequate enough to obtain the target outcome. We therefore extract the new feature from existing features, removed irrelevant and redundant features, and finally, we have identified ten categorical and 21 numerical features which have shown in Tables 2 and 3 accordingly. The credit dataset has 4600 instances in where the majority of samples (3698) are positive class and the minority of samples (902) are negative class. It is clear that the ratio of position and negative classes is unbalanced.

Before doing any more feature engineering, it is important to establish a baseline performance measure. In this study, we used a gradient boosting framework implemented in Light-GBM library to assess the performance of new feature. Our feature engineering approach is outlined below:

- One-hot encode categorical variables.
- Make a baseline model to establish a benchmark.
- Build a new feature by manipulating columns of the base (main) data frame.
- Assess performance of new feature set in a machine learning model.

**Table 2** Categorical feature description

SL	Features	Description
1	V01	EMI facility (1 = yes, 0 = no)
2	V02	Individual loan (1 = yes, 0 = no)
3	V03	Personal loan (1 = yes, 0 = no)
3	V03	Agriculture loan (1 = yes, 0 = no)
4	V04	Simple or compound interest (1 = yes, 0 = no)
5	V05	Staff loan (1 = yes, 0 = no)
6	V06	Customer sex (1 = male, 0 = female)
7	V07	Marital status (married, single, others)
8	V08	Secured or unsecured loan (1 = secured, 0 = unsecured)
9	V09	Repayment frequency (monthly, others)
10	V10	Number of bad loan (1 = one, 0 = more than one)



**Table 3** Confusion matrix of binary classification

SL	Features	Description
1	V11	Annual income slave
2	V12	Number of days of regular payment
3	V13	Days of customer relationship with bank
4	V14	Initial contract amount
5	V15	Customer age in year
6	V16	Interest rate
7	V17	Amount of the outstanding balance
8	V18	Total recovery amount
9	V19	Total interest received from customer
10	V20	Monthly installment amount (EMI amount)
11	V21	Total number of installment
12	V22	Number of installment paid
13	V23	Number of loan account
14	V24	Remaining day of contract maturity
15	V25	Total amount of last six month bill
16	V26	Total amount of last six month payment
17	V27	Mean value of last six month bill
18	V28	Mean value of last six month payment
19	V29	Ratio of contract and recover amount
20	V30	Ratio of past six month bill and recover amount
21	V31	Ratio of contract value and days of regular payer

The last step we aimed to reduce the dimensions of our dataset by removing irrelevant and redundant features did not have a significant impact on prediction power.

## 5 Experimental Results

The result that we have achieved after implementing different classification algorithm is promising enough. Table 4 describes our achieved result. In the experiments reported in this paper, *Keras*, which is a high-level neural networks API, is used to run the neural network, *Python v3.6* and *scikit learn* [20] were used for the implementation. The accuracy, precision, recall, *F*-score and AUC values are extremely good for most of the algorithms. Thus, these algorithms are very suitable for analyzing bank credit data and build a classifier model. Our experimental result shows that the random forests model is outperformed for credit delinquent dataset in term of

**Table 4** Experimental results

Classifier	ACC	PR	RE	F1-score	AUC
Logistic regression	0.87	0.89	0.34	0.49	0.66
Naive Bayes	0.24	0.19	0.94	0.31	0.51
KNN classification	0.87	0.80	0.42	0.55	0.70
Decision tree	0.85	0.60	0.50	0.55	0.71
Random forests	<b>0.90</b>	0.85	0.56	0.67	0.77
Support vector machine	0.87	0.85	0.35	0.49	0.67
Multilayer perceptron	0.87	0.94	0.33	0.49	0.66
AdaBoost	0.87	0.79	0.38	0.51	0.68
XGboost	<b>0.89</b>	0.88	0.46	0.61	0.72
Neural networks	0.87	0.75	0.41	0.53	0.69
Linear discriminant analysis	0.87	0.79	0.37	0.50	0.67

accuracy (**90%**), precision, recall, *F*-score and AUC values and AUC score. Naive Bayes model is worst performing model than others models.

The XGboost model achieves 89% accuracy and AUC value 0.72 which is closer to best performing model. It indicates that the random forest model can better distinguish the delinquent client for short-term credit recovery instances from credit defaulter dataset.

## 6 Conclusion

NPL refers to those financial assets from which bank no longer receives neither interest nor installment amount and it is a big obstacle for economic development of a country. The elevated rate of NPL amount needs to reduce for sustainable economic development, and in the era of artificial intelligent, machine learning technique can be used to reduce the ratio of NPL amount. Therefore, in this research, we have applied different machine learning approach in credit defaulter dataset to predict the delinquent clients who possessed the highest probability of short-term credit recovery. We have evaluated 11 different classification algorithms to determine the algorithm which is best fit for credit recovery dataset. Basically, we have tried to establish a solid comparison between different classification algorithms and improve the accuracy of the prediction by increasing accuracy and minimizing errors, bias and variance. As a result, we got random forests as the best performing model in terms of accuracy and AUC, and Naive Bayes is the worst. We also identified important feature that needed to build an optimal predictive model in order to formulate banks credit recovery automated system. The classifier identified by supervised machine learning techniques will be very much supportive for the financial institution to

identify delinquent client for short-term credit recovery. The proposed model would be helpful with existing credit risk assessment approach to reduce the ratio of NPL.

## References

1. Mujeri MK (2018) Rising non-performing loans [Online]. Available <https://www.thedailystar.net/news/opinion/perspective/rising-non-performing-loans-1619257>. Accessed 25 Nov 2018
2. World Bank, International Monetary Fund (2018) Global financial stability report [Online]. Available <https://data.worldbank.org/indicator/FB.AST.NPER.ZS>. Accessed 10 Dec 2018
3. Sarker MAA (2017) Study on credit risk arising in the banks from loans sanctioned against inadequate collateral. Research Department, Bangladesh Bank
4. Li FC, Wang PK, Wang GE (2009) Comparison of primitive classifier with ELM for credit scoring. In: Proceeding of IEEE IEEM, pp 685–688
5. Wang Y, Wang S, Lai KK (2005) A new fuzzy SVM to evaluate credit risk. *IEEE Trans Fuzzy Syst* 13:820–831
6. Zhou H, Lan Y, Soh YC, Huang GB (2012) Credit risk evaluation using extreme learning machine. In: Proceeding of IEEE international conferences on system, man and cybernetics, pp 1064–1069
7. Zhu B, Yang W, Wang H, Yuan Y (2018) A hybrid deep learning model for consumer credit scoring. In: Proceeding of IEEE international conference on artificial intelligence and big data, pp 205–208
8. Pandey TN, Mohapatra SK, Jagadev AK, Dehuri S (2017) Credit risk analysis using machine learning classifiers. In: Proceeding of IEEE international conference on energy, communication, data analytics and soft computing, pp 1850–1854
9. Li J, Wei H, Hao W (2013) Weight-selected attribute bagging for credit scoring. *Math Probl Eng* 2013
10. Li Y, Li X, Wang X, Shen F, Gong Z (2017) Credit risk assessment algorithm using deep learning neural networks with clustering and merging. In: Proceeding of 13th international conference on computational intelligence and security (CIS), pp 173–176
11. Chowdhury R, Dhar BK (2012) The perspective of loan default problems of the commercial banking sector of Bangladesh. *Univ Sci Technol Annu (USTA)* 18(1):71–87
12. Adhikary BK (2015) Nonperforming loans in the banking sector of Bangladesh: realities and challenges. *J BIBM* 75–95
13. Banik BP, Das PC (2015) Classified loans and recovery performance: a comparative study between SOCBs and PCBs in Bangladesh. *J Cost Manag* 43:20–26
14. McKinney W (2013) *Python for data analysis*. O'Reilly
15. Heaton J (2017) An empirical analysis of feature engineering for predictive modeling. arXiv preprint [arXiv:1701.07852v1](https://arxiv.org/abs/1701.07852v1)
16. Raschka S (2015) *Python machine learning*. Packt Publishing
17. Guido S, Muller AC (2016) *Introduction to machine learning with python*. O'Reilly Media, Inc
18. van Rijn JN, Hutter F (2018) Hyperparameter importance across datasets. arXiv preprint [arXiv:1710.04725v2](https://arxiv.org/abs/1710.04725v2)
19. Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. *J Mach Learn Res* 13:281–305
20. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830