Enhancing Credit Risk Assessment Through Machine Learning

Oluseyi Lawal
Business insights and analytics
Humber College
Ontario, Canada
N01488609@humber.ca

Rogy Chacko Joseph
Business insights and analytics
Humber College
Ontario, Canada
N01510996@humber.ca

Ayotunde Olowookere
Business insights and analytics
Humber College
Ontario, Canada
N01203097@humber.ca

Nirav Sanjay Modi
Business insights and analytics
Humber College
Ontario, Canada
N01514238@humber.ca

Abstract— **Financial institutions, particularly banks, grapple with accurately predicting credit risk in loan applications, a challenge intensified by multifaceted factors and ever-changing economic conditions [1]. This study proposes a sophisticated machine learning algorithm as a solution. The algorithm, rooted in thorough borrower analysis encompassing financial history, employment status, and credit score, aims to provide an effective tool for informed lending decisions.**

The proposed solution addresses vital needs, including refining lending decisions, curtailing non-performing loans, fostering financial inclusion, minimizing credit losses, automating credit assessments, elevating customer experiences, spotting growth prospects, and mitigating systemic risks [2]. Crucial steps encompass data preparation, feature engineering, algorithm selection, model training, validation, and performance evaluation.

By achieving this, the algorithm enhances credit risk prediction, supporting banks in sound lending practices, comprehensive risk management, and prudent decision-making. The potential impact is significant: it encourages responsible lending, advances financial inclusion efforts, and fortifies the industry's stability. Ultimately, this predictive model holds promise in transforming financial decision landscapes, reducing vulnerabilities, and nurturing financial well-being.

Keywords—Credit Risk Prediction, Machine Learning, Lending Decisions, Financial Inclusion, Risk Management, Systemic Risks, Non-Performing Loans

I. Introduction

Lending decisions can have a big impact on people and businesses. Financial institutions need to be able to accurately assess the risk of default in order to make sound lending decisions. Traditional methods of credit risk assessment are often not enough, as they can't account for all the factors that contribute to default risk.

Machine learning can be used to develop more accurate credit risk prediction models. These models can analyze a variety of factors, including financial history, employment status, and credit scores, to predict the likelihood of default [3]. This can help financial institutions make better lending decisions and reduce the risk of non-performing loans.

The goal of this analytics project is to develop a machine learning algorithm that accurately predicts whether a credit request poses a risk to the bank. The model will use historical credit application data and applicant information to assess creditworthiness and classify credit requests into either low-risk or high-risk categories. The main objective is to provide the bank with a data-driven tool for making informed lending decisions and effectively managing credit risk.

We believe that machine learning has the potential to revolutionize credit risk assessment. By using machine learning, financial institutions can make better lending decisions, reduce the risk of non-performing loans, and improve their overall financial performance.

About the Dataset

The Statlog (German Credit Data) **dataset** used in this study was created by Hans Hofmann. It is part of the UCI Machine Learning Repository and is used for credit risk classification. The dataset consists of 1000 instances, each described by 20 attributes. The attributes can be categorical or integer, making the dataset multivariate and well-suited for classification tasks.

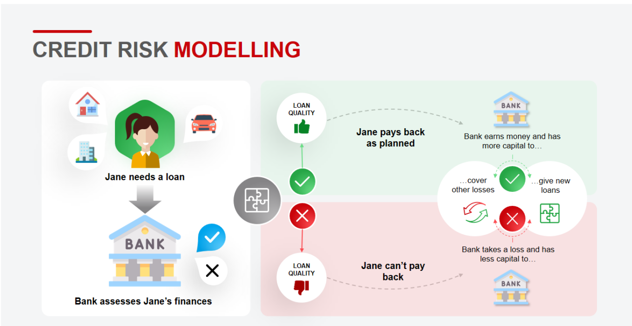
The purpose of the dataset is to classify individuals as either good or bad credit risks. The objective of building a predictive model using this dataset is to accurately predict the credit risk associated with each individual. This information can then be used by financial institutions to make informed lending decisions.

The dataset is publicly available and can be downloaded from the UCI Machine Learning Repository. It is a valuable resource for researchers and practitioners who are interested in credit risk classification.

We hope that this paper will provide a better understanding of the potential of machine learning for credit risk prediction. We believe that this technology has the potential to make a significant positive impact on the financial industry.

II. BACKGROUND STUDY

Lending and borrowing are essential parts of the economy. They allow people and businesses to get the money they need to make investments and grow. However, lending also involves risk. If a borrower defaults on a loan, the lender may not get their money back [4].



Credit Risk Modelling [5]

Traditional methods of credit risk assessment have been based on a borrower's financial history and credit score. However, these methods are not always accurate. They may not take into account all the factors that can affect a borrower's ability to repay a loan.

Machine learning can be used to develop more accurate credit risk prediction models. These models can analyze a variety of factors, including financial history, employment status, and credit scores, to predict the likelihood of default. This can help lenders make better lending decisions and reduce the risk of non-performing loans.

In recent years, there has been a growing interest in using machine learning for credit risk prediction. This is due to the increasing complexity of the financial landscape and the need for more accurate methods of risk assessment.

The benefits of using machine learning for credit risk prediction include [6]:

- Enhancing lending decisions: A robust machine learning algorithm can provide an objective and data-driven assessment of a borrower's creditworthiness, improving the overall lending process.
- Reducing non-performing loans: By using machine learning to predict credit risk, banks can proactively identify potential high-risk borrowers and reduce the likelihood of loan defaults.
- Increasing financial inclusion: A well-calibrated credit risk prediction algorithm can help banks extend credit to previously underserved populations while maintaining responsible lending practices.
- Minimizing credit losses: A machine learning algorithm that accurately predicts credit risk can assist banks in managing their loan portfolios more effectively, thereby minimizing credit losses and enhancing overall financial stability.
- Automating credit assessment: Manual credit assessment processes can be time-consuming and labor-intensive. Implementing a machine learning algorithm allows for automation of credit risk assessment, enabling faster decision-making and streamlining operational efficiency.
- Improving customer experience: A reliable credit risk prediction model can expedite loan approvals for low-risk applicants, leading to an improved customer experience. This can attract more customers to the bank and increase customer satisfaction.
- Identifying growth opportunities: A machine learning algorithm that effectively predicts credit risk can provide insights into profitable lending segments and identify potential growth opportunities for the bank.
- Mitigating systemic risks: By using data-driven credit risk prediction, banks can contribute to mitigating systemic risks in the financial sector, as a more accurate assessment of credit risk reduces the potential for financial crises.

A. Traditional Method of Credit assessment

Traditional credit assessment is the process of evaluating a borrower's creditworthiness using factors such as credit score, debt-to-income ratio, and collateral [7]. It is a manual process that can be time-consuming and subjective.

Here are several conventional approaches employed for credit assessment, along with the challenges they commonly encounter: -

- **The Credit score**: A credit score is a numerical summary of a borrower's credit history. It is calculated using a variety of factors, including payment history, debt-to-income ratio, and length of credit history. Credit scores are a good indicator of a borrower's creditworthiness, but they are not perfect. They can be inaccurate for borrowers with limited credit history, and they may not take into account all the factors that can affect a borrower's ability to repay a loan.
- **Debt-to-income ratio:** A debt-to-income ratio is a measure of how much debt a borrower has compared to their income. A high debt-to-income ratio is often seen as a sign of financial risk, as it means that the borrower may have difficulty repaying their debts. However, a debt-to-income ratio can be misleading for borrowers who have high-interest debt, such as credit card debt.
- **Collateral**: Collateral is an asset that a borrower pledges to a lender in the event of default. If the borrower defaults on the loan, the lender can seize the collateral to recoup their losses. Collateral can be a good way to reduce credit risk, but it is not always

available. For example, borrowers who are self-employed may not have any assets that they can pledge as collateral.

- **Personal interview**: A personal interview is a traditional method of credit assessment that allows the lender to get to know the borrower and assess their character. Personal interviews can be helpful in identifying borrowers who are likely to be good credit risks, but they can also be subjective and time-consuming.

B. Machine Learning for Credit Risk Assessment

Machine learning is a type of artificial intelligence that can be used to develop more accurate models of credit risk. Machine learning algorithms can analyze large amounts of data, including historical loan data, borrower characteristics, and economic indicators, to identify patterns that can be used to predict the likelihood of default [8].

The advantages of using Machine Learning for Credit Risk Assessment are stated in the Background Study above.

III. Methodology

Creating a credit risk prediction model is a complex process that requires a systematic approach. The methodology involves several steps, including data preprocessing, feature selection, model training, and evaluation.

A. Data Loading and Initial Exploration

The dataset is imported from the Statlog (German Credit Data) dataset. It is checked for duplicate rows and missing values. The dataset used in this study is a publicly available dataset that is often used for credit risk prediction tasks.

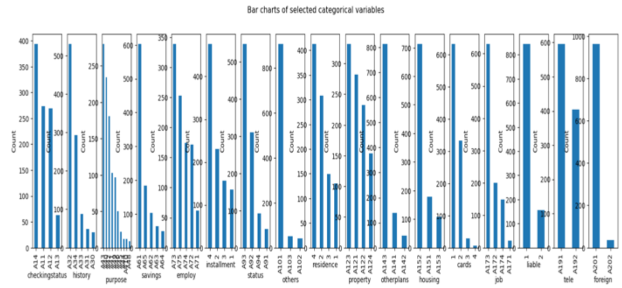
	GoodCredit	checkingstatus	duration	history	purpose	amount	savings	employ	installment	status	...	residence	property	age	otherplans	housing
count	1000.000000	1000	1000.000000	1000	1000	1000.000000	1000	1000	1000.000000	1000	...	1000.000000	1000	1000.000000	1000	1000
unique	NaN	4	NaN	5	10	NaN	5	5	NaN	4	...	NaN	4	NaN	3	3
top	NaN	A14	NaN	A32	A43	NaN	A61	A73	NaN	A93	...	NaN	A123	NaN	A143	A152
freq	NaN	394	NaN	530	280	NaN	603	339	NaN	548	...	NaN	332	NaN	814	713
mean	0.300000	NaN	20.903000	NaN	NaN	3271.258000	NaN	NaN	2.973000	NaN	...	2.845000	NaN	35.546000	NaN	NaN
std	0.458487	NaN	12.058814	NaN	NaN	2822.736876	NaN	NaN	1.118715	NaN	...	1.103718	NaN	11.375469	NaN	NaN
min	0.000000	NaN	4.000000	NaN	NaN	250.000000	NaN	NaN	1.000000	NaN	...	1.000000	NaN	19.000000	NaN	NaN
25%	0.000000	NaN	12.000000	NaN	NaN	1365.500000	NaN	NaN	2.000000	NaN	...	2.000000	NaN	27.000000	NaN	NaN
50%	0.000000	NaN	18.000000	NaN	NaN	2319.500000	NaN	NaN	3.000000	NaN	...	3.000000	NaN	33.000000	NaN	NaN
75%	1.000000	NaN	24.000000	NaN	NaN	3972.250000	NaN	NaN	4.000000	NaN	...	4.000000	NaN	42.000000	NaN	NaN
max	1.000000	NaN	72.000000	NaN	NaN	18424.000000	NaN	NaN	4.000000	NaN	...	4.000000	NaN	75.000000	NaN	NaN

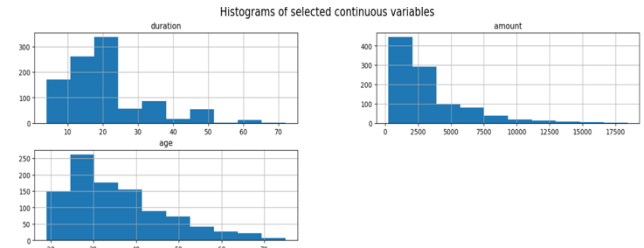
11 rows × 21 columns

Descriptive statistics for the dataset

B. Data Exploration and Visualization

The distribution of the target variable is visualized using a bar chart and a summary information about the dataset is presented, including data types and missing values. Columns with substantial missing values are eliminated. Also, qualitative variables are excluded, and descriptive statistics are calculated and reported.





Notably, all categorical columns, except "foreign" and "others," exhibit satisfactory distributions for machine learning, highlighting the importance of balanced data for effective model training and prediction [9]. For the continuous variable in the Histogram plot, the preferred outcome is a bell-shaped or slightly skewed bell-shaped curve which indicates a normal distribution. Age, Amount, and Duration are acceptable from the plot.

C. Feature Selection

The nature of each column is determined based on the count of unique values. Categorical variables are visualized using bar charts and histograms are used to visualize the frequency distribution of selected continuous variables.

The feature selection steps in this study are designed to identify the features that are most likely to be predictive of credit risk. This is done by using statistical methods and domain knowledge.

D. Statistical Tests for Feature Selection

The ANOVA technique is used to assess the correlation between continuous predictors and the categorical target variable.

Chi-Square tests are used to investigate potential associations between categorical predictors and the categorical target variable.

E. Data Preprocessing

Ordinal categorical variables are converted into numeric values while binary nominal categorical variables are transformed into numeric representations. Also, dummy variables are generated for non-binary nominal categorical columns.

The data preprocessing steps are designed to ensure that the data is clean, consistent, and in a format that is compatible with the machine learning algorithms that will be used to build the model.

F. Data Normalization, Splitting, and Sanity Check

The Min-Max normalization is applied to ensure uniform scaling of predictor variables and the dataset is divided into training and testing subsets of 70% and 30% respectively. The sizes of the training and testing datasets are verified.

G. Model Evaluation

Seven classification models were used to evaluate the given dataset and the performance metrics such as precision, recall, F1-score, and accuracy are calculated. Cross-validation and learning curves are used to assess model stability and generalization performance. The models used are Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbours, Support Vector Machines, Naïve Bayes, Neural Network (MLP Classifier).

The model evaluation steps in this study are designed to assess the performance of the models on a holdout dataset. This allows the model to be evaluated on data that it has not seen before, which provides a more accurate assessment of its performance.

IV. RESULTS

A. Results of the models are as follows: -

i. Logistic Regression

The logistic regression model achieved an accuracy of 0.76 on the testing data. This means that the model correctly predicted the credit risk of 76% of the test data points. The model also had a precision of 0.80 and a recall of 0.90, which means that it was 80% accurate in predicting good loans and 90% accurate in predicting bad loans.

The confusion matrix for the logistic regression model shows that the model correctly classified 189 good loans and 43 bad loans. The model incorrectly classified 20 good loans as bad loans and 48 bad loans as good loans.

The 10-fold cross validation results showed that the average accuracy of the logistic regression model was 0.74. This means that the model was able to achieve an accuracy of 74% on 10 different folds of the data. The standard deviation of the accuracy values was 0.05, which indicates that the model was relatively consistent in its performance.

ii. Decision Tree

The decision tree model achieved an accuracy of 0.67 on the testing data. This means that the model correctly predicted the credit risk of 67% of the test data points. The model also had a precision of 0.74 and a recall of 0.85, which means that it was 74% accurate in predicting good loans and 85% accurate in predicting bad loans.

The confusion matrix for the decision tree model shows that the model correctly classified 177 good loans and 32 bad loans. The model incorrectly classified 32 good loans as bad loans and 61 bad loans as good loans.

The 10-fold cross validation results showed that the average accuracy of the decision tree model was 0.7. This means that the model was able to achieve an accuracy of 70% on 10 different folds of the data. The standard deviation of the accuracy values was 0.03, which indicates that the model was relatively consistent in its performance.

iii. Random Forest

The random forest model achieved an accuracy of 0.75 on the testing data. This means that the model correctly predicted the credit risk of 75% of the test data points. The model also had a precision of 0.78 and a recall of 0.93, which means that it was 78% accurate in predicting good loans and 93% accurate in predicting bad loans.

The confusion matrix for the random forest model shows that the model correctly classified 195 good loans and 14 bad loans. The model incorrectly classified 14 good loans as bad loans and 55 bad loans as good loans.

The 10-fold cross validation results showed that the average accuracy of the random forest model was 0.74. This means that the model was able to achieve an accuracy of 74% on 10 different folds of the data. The standard deviation of the accuracy values was 0.02, which indicates that the model was relatively consistent in its performance.

iv. K-Nearest Neighbors (KNN)

The KNN model achieved an accuracy of 0.74 on the testing data. This means that the model correctly predicted the credit risk of 74% of the test data points. The model also had a precision of 0.79 and a recall of 0.87, which means that it was 79% accurate in predicting good loans and 87% accurate in predicting bad loans.

The confusion matrix for the KNN model shows that the model correctly classified 181 good loans and 28 bad loans. The model incorrectly classified 28 good loans as bad loans and 47 bad loans as good loans.

The 10-fold cross validation results showed that the average accuracy of the KNN model was 0.7. This means that the model was able to achieve an accuracy of 70% on 10 different folds of the data. The standard deviation of the accuracy values was 0.02, which indicates that the model was relatively consistent in its performance.

V. SUPPORT VECTOR MACHINES (SVM)

The SVM model achieved an accuracy of 0.73 on the testing data. This means that the model correctly predicted the credit risk of 73% of the test data points. The model also had a precision of 0.78 and a recall of 0.89, which means that it was 78% accurate in predicting good loans and 89% accurate in predicting bad loans.

The confusion matrix for the SVM model shows that the model correctly classified 185 good loans and 24 bad loans. The model incorrectly classified 24 good loans as bad loans and 53 bad loans as good loans.

The 10-fold cross validation results showed that the average accuracy of the SVM model was 0.75. This means that the model was able to achieve an accuracy of 75% on 10 different folds of the data. The standard deviation of the accuracy values was 0.01, which indicates that the model was relatively consistent in its performance.

vi. Naive Bayes (NB)

The Naive Bayes model achieved an accuracy of 0.72 on the testing data. This means that the model correctly predicted the credit risk of 72% of the test data points. The model also had a precision of 0.85 and a recall of 0.72, which means that it was 85% accurate in predicting good loans and 72% accurate in predicting bad loans.

The confusion matrix for the Naive Bayes model shows that the model correctly classified 150 good loans and 59 bad loans. The model incorrectly classified 27 good loans as bad loans and 64 bad loans as good loans.

The 10-fold cross validation results showed that the average accuracy of the Naive Bayes model was 0.7. This means that the model was able to achieve an accuracy of 70% on 10 different folds of the data. The standard deviation of the accuracy values was 0.02, which indicates that the model was relatively consistent in its performance.

vii. Neural Network Model

The Neural Network model achieved an accuracy of 0.73 on the testing data. This means that the model correctly predicted the credit risk of 73% of the test data points. The model also had a precision of 0.81 and a recall of 0.79, which means that it was 81% accurate in predicting good loans and 79% accurate in predicting bad loans.

The confusion matrix for the Neural Network model shows that the model correctly classified 166 good loans and 43 bad loans. The model incorrectly classified 27 good loans as bad loans and 38 bad loans as good loans.

The 10-fold cross validation results showed that the average accuracy of the Neural Network model was 0.71. This means that the model was able to achieve an accuracy of 71% on 10 different folds of the data. The standard deviation of the accuracy values was 0.02, which indicates that the model was relatively consistent in its performance.

Model	Accuracy	Precision	Recall	Avg. Accurac	SD (CV)
Logistic Regression	76%	80%	90%	74%	0.05
Decision Tree	67%	74%	85%	70%	0.03
Random Forest	75%	78%	93%	74%	0.02
K-Nearest Neighbors	74%	79%	87%	70%	0.02
Support Vector Machines	73%	78%	89%	75%	0.01
Naive Bayes	72%	85%	72%	70%	0.02
Neural Network Model	73%	81%	79%	71%	0.02

Summary of Results

B. Comparison and Model selection

The neural network model, the NB model, the SVM model, the KNN model, the random forest model, the decision tree model, and the logistic regression model all achieved accuracies of between 0.70 and 0.76 on the testing data. The neural network model had the highest average accuracy across the 10-fold cross validation, followed by the SVM model, the KNN model, the logistic regression model, the random forest model, the decision tree model, and the NB model.

The neural network model is a more complex model than the other models. This means that it can potentially achieve higher accuracy, but it also requires more training data and time to train [10]. The NB model is a simple model that is easy to interpret. However, it is not as accurate as the other models. The SVM model is a more complex model that can achieve higher accuracy than the NB model. However, the SVM model can be more difficult to interpret. The KNN model is a simple model that is easy to interpret. It is also relatively fast to train. However, the KNN model may not be as accurate as the other models. The random forest model is a more complex model that can achieve higher accuracy than the KNN model. It is also relatively fast to train. The decision tree model is a simple model that is easy to interpret. However, the decision tree model may not be as accurate as the other models. The logistic regression model is a relatively simple model that is easy to interpret. It is also relatively fast to train. However, the logistic regression model may not be as accurate as the other models.

C. Overview of the Neural Network Model

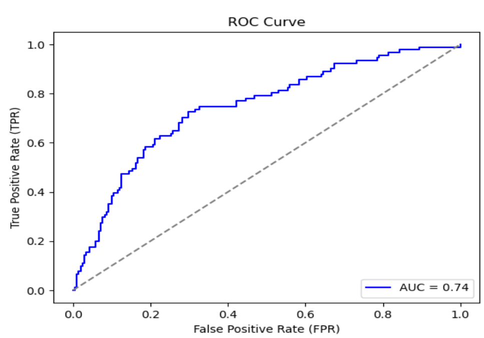
The Neural Network model has two hidden layers. The first hidden layer has 64 neurons, and the second hidden layer has 32 neurons. The activation function for each layer is relu, which is a non-linear function that helps the model learn complex relationships between the features and the target variable. The solver is set to adam, which is a popular optimization algorithm for training neural networks. The random state is set to 42, which ensures that the results are reproducible.

The Neural Network model was chosen because it has the following advantages over other models: -

- Higher accuracy: The Neural Network model achieved a relatively high average accuracy (71%) during cross-validation. This indicates that the model consistently performs well across different data folds, making it reliable for predicting credit risk.

- Balanced precision and recall: The Neural Network model demonstrated a balanced trade-off between precision (81%) and recall (79%). This indicates that the model effectively identifies both good and bad credit risks, minimizing the risk of false positives (approving risky loans) and false negatives (rejecting potential good loans).
- Capability for complex patterns: Neural Networks are known for their ability to capture intricate relationships and patterns within data. This is essential for credit risk prediction, as creditworthiness is influenced by a multitude of factors that may not have linear relationships. The Neural Network's inherent complexity allows it to learn and adapt to these complex interactions.
- Potential for model improvement: The Neural Network's performance can potentially be enhanced with more training data and further optimization of hyperparameters. As more data becomes available and the model is fine-tuned, it has the potential to achieve even higher accuracy.

In addition to these advantages, the Neural Network model is also scalable and adaptable to evolving scenarios. This makes it a good choice for financial institutions that are looking for a long-term solution for credit risk prediction.



Further analysis of the NN model shows that the ROC (Receiver Operating Characteristic) curve shows the trade-off between the true positive rate (TPR) and the false positive rate (FPR) of the binary classifier system as its discrimination threshold is varied. The area under the ROC curve (AUC) is 0.74 meaning that the classifier is better than a random classifier, but it is not perfect.

vii. Conclusion

The Neural Network model emerged as a frontrunner with the highest average accuracy during cross-validation. Its ability to capture intricate relationships in the data presents an opportunity for further improvement. With increased data and training efforts, the Neural Network model's accuracy could be further enhanced.

The selection of the Neural Network model as the preferred choice for credit risk prediction marks the beginning of an iterative and dynamic process. This process involves fine-tuning the hyperparameters of the model, identifying the most influential features, evaluating the model's performance, addressing class imbalance issues, exploring ensemble techniques, enhancing interpretability, refining and engineering features, addressing ethical considerations, deploying the model into production, collaborating with stakeholders, and continuously monitoring the model's performance. The goal of this process is to refine, optimize, and effectively leverage the model's capabilities to make informed credit risk decisions.

NB- “Detail code sheet is available in one of the attached files.”

References

[1] David Hillier, Mark Grinblatt and Sheridan Titman, “Financial markets and corporate strategy,” Second European Edition, pp. 28-57, 2012.

[2] Ogundokun RO, Misra S, Maskeliunas R, Damasevicius R. A Review on Federated Learning and Machine Learning Approaches: Categorization, Application Areas, and Blockchain Technology. *Information*. 2022; 13(5):263. https://doi.org/10.3390/info13050263

[3] L. Yu, S. A. Wang, and K. K. Lai, “Credit risk assessment with a multistage neural network ensemble learning approach,”Expert systems with applications, vol. 34, pp. 1434-1444, Feb. 2008, doi: 10.1016/j.eswa.2007.01.009

[4] A. A. Turjo, Y. Rahman, S. M. M. Karim, T. H. Biswas, I. Dewan and M. I. Hossain, "CRAM: A Credit Risk Assessment Model by Analyzing Different Machine Learning Algorithms," 2021 4th International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2021, pp. 125-130, doi: 10.1109/ICOIACT53268.2021.9563995.

[5] Powerslide (2022, January 9). *Credit Risk Modelling | Download editable diagrams & templates*. Powerslides. https://powerslides.com/powerpoint-business/business-models/credit-risk-modelling/

[6] Ghodselahi, Ahmad, and Ashkan Amirmadhi. "Application of artificial intelligence techniques for credit risk evaluation." *International Journal of Modeling and Optimization* 1, no. 3 (2011): 243.

[7] Richeson, L. “Predicting consumer credit performance: Can neural networks outperform traditional statistical methods”. *unpublished*

[8] Huang, Shian-Chang, and Cheng-Feng Wu. "Customer credit quality assessments using data mining methods for banking industries." *African Journal of Business Management* 5, no. 11 (2011): 4438.

[9] L. Nanni, and A. Lumini,” An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring,” Expert systems with applications, vol. 36, March. 2009, pp. 3028-3033, doi: 10.1016/j.eswa.2008.01.018

[10] T. G. Dietterich, “Machine learning research: Four current directions,” AI Magazine, vol. 18, no.4, pp. 97–136