Enhancing Credit Risk Analysis of SME Loans by Using Data Mining Techniques

Thamali D. Chandrasiri and Saminda C. Premaratne

Abstract Analyzing credit risk is important in banking systems to ensure that debtors pay the loans regularly, on schedule. The inability to manage the credit risk may lead to severe losses in financial institutes. Every financial institute must predict and manage credit risk to avoid financial crisis. Hence, finding an effective method for credit risk analysis is imperative. Among various types of loans, Small and Medium-sized Enterprise (SME) loans dominate since SMEs are considered the backbone of any economy. Despite, the amount of non-performing SME loans has increased at a higher pace throughout the last few quarters in Sri Lanka. Hence, this study analyzes the credit risk of SME loans by using a data set received from a financial institute in Sri Lanka by using data mining techniques. Data mining is used widely for financial analysis since it facilitates knowledge extraction from large data sets and making effective decisions. A selected set of classification algorithms were used to analyze the data and extract knowledge. In the first phase of the analysis, it predicts whether a particular borrower will pay a loan fully or not. Further, if a particular loan is identified as a non-performing loan, then the model predicts the percentage of the non-payments. In this way, the study helps financial institutes to predict the credit risk of potential SME borrowers and avoid inefficiencies in the lending process. by identifying the credit risk of debtors in different aspects.

***Index Terms*— Credit Risk, Data Mining, Non-performing loans, SME sector**

I. Introduction

CREDIT risk is the most important risk that every financial institution should endure. Analyzing the credit risk of loans is a crucial operation in financial institutes since it ensures their performance and survival. Banks must confirm the lender's ability to pay a loan before granting a loan [1]. The volume of loan facilities has increased daily, making more challenges to banks and financial institutes. The excessive amount of credit and poor credit risk management may lead to a financial crisis. Hence, effective credit management is imperative for the stability of the financial system of a country [2]. Finding an effective method for credit risk evaluation is critical for financial institutions when they are deciding credit facilities and reviewing the facilities granted. Credit risk analysis helps banks to improve their cash flow and make managerial decisions effectively by reducing the risk. Further, the credit rating agencies may need this kind of credit risk analysis mechanism to rate the creditworthiness of lenders of debt agreements [3].

The need for planning, financing and credit management is mandatory for banks with growing loans [4]. Among different types of loans, small and medium entrepreneur (SME) loans are imperative. SMEs play a critical role in economic growth of any country [5]. In Sri Lanka, SMEs are also considered the backbone of the economy. SME contributes 52% of the Gross Domestic Production (GDP) and also it offers 45% of employment [6]. Therefore, stimulating SMEs are necessary for the economic growth in the country. Licensed and specialized banks provide credit support to eligible SMEs to inspire them [6]. The incapability of bankers to foresee the credit risk of potential SME borrowers may adversely affect the financial system and economic activities. [7] have found that credit risk is an important determinant of assessing the profitability of the banking sector in Sri Lanka. Moreover, banks perceive the SME sector is risky and costly than the large enterprises [8].

Many studies have been conducted on building reliable credit scoring models to analyze the credit risk of new loan applicants. These attempts seem to be not significant due to major changes of the weights around the optimum model having a small impact on its implementation. Some misclassification patterns also could arise, due to the impact of economic conditions and highly nonlinear characteristics of loan data, even the scoring models are accurate [9]. Though financial institutes use different methods to assess the credit risk, the number of non-performing loans have increased throughout the past few years in the Sri Lankan context. Specifically, the number of NPLs has been drastically increased in SME loans from the fourth quarter of 2020 to the second quarter of 2021 [10].

The Department of Supervision of Non-Bank Financial Institutions under the Central Bank of Sri Lanka has also recognized that existing directions on credit risk

G.D.T.D. Chandrasiri, was with University of Sri Jayewardenepura, Nugegoda, Colombo, Si Lanka. She is now with the School of Information Systems, Faculty of Science, Queensland University of Technology, Australia (e-mail: thamali.don@hdr.qut.edu.au).

S.C. Premaratne is with the Department of Information Technology, Faculty of Information Technology, University of Moratuwa, Katubedda, Moratuwa, Sri Lanka. (e-mail: samindap@uom.lk).

ISSN: 2961-5992

management are outdated and one of the main reasons for non-performing loans in the leasing and finance companies. They have mentioned that the credit risk analysis methods should be updated regularly in order to cope up with local and international developments, standards and the best practices arising at the Basel Committee on banking supervision [11]. Further, previous studies have recommended that implementing effective tools and techniques in credit risk management will reduce the credit risk [12]. Besides, there are limited research in the Sri Lankan context to predict the credit risk by using data mining techniques. Hence, this study tries to incorporate appropriate factors in assessing the credit risk and analyze the credit risk using data mining techniques. Further, the non-performing loans will be analyzed based on different aspects which will be useful for the financial sector to identify factors affecting the non-payments and to predict the credit risk accordingly.

II. Literature review

Loan categorization is a process of evaluating loans, allocating loans to groups and assigning grades based on the perceived risk and other related properties [13]. Among various types of loans such as commercial loans, SME loans, personal loans, educational and industrial loans, SME loans are imperative since SME financing critically impacts on economic growth and it creates employment opportunities. Evidence in SME financing shows that enterprises who are unable to produce adequate operating cash flow are more vulnerable to bankruptcy [8]. SMEs in developing countries are facing many challenges such as higher competition, administrative inefficiencies, economic conditions, the long-lasting threat from globalization, and financial crisis have initiated distress [14]. On the other hand, small businesses are the primary sources of the stability and development of financial systems in most nations, fueling the engine of economic advancement and expansion [15]. SMEs are playing a key role in encouraging economic development and industrial production all over the world. Specially, SMEs provide the foundation for sustainable economic growth and increasing income in less developed and developing countries [16].

SMEs are defined based on the annual turnover and the number of employees in Sri Lanka. The Sri Lankan policy framework defines SME as an enterprise which have an annual turnover below 750 million and which have less than 300 employees. It consists of micro, small and medium enterprises. The contribution of SMEs to Sri Lankan economy is more than 75% of total enterprises and contribution to the GDP is 52%. It provides 45% of employment also [6]. Hence, analyzing the credit risk of SME loans is crucial in any economy, to encourage them and grow the economy [16]. According to [10] bankers' willingness to lend has been increased in 2021, compared to fourth quarter in 2 020, and on the other hand, all the loan categories (corporate, retail, State Owned Enterprises (SOE), SME) has been reported a higher readiness to lend. Further, it emphasizes that the Covid-19 pandemic in the first quarter of 2020 stressed more increase in willingness to lend. With this increasing amount of lending, the number of Non-Performing Loans (NPL) were increased at a higher phase in 2021 [10]. Further, it is expected more growth of NPL, with the expiration of moratorium facilities during uncertainty in income levels. [11] has stressed the need to change credit risk directions regularly to keep pace with local and international bodies, standards, practices proposed by the Basel Committee in banking supervision. This highlights the necessity for proper management of the credit risk of SME loans in Sri Lankan economy evidently.

It is important to identify attributes that impact on the credit risk [17]. [18] have introduced 6C's analysis for credit evaluation. Those are character, capacity, collateral capital, condition of economy and constraint. Character refers to the lender's credit rating, which evaluates the ability to pay on the agreement. Capital is the amount of funds available at the lender. Capacity is estimated by considering the current income and expected income during the loan period, and it is the ability of lenders to pay the installments. Collateral is a property of a lender which should be considered as a warranty of the loan. The condition of economy represents the social, political, cultural and economic conditions that affect the sustainability of businesses. Constraint refers to all the barriers and limitations that affect a business, such as scarcity of resources and regulations.

[17] have emphasized different kinds of credit scoring methods such as application credit scoring, collection scoring, behavioral scoring and fraud detection. The manual credit scoring methods consist of many troubles in the industry. Moreover, the process of credit scoring is not standardized in many banks. Such non-standard models are associated with serious problems of analyzing expensive data repeatedly. It is an obstacle for building an optimal model for credit risk analysis [17]. [4] have investigated factors in measuring the credit of bank customers such as current capital turnover, loan interest rate, lender's annual income, customer capital, loan amount and the history of corporate with the bank. According to the results, the average turnover of the account has the highest weightage among the factors.

Though many financial institutes use credit scoring methods, those are not standardized. There are critical issues with these non-standardized models since they analyze repetitive and very expensive data. Indeed, this issue inhibits the development of optimal models in credit scoring. Furthermore, most credit risk models are static, therefore incapable of functioning effectively in economic crises. While these traditional static models have performed quite well in stable situations, it fails to work so with political and economic fluctuations [19]. Thus, usage of data mining methods to assess credit risk is dominant in the field. Data mining is an emerging technology that provides significant advantages in making correct decisions [20]. It contains set of algorithms and methods to extract

ISSN: 2961-5992

knowledge and information from large databases. According to [21] data mining is "the process of discovering meaningful new correlations, patterns and trends by examining through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques". It can be supervised or unsupervised based on the data that are going to be analyzed. The rapid development of internet and telecommunication technologies marks the arrival of big data, as structures, semi-structured and unstructured data generated from different sources. Many countries have identified the importance of big data as a global opportunity [22].

Moreover, different data mining techniques have been stated in the literature for credit risk assessments including K Nearest Neighbors (KNN) [23], [24], Support Vector Machines (SVM) [2] [25], Neural networks [2] , and Logistic Regression analysis [26]. [5] has analyzed the credit risk of SME loans in China by using the Random Forest algorithm and they have ensured the accuracy of the results. Further they have mentioned that analyzing the credit risk of SMEs are very important in this pandemic situation since governments have paid their attention to simulate SMEs. [27] has emphasized that artificial neural networks are more suitable with credit scoring methods as a data mining technique. The author has suggested emotional neural network as a more successful method for pattern recognition when compared to neural networks. Regression models also have been used successfully in predicting credit risk, such as linear regression and logistic regression [28]. Some authors have introduced hybrid or combined data mining techniques by integrating two or more techniques together. [9] have introduced such a hybrid method by combining clustering and classification algorithms. They have used K-Means cluster as a clustering algorithm and support vector machines as a classification algorithm in predicting credit risk of local banks in China. [15] have introduced a data mining approach to predict imbalanced set of small businesses' loan data by combing weighted SMOTE ensemble algorithms and bagging algorithms. [20] have also proposed a hybrid approach by combining genetic algorithms with support vector machine classifiers.

III. METHODOLOGY

In this study, the main process is to identify the most accurate classification algorithm which can be used in analyzing the credit risk of SMEs. A thorough literature review was done to identify existing algorithms that can be used in the process. Within this main process, several sub processes are included. Those are data selection, data preprocessing, data transformation and evaluation. The process will be described further in the next few topics.

The sample data related to SME loans were collected from one of the leading financial institutes in Sri Lanka. It contains nearly 3000 records of SME loan data pertaining to the years 2017 – 2020. Since the Covid 19 pandemic has adversely affected in financial sector, three years were selected (before the pandemic) for a better prediction of credit risk. The data is in the format of MS Excel, thereby the manual preprocessing activities were easy to perform. Further 15 SME types are available in the data set such as plant nursery, weaving, food manufacturing, cane industry, rubber products, batik industry, food manufacturing, garage, craft designing, rubber products, aquarium, wood carving, apparel, mills, diary processing and salon. The factors identified from the data set are gender, age, marital status, net monthly income, business type, loan amount, interest, loan term, monthly installment, status of the payment (Whether the loan will get arrears or not), payment percentage (If the loan is not fully paid, the extent to which the payments have been occurred. Very Low – Pay between 0% - 25%, Low – Pay between 25% - 50%, Medium – Pay between 50% - 75%, High – Pay between 75% - 100%).

The output of this study was a designed simulation which can be used to identify non- performing loans in two aspects. First, it predicts whether a particular SME loan will be fully paid or not. It is not fully paid, then it predicts the percentage of the arrears amount. The classification models were created by using selected classification algorithms and model with the highest accuracy was selected to predict the non-performing loans in this selected data set.

IV. DESIGN AND ANALYSIS

This study follows the scientific research methods to achieve its aim of efficiently predicting the credit risk of SME loans. It adopts the quantitative approach since it relies on numeric data. The quantitative approach requires a set of hard data which are then manipulated and analyzed by using statistical methods to check whether the hypothesis is proved or not. This study adopts the data collection methods such as observations and documentations to identify the issues related to credit risk management of financial institutions. Then, a background study was carried out to identify solutions introduced to mitigate the associate risk with credit. Due to the inefficiencies of those solutions and due to the higher growth of non-performing loans in SME sector, the data mining approach was selected to analyze the credit risk of SME loan data. Previous studies have also used data mining techniques in different contexts successfully.

The main objective of this study is to analyze the credit risk of SME loan data by using data mining techniques. RapidMiner Studio was used for the analysis as a data mining tool. The high-level research design of the study is depicted in Figure 4.1. A sample set of training data was inserted as the input and several classification algorithms were applied on that. The model was developed and evaluated by looking at the performance of different machine learning algorithms. Then, a test data set was inserted to the model and it predicts the credit risk of the given data set.

ISSN: 2961-5992

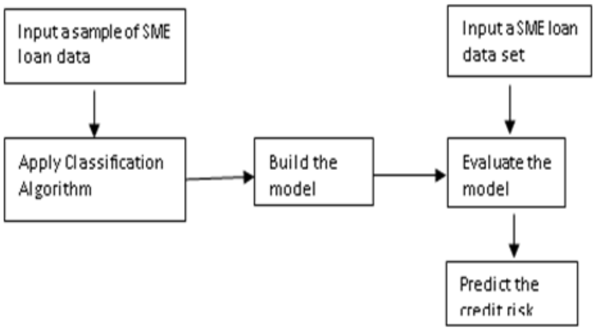

Figure 01: The High Level Design of the Proposed Model

V. IMPLEMENTATION

A. Data Preprocessing:

In this study, a few missing values were identified under the status, gender and business type attributes. By using RapidMiner, first, those missing values were replaced with the average values. Then, the duplicate values were removed as it generates an overfit model. Discretization was done only by selecting the percentage attribute to arrange the percentage values into ranges. Here, discretization by user specification method was used with 4 classes as high (will pay above 75%), medium (will pay between 50%-75%), low (will pay between 25%-50%) and very low (will pay below 25%). Then normalization was done for the interest and loan term attributes to rescale them. Z-transformation or the statistical transformation was used here in order to verify all the attributes have a similar scale for a reasonable comparison between them. The outliers were also detected with the Euclidian distance function with 10 number of neighbors. Then by using a filter example operator, the outliers were filtered out and only the non-outlier examples were selected.

Then the attribute selection was done. The borrower's birth date, the total receivable, and the total amount paid are not directly relevant to the study, hence not selected for further proceedings. Only the relevant attributes were selected including age, gender, interest rate, loan amount, loan term, monthly installment, net monthly income, marital status, payment percentage and the business type. The class label was set to whether arrears or not attribute by using the set role operator and cross validation was used to split the test and training data set. Afterwards different machine learning algorithms were applied to identify the most accurate predictions.

B. Data Mining Algorithms:

The classification algorithms were selected for the analysis since the class label is available in the data set. Among different classification algorithms, five algorithms were selected by referring to the literature. Many researchers have used decision tree, neural networks, SVM, naïve bayes, KNN, linear regression, Rule Induction and Random Forest algorithms in predicting the credit risk. Hence those algorithms were applied and tested for their accuracy in this study with special reference to SME loans which is dearth in literature. To evaluate each algorithm the performance operator was used with different parameters such as the accuracy, classification error, correlation, squared correlation and the Kappa statistic.

C. Parameter Tuning:

Different parameters can be identified under each algorithm. These parameters can be changed to increase each algorithm's performance and reduce the error rate. In decision tree algorithm there are parameters like criterion, maximal depth, confidence, minimal gain, leaf size etc. The criterion and maximal depth were changed to increase the performance. Different criterion can be selected such as gain ratio, information gain, Gini index, accuracy and least square. The Gini index shows the highest accuracy here. The depth of a tree differs depending on the size and characteristics of the example set [29]. Further, this can be applied to limit the depth of the decision tree and the default value is 10, that keep as it is.

Under the KNN algorithm the K value was decreased, which was 5 by default. It finds the K number of training examples which are closest to the unknown examples. It should be a small, positive and an odd number. Then in SVM, the Kernel type and Kernel cache can be changed to enhance the accuracy. It supports different Kernel types such as Anova, Dot, Polynomial, Neural etc. Different kernel types were tried in order to improve the accuracy of the model. In linear regression algorithm the feature selection can be changed into none, greedy, M5 prime, T-test and iterative T-test. Further, the minimum tolerance ratio can be specified. In Random Forest algorithm, the number of trees can be specified starting from 1 to infinity. The default value is 100 and it was kept as it is. The criterion can be again set to gain ration, Gini index, information gain, accuracy or least square. Based on the performance, it was set to Gini index. Under LR, feature selection parameter describes the feature selection method to be used in regression such as M5 prime, greedy, T-test and so on [29]. It can be set to none as well.

D. Cross Validation

Cross validation operator can be used to assess how accurately a model may perform in practical usage. It contains two sub parts the training sub process and the testing sub process. A model can be trained by using the training sub process. Then, that trained model will be used for predictions under the testing sub process. During the Testing phase, the model's performance is evaluated. Here it creates the number of validation subsets internally by itself. The input data set is divided into K number of samples in equal size. Among these K number of samples, one subset is considered as the test data and the rest (K-1) is

ISSN: 2961-5992

considered as the training data set. This process is then iterated K number of iterations. Then, the outcomes of K iterations are averaged in order to generate the final output. In this study, cross validation was used.

VI. EVALUATION

A. Evaluation Parameters

Since this study has two stages of predictions different performance criteria have been used according to the model. In the first phase of analysis, accuracy, precision, recall, and F-measure were used as it is a binominal prediction. Accuracy was used in measuring the correct rate of predictions. Precision was used as measure which indicates the number of true positive predictions among all positive predictions where recall measures the number of true positive predictions among number of positive examples. Then F-measure was used which generates a single score by combining precision and recall. Since the second prediction is a polynomial one, accuracy, classification error, correlation, squared correlation and Kappa statistic were used. Classification error measures the relative number of misclassified examples. Kappa statistic is a more robust measure since it considers the correct prediction occurring by chance rather than a simple percentage of correct prediction. Correlation was used, which is also an important aspect to evaluate a model which returns the correlation coefficient between the class label and the independent variables. Another important metric used is squared correlation that returns the squared correlation coefficient between the label and the prediction attributes. There are various other performance metrics available such as absolute error, actual values, class weight, root mean squared error, weighted mean recall etc.

B. Prediction Results

The first prediction predicts the status of the payment (predicting whether a loan will be fully paid or not). Five algorithms were selected to apply here, based on the thorough literature review done. A balanced data set with around 3050 records was used with the previously mentioned attributes. The results of different algorithms applied in the first stage were depicted in the table below. By considering the performance metrices the Random Forest algorithm was selected for the first stage of the analysis to predict whether a loan will get paid fully or not.

In the second stage it predicts the non-performing loan in another perspective. If a particular loan is identified as a non-performing loan, then the model predicts the percentage of the non-payments. The results obtained from the first prediction were imported for this second phase by exporting the results into an excel file. Then the samples are filtered out to get the records which are actually arrears and the prediction is also arrears.

TABLE 1:
PREDICTION 01- PERFORMANCE METRICS OF DIFFERENT ALGORITHMS

Algorithm	Accuracy	Precision	Recall	F Measure
Naïve Bayes	58.23%	58.83%	63.79%	61.11%
Rule Induction	66.22%	71.82%	60.44%	64.35%
Decision Tree	63.98%	97.05%	30.94%	46.69%
KNN	61.27%	61.88%	64.36%	63.05%
Random Forest	70.00%	72.74%	66.57%	69.45%

As the percentage was discretized into four categories as high, medium, low and very low the prediction will be polynomial. All the algorithms applied in prediction one was applied here also. Parameters of the algorithms were tuned to get more performance and the KNN algorithm was selected for this stage since it showed the highest performance (Refer Table 2).

TABLE 2:
PREDICTION 02- PERFORMANCE METRICS OF DIFFERENT ALGORITHMS

Algorithm	Accuracy	Classification Error	Correlation	Squared Correlation	Kappa Statistic
Naïve Bayes	53.9%	46.11%	19.1%	3.6%	13.2%
Rule Induction	62.06%	37.94%	32.0%	10.3%	29.5%
Decision Tree	62.63%	37.37%	36.2%	13.1%	28.7%
Random Forest	72.81%	27.19%	56.4%	31.8%	50.0%
KNN	73.87%	26.13%	57.3%	32.8%	52.3%

VII. CONCLUSION

Predicting the credit risk is essential in the financial sector for the reduction of non-performing loans and for better decision making. Among various types of loans SME loans are imperative. SMEs laid the foundation of the economic growth of a country. Besides, in Sri Lankan context, the number of non-performing loans have increased drastically throughout the last few years. Banks and other financial institutes use various techniques to predict SME loan credit risk, but there are some inefficiencies in the process. This study adopted a comprehensive model to predict different aspects of SME credit risk by using data mining algorithms to address that issue. The predictions were done in two stages to analyze the non-performing loans in different perspectives. First it predicts whether a particular loan borrower will pay a loan fully or not. According to the performance of classification algorithms, Random Forest algorithm shows the highest accuracy for this prediction. Then, if a loan is identified as a non-performing loan the second phase predicts the

ISSN: 2961-5992

percentage of non-payments. KNN algorithms was identified as the most accurate algorithm in this level.

A. Limitations

There are a few limitations to this study as in any research. The unavailability of more SME loan data can be a limitation since it focuses on big data analytics. If there is a large amount of data, then the predictions will be more accurate since this study uses data mining techniques as a machine learning approach. There were some overfitting issues when came to the later stages of predictions. It can be solved with a large set of data. Further, if data can be obtained from different financial institutes, including banks, it will give more insights to this study.

Variables were also selected considering only the available dataset from one financial institution. Hence, the results to the research problem can be limited due to the unobtainability of data from different financial institutions. Furthermore, this study has adopted only classification algorithms for the analysis. Though previous studies have emphasized classification as the most used data mining technique in predicting credit risk, other types of data mining techniques can be applied such as clustering to identify the distinct clusters within that data set. By addressing these limitations, a more accurate and efficient data mining approach can be created to predict SME loan credit risk.

B. Future Development

As future enhancements of this study, the model will be finetuned with more examples since this study analyzed only 3000 SME loan data. Data can be obtained from different types of financial institutes such as state and private banks and other institutes. Further, it can be expanded for other types of loans including personal loans, professional loans, property loans, educational loans etc. In addition to that, an integrated model can be developed in predicting different aspects of the credit risk which are not addressed here. It will enhance the overall performance of the model and ultimately it provides a valuable contribution for the financial sector in Sri Lanka by enhancing effective decision making.

REFERENCES

[1] I. G. N. N. Mandalaa, C. B. Nawangpalupi and F. R. Praktikto, "Assessing Credit Risk: an Application of Data Mining in a Rural Bank," Procedia Economics and Finance, pp. 406-412, 2012.

[2] S. Khemakhem, F. B. Said and Y. Boujelbene, "Credit risk asseessment for unbalanced datasets based on data mining, artificial neural network and support vector machines," Journal of Modelling in Management, 2018.

[3] F. Ciampi, "Corporate governance characteristics and default prediction modeling for small enterprises. An empirical analysis of Italian firms," Journal of Business Research, vol. 68, no. 5, pp. 1012-1025, 2015.

[4] M. Delivand, M. Moghadam, N. Shamami and M. Taghipour, "Investigating the effective factors in measuring customers' credibility with a combined approach of data mining and multidisciplinary decision making Problem," Journal of Military Operations Research, vol. 1, no. 1, 2021.

[5] Liu-yi and Z. Li-Gu, "Research and application of credit risk of small and medium-sized enterprises based on random forest model," IEEE International Conference on Consumer Electronics and Computer Engineering, 2021.

[6] M. o. I. a. Commerce, National Policy Framework for SME Development, Ministry of Industry and Commerce, 2015.

[7] H. M. K. S. Bandara, A. L. M. Jameel and H. Athambawa, "Credit Risk and Profitability of Banking Sector in Sri Lanka," Journal of Economics, Finance and Accounting Studies, vol. 1, p. 3, 2021.

[8] J. Gupta, N. Wilson, A. Gregoriou and J. Healy, "The value of operating cash flow in modelling credit risk for SMEs," Applied Financial Economics, vol. 24, no. 9, pp. 649-660, 2014.

[9] W. Chen, G. Xiang, Y. Liu and K. Wang, "Credit risk Evaluation by hybrid data mining technique," Systems Engineering Procedia, vol. 3, 2012.

[10] S. Department, "CREDIT SUPPLY SURVEY," Central Bank of Sri Lanka, 2021.

[11] D. o. S. o. N.-B. F. Institutions, "Proposed New Direction on Credit Risk Management for Liscened Finance Companies," Central Bank of Sri Lanka, 2019.

[12] S. Kodithuwakku, "Impact of credit risk management on the performance of commercial banks in Sri Lanka.," 2015.

[13] A. Hamid and T. Ahmed, "DEVELOPING PREDICTION MODEL OF LOAN RISK IN BANKS USING DATA MINING," Machine Learning and Applications: An International Journal (MLAIJ), vol. 3, no. 1, 2016.

[14] Koyuncugil, A. S. and Ozgulbas, N., "Financial early warning system model and data mining application for risk detection," Expert systems with Applications, vol. 39, no. 6, pp. 6238-6253, 2012.

[15] Abedin, M. Z., Guotai, C., Hajek, P. and Zhang, T. , "Combining weighted SMOTE with ensemble learning for the class-imbalanced prediction of small business credit risk," Complex & Intelligent Systems, pp. 1-21, 2022.

[16] A. Islam, S. Yousuf and I. Rahman, "SME Financing in Bangladesh: A Comparative Analysis of Conventional and Islamic Banks," Journal of Islamic Banking and Finance, vol. 2, no. 1, pp. 79-92, 2014.

[17] S. Khemakhem and Y. Boujelbene, "Predicting credit risk on the basis of financial and non-financial variables and data mining," Review of Accounting and Finance, vol. 17, no. 3, 2018.

[18] V. Rivai, A. P. Veithzal and N. I. Ferry, "Bank dan Financial Institution Management convensional dan Shariah System.," 2007.

[19] S. Moradi and F. M. Rafiei, "A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks," Financial Innovation, vol. 5, no. 1, pp. 1-27, 2019.

[20] C. Huang, M. Chen and C. Wang, "Credit scoring with a data mining approach based on support vector machines," Expert Systems with Applications, vol. 33, no. 4, pp. 847-856, 2007.

[21] D. T. Larose, An introduction to data mining, John Wiley & Sons, 2005.

[22] Chen, Tsung-Kang, H.-H. Liao and W.-H. Chen, "CEO ability heterogeneity, board's recruiting ability and credit risk," Review of Quantitative Finance and Accounting, vol. 49, no. 4, pp. 1005-1039, 2017.

[23] A. K. Abdelmoula, "Bank credit risk analysis with k-nearestneighbor classifier: Case of Tunisian banks," Accounting and Management Information Systems, vol. 14, no. 1, 2015.

[24] M. A. Mukid, T. Widiharih, A. Rusgiyono and A. Prahutama, "Credit scoring analysis using weighted k nearest neighbor," IOP Conf. Series: Journal of Physics, 2018.

[25] P. Danenasa, G. Garsva and S. Gudas, "Credit Risk Evaluation Model Development Using Support Vector Based Classifiers," International Conference on Computational Science, 2011.

[26] B. W. Yap, S. H. Ong and N. H. M. Husain, "Using data mining to improve assessment of credit worthiness via credit scoring models," Expert Systems with Applications, vol. 38, no. 10, 2011.

[27] A. Khashman, "Credit risk evaluation using neural networks: Emotional versus conventional models," Applied Soft Computing, vol. 11, no. 8, pp. 5477-5484, 2011.

[28] S. Klotz and A. Lindermeir, "Multivariate credit portfolio management using cluster analysis," The Journal of Risk Finance, vol. 16, no. 2, pp. 145-163, 2015.

[29] "RapidMiner Documentation," RapidMiner, 2021. [Online]. Available: https://docs.rapidminer.com/. [Accessed 06 12 2021].

ISSN: 2961-5992

[30] S. M. Sadatrasoul, M. R. Gholamian, M. Siami and Z. Hajimohammadi, "Credit scoring in banks and financial institutions via data mining techniques: A literature review," Journal of AI and Data Mining, vol. 1, no. 2, 2013.

ISSN: 2961-5992