An Intelligent Model to Assess the Credit Risk in Egyptian Banks

Khaled Fathy[1] *Mohamed Marie*[2] *Engy Yehia*[3]

Abstract

In the realm of financial and banking institutions, the art of forecasting and assessing banking risks holds paramount significance. Preserving the financial stability of banks is contingent upon adept risk management, a cornerstone in enhancing overall bank performance. Moreover, the effectiveness of financial and banking institutions can be gauged by their ability to systematically evaluate and mitigate risks. Among these risks, the assessment of banking credit risks looms large in contemporary times, given the heightened necessity for decision-makers to anticipate the likelihood of loan defaults. However, one formidable challenge persists: the inadequate assessment of banking credit risks. This challenge stems from the multifaceted factors that influence risk assessment and the soundness of credit decisions. In response to this pressing issue, our research presents a predictive model employing machine learning (ML) algorithms. Our objective is to facilitate informed credit decision-making and safeguard the financial assets of banks. In pursuit of this aim, we employed five machine learning classification algorithms: Artificial Neural Networks (ANN), Random Forest (RF), Logistic Regression (LR), Decision Tree (DT) and XGBoost (XGB). To ensure the robustness of our study, we utilized a real-world dataset gleaned from the historical records of a prominent Egyptian bank. Subsequently, we assessed the performance of our model based on key metrics such as accuracy, precision, recall, and the f1 score. The results showed that XGB exhibited the highest accuracy, underlining the potential for ML algorithms to revolutionize the assessment of banking credit risks.

Keywords: Machine learning; Financial institutions; Risk assessment; Loan defaults; Predictive model.

----

[1]- *Business Information Systems Department, Faculty of Commerce and Business Administration, Helwan University, Helwan, Cairo, Egypt.*
[2]- *Information Systems Department, Faculty of Computers and Artificial Intelligence, Helwan University, Helwan, Cairo, Egypt.*
[3]- *Information Systems Department, Faculty of Commerce and Business Administration, Helwan University, Helwan, Cairo, Egypt.*

نموذج ذكي لتقييم مخاطر الائتمان في البنوك المصرية

الملخص

في عالم المؤسسات المالية والمصرفية، يحتل فن التنبؤ وتقييم المخاطر المصرفية أهمية قصوى. ويتوقف الحفاظ على الاستقرار المالي للبنوك على مهارة ادارة المخاطر، وهو حجر الزاوية في تعزيز الأداء العام للبنوك. علاوة على ذلك، يمكن قياس فعالية المؤسسات المالية والمصرفية من خلال قدرتها على تقييم المخاطر وتقليلها بشكل منهجي وذكي. ومن بين هذه المخاطر، مخاطر الائتمان المصرفي التي تعد من من اهم المخاطر التي تشغل البنوك في الأوقات المعاصرة، نظرا للحاجة المتزايدة لصانعي القرار لتوقع احتمالية تعثر عن سداد القروض. ومع ذلك، لا يزال هناك تحدٍ هائل قائم، ألا وهو التقييم غير الكفؤ لمخاطر الائتمان المصرفي. وينبع هذا التحدي من العوامل المتعددة الأوجه التي تؤثر على تقييم المخاطر وسلامة القرارات الائتمانية. استجابة لهذه المشكلة، يقدم هذ البحث نموذجًا تنبؤيًا يستخدم خوارزميات التعلم الآلي لتنبؤ احتمالية تعثر العملاء عن سداد القروض. يهدف النموذج المقترح الى تسهيل عملية اتخاذ القرارات الائتمانية المستنيرة مما يساعد على حماية الأصول المالية للبنوك. ولضمان قوة دراستنا، استخدمنا مجموعة بيانات واقعية مستمدة من السجلات التاريخية لأحد البنوك المصرية البارزة. وبعد ذلك، قمنا بتقييم أداء النموذج المقترح استنادًا إلى المقاييس الأساسية التي تستخدم في قياس أداء نماذج التعلم الآلي. وأظهرت النتائج أن نموذج XGB أظهر أعلى دقة، مما يؤكد قدرة خوارزميات تعلم الآلة على إحداث ثورة في تقييم مخاطر الائتمان المصرفي.

الكلمات المفتاحية: التعلم الآلي، المؤسسات المالية، تقييم المخاطر، التخلف عن سداد القروض، نموذج تنبؤي.

1. Introduction

The significance of credit risk assessment is emphasized by the responsibility of banks to acknowledge and accept risk when providing loans and credit cards to consumers, as these kinds of financial products play a vital role in driving the economy. The need for this derives from the unstable balancing on which a bank's survival depends, where both cautious and highly risky approaches can leave a bank vulnerable (Moradi & Rafiei, 2019). Within the dynamic realm of finance and banking, a major barrier involves accurate assessment of credit risk. The importance of this effort has grown significantly as financial organizations increasingly depend on credit risk assessment models as the core of their decision-making processes, especially in the areas of lending and investing. Lack of addressing this issue can lead to serious consequences, such as making incorrect credit judgements that lead to loan defaults and significant financial losses for the bank. Financial organizations must prioritize accurate credit risk forecasting in order to reduce the possibility of loan defaults and protect the integrity of their assets. This involves implementing the latest technologies and procedures to guarantee that lending and investment choices are made based on an accurate and reliable assessment of credit risk (Lili Lai, 2020). Thereby, to avoid this loan defaults, the Banks and financial institutions have to assess the credit risk. To mitigate the risk of loan defaults and preserve the integrity of their funds, financial institutions must prioritize precise credit risk forecasting. This entails deploying sophisticated models and strategies to ensure that lending and investment decisions are based on a comprehensive and dependable assessment of credit risk (Madaan et al., 2021). Applying machine learning (ML) methods allows for the anticipation of credit risk in banking and the detection of crucial elements that contribute to loan default. This enables educated decision-making to enhance investments and protect the bank's capital. Prior research indicates that there is increasing curiosity in applying machine learning for banking risk assessment in order to predict loan defaults. Due to the emergence of the big data era and improvements in machine learning techniques, we now have a greater range of options for categorizing and predicting loan defaults, instead of strictly depending on manual processes. The financial industry may greatly benefit from the application of machine learning techniques, which consist of various models that can effectively tackle complicated prediction problems, especially in cases where the relationship between input factors and outputs is unclear or not revealed. Previously, credit scoring utilized a combination of specialists and statistical algorithms to accurately evaluate an individual's creditworthiness. However, in recent years, researchers and banking regulators have adopted the use of machine learning and deep learning algorithms to train classifiers that can independently forecast an applicant's credit score by analyzing their credit history and other historical information. This approach significantly simplifies the process of choosing suitable candidates prior to loan acceptance, resulting in a substantially more efficient process (Kalyani & Suri, 2023; Moscatelli et al., 2020). Prior studies have shown a significant predilection towards investigating the prediction of loan defaults through the application of diverse machine learning methods. However, these investigations mainly focused on individual or restricted aspects. Generally, these machine learning techniques are not focused on the analysis of economic aspects that influence the likelihood of default. Nevertheless, several studies have shown that they can achieve high levels of accuracy when tested on new data samples (Madaan et al., 2021; Kalyani & Suri, 2023). This study aims to explore the main factors that have a significant influence on predicting loan default by borrowers and determine the machine learning model exhibit the highest accuracy in forecasting loan defaults. Consequently, this study demonstrates how machine learning methods are utilized to calculate the likelihood of a borrower being categorized as defaulted or Not-defaulted, based on the use of machine learning models including Artificial Neural Networks (ANN), Decision Trees (DT), Random Forest (RF), XGBoost (XGB), and Logistic Regression (LR). These ML models were evaluated, and their outcomes assessed against each other. XGBoost (XGB) exhibited marginally superior performance compared to the other models, leading to the conclusion that it could serve effectively as a predictive tool for determining the likelihood of loan defaults. Consequently, banks could implement this machine learning model to streamline the loan default prediction process for new loan applicants, allowing for faster loan processing with minimal time and decreased risk. The findings of this study will contribute to Improve the lending portfolio in order to maximize returns, determine the crucial characteristics or factors that have a substantial impact on forecasting loan defaults, provide forecasting which help credit evaluators in making well-informed decisions, and ensure the protection of the monetary assets of both financial institutions and banks. The paper is structured to explore loan default prediction, starting with a review of relevant literature in Section 2 and classification methods in Section 3. Section 4 elucidates the proposed model, followed by Section 5, which represents the results and discussion of risk assessment features and ML models evaluation. The concluding remarks, and potential improvements are outlined in Section 6.

2. Related Work

This section provides a concise overview of prior efforts in developing machine learning models utilizing diverse algorithms to enhance the loan prediction procedure, aiding banking authorities and financial institutions in identifying suitable candidates with minimal credit risk. To discuss the contribution of ML in continuous quality improvement. We focused on some of the previous works that used different machine learning techniques such as Artificial Neural Network (ANN), Decision Tree (DT), K-Nearest Neighbor (KNN), Gaussian Naïve Bayes (Gaussian NB), Logistic Regression (LR), Neural Network (NN), Random Forest (RF), Naïve Bayes (NB) and Support Vector Machine (SVM).

In (Sun & Vasarhelyi, 2018), the researcher aimed to achieve a dual objective: firstly, the development of a prediction system for credit card delinquency risk modeling, and secondly, an exploration of the potential of deep learning in the credit risk domain. Utilizing real-world credit card data from a large Brazilian bank, the study employed a deep neural network to assess delinquency risk based on client attributes and spending behaviors. In comparison to various machine-learning algorithms, such as logistic regression, naive Bayes, traditional artificial neural networks, and decision trees, deep neural networks demonstrated superior predictive performance with higher F scores and area under the receiver operating characteristic curve (AUC-ROC). This successful application of deep learning signifies the significant potential of artificial intelligence in enhancing credit risk assessment for financial institutions and credit bureaus. In (Zhu et al., 2019), the objective was to address the risk of user loan default on P2P online lending platforms using machine learning techniques. Employing the Random Forest algorithm, the study utilized real-world user loan data from Lending Club. To tackle class imbalance, the SMOTE method was applied, followed by data cleaning and dimensionality reduction. The results demonstrated the superiority of the Random Forest algorithm over other methods, including logistic regression and decision trees, in accurately predicting loan defaults. This research carries significant implications for enhancing the risk assessment and sustainability of online lending platforms. In (Alsaleem & Hasoon, 2020), the research objective was to assess the performance of various machine learning algorithms in classifying bank loan risks. With the increasing importance of bank loans and the challenges posed by the abundance of borrower data, this research aimed to aid banks in making informed grant decisions. The study utilized machine learning techniques, including Multilayer Perceptron, RandomForest, BayesNet, NaiveBayes, and DTJ48, to classify loans based on risk. Among these algorithms, Multilayer Perceptron demonstrated the highest accuracy, making it the preferred choice for risk classification in comparison to the other methods. This research has significant implications for enhancing banks' loan management and ultimately contributes to informed decision-making in the financial sector. In (Rath et al., 2020), the research objective aimed to streamline and expedite the loan approval process in banks by leveraging machine learning techniques and classification algorithms. The methodology involved training a model with historical loan applicant data and approval outcomes. Key findings revealed that the logistic regression model outperformed other algorithms, effectively categorizing applicants as deserving or undeserving of loans. This approach not only reduced the risk factor in applicant selection but also offered a potential solution for faster loan processing in the banking sector, ultimately saving time and bank resources. In conclusion, implementing machine learning tools in the loan approval process holds promising implications for improving decision-making efficiency and minimizing defaults. In (Dosalwar et al., 2021), the research objective was to enhance the banking system's profitability by accurately predicting loan defaulters, recognizing the significant impact of defaulting on a bank's profit and loss. The methodology employed logistic regression models and utilized data from Kaggle, incorporating variables such as customer personal attributes and checking account details. The results indicated that this approach significantly improved the accuracy of identifying potential loan defaulters, emphasizing the importance of considering various attributes in credit decisions. This research implies that banks should consider a creditor's broader characteristics when issuing loans, ultimately aiding in better forecasting and decision-making within the banking sector. In (Anand et al., 2022), the primary research objective was to improve loan default prediction, a critical factor influencing credit scores and financial organizations' earnings. Traditional and machine learning techniques, including Multiple Logistic Regression, Decision Tree, Random Forests, Gaussian Naive Bayes, Support Vector Machines, and ensemble methods, were employed. The dataset included loan data from various internet sources and applicants' loan applications. The results showcased comprehensive evaluation metrics, including Confusion Matrix, Accuracy, Recall, Precision, F1-Score, ROC analysis area, and Feature Importance. The study concluded that the Extra Trees Classifier and Random Forest exhibited the highest accuracy in predictive modeling, offering valuable insights for more effective credit disapproval decisions in a large number of loan applications, particularly for vulnerable consumers. In (Wang et al., 2023), the research objective aimed to compare the effectiveness of Random Forest and Decision Tree algorithms in predicting loan approvals using novel random forest classifiers. They employed loan prediction datasets from Kaggle, with a total sample size of 20 divided into two groups, Random Forest (N=10) and Decision Tree (N=10), using G-power at 80% computation. Results revealed that the Random Forest method achieved a higher precision of 79.4490% and lower loss of 21.0310% compared to the Decision Tree's precision of 67.2860% and loss of 32.7140%. The study concludes that Random Forests outperform Decision Trees in accurately predicting loan acceptance, supported by a statistically insignificant independent sample T-test result (p=0.33, p>0.05) with a 95% confidence level. (Dasari et al., 2023) aimed at addressing the challenge of identifying eligible loan recipients for banks and financial organizations. They recognized that existing models had limited accuracy, with a maximum of 80%. To improve this, the researchers proposed a model that utilized various machine learning techniques and ensemble algorithms such as bagging and voting classifiers. The primary research objective was to predict loan eligibility, reducing human effort and processing time while achieving more accurate results than previous models. Their experimental results demonstrated a significant improvement, surpassing the 80% accuracy threshold, offering promising implications for the financial sector. In (Abdullah et al., 2023), the research objective focused on forecasting nonperforming loans in financial institutions, recognizing their significance in overall performance. Employing machine learning techniques, the study utilized quarterly cross-sectional data from 322 banks in 15 emerging countries. Notably, advanced machine learning models surpassed linear techniques, with the random forest model achieving an impressive 76.10% accuracy in forecasting nonperforming loans. This robust result across various performance metrics highlights the model's efficacy. The variable importance analysis identified bank diversification as the key determinant, emphasizing its role in predicting future nonperforming loans. Additionally, the study underscored the limited influence of macroeconomic factors compared to bank-specific factors in this prediction process, offering valuable insights for financial institutions.

The conclusion of the mentioned studies highlights the importance of using machine learning models, particularly artificial neural networks, in assessing credit risk in banks. These studies demonstrate the effectiveness of various machine learning algorithms in predicting loan default probabilities, assisting decision-makers in credit evaluation, and improving credit risk assessment systems. However, there exists a notable gap in these studies regarding a comprehensive and unified approach to credit risk assessment that integrates the strengths of different machine learning methods, considering various data characteristics. This research aims to bridge this gap by building a predictive model using machine learning to assess credit risk in banks.

3. Classification Methods

Five distinct binary classification methods have been employed for the purpose of predicting loan defaults and making well-informed credit decisions. These algorithms operate by categorizing new data points into one of two classes, where the binary class in our dataset represents "0" for loans that did not default, signifying the eligibility of borrowers for loans, and "1" for defaulted loans, indicating the non-eligibility of borrowers for loans. The study utilized a dataset comprising 166239 records and leveraged Python programming language libraries including Scikit Learn, Pandas, NumPy, Matplotlib, and

Seaborn. The five classification algorithms employed are described in the following sections.

3.1 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is a mathematical model designed to emulate the structure and operations of biological neural networks. The fundamental components of any artificial neural network are artificial neurons, which serve as the fundamental mathematical units or functions. This model comprises three key sets of operations: multiplication, addition, and activation. Within an artificial neuron, each input value is multiplied by its respective weight. In the middle stage of the artificial neuron, a total function is computed, which encompasses the weighted sum of all input values. Finally, at the output of the artificial neuron, the weighted total input has undergone an activation phase, often referred to as a transfer function (Krenker et al., 2011).

3.2 Decision Trees (DT)

Decision Trees (DT) are a widely used method for classification in the field of credit scoring, featuring multiple branches, root nodes, and leaf nodes. This approach creates a tree-like structure that classifies instances by employing a recursive partitioning algorithm (RPA), as suggested by its name (Aslam et al., 2019) [17]. In our suggested approach, we employ the Gini method to establish splitting points by identifying a decision rule that maximizes the reduction in impurity at a given node. The Gini impurity at node t, denoted as G(t), is defined as follows:

$$G(t)=1-\sum_{i=1}^{c} Pi^2 \quad (1)$$

where i ranges from 1 to the number of classes, and Pi represents the proportion of observations belonging to class c at node t. This decision-making process is executed recursively until either all leaf nodes become pure or a predefined cut-off criterion is met.

3.3 Random Forest

Random Forest is a supervised learning algorithm and an ensemble method used for classification, regression, and other tasks. It constructs multiple decision trees during training and makes predictions by aggregating the outcomes (mode of the classes for classification or mean prediction for regression) of these individual trees, enhancing prediction accuracy and overfitting resistance (Zhu et al., 2019). The calculation of the Gini Index at an internal node of a tree is performed in the following manner: For a candidate (nominal) split attribute Xi, denote possible levels as $L1;\ldots;L$j.

$$G(X_i) := \sum_{j=1}^{J} \Pr(X_i = L_j)(1 - \Pr(X_i = L_j)) \quad (2)$$

3.4 XGBoost

XGBoost (XGB) as introduced by Chen and Guestrin in their work (Chen, T. & Guestrin, C., 2016), is a model based on gradient boosted decision trees. The XGBoost algorithm sequentially trains decision trees on the provided training data. In each iteration, the algorithm incorporates a new decision tree into the existing ensemble of trees, progressively enhancing the value of the objective function. XGBoost can effectively address real-world, large-scale problems while utilizing only a modest amount of computing resources.

3.5 Logistic Regression

Logistic Regression is a statistical algorithm used to predict the likelihood of an event through a classification function. This function computes statistics for a logistic response function, illustrating how the dependent variable is related to one or more independent variables (Rath et al., 2020). Within our proposed model, we incorporate a linear model within a logistic function, expressed as follows:

$$P(y) = 1/1 + e^{-z} \qquad (3)$$

In this context, P(y) represents the outcome, calculated using the dependent variable y, with z symbolizing the function based on independent variables in the dataset. P(y)'s predicted values fall between 0 and 1, assisting in categorizing the results into 'no' or 'yes'.

4. The proposed Intelligent Model

In this section, we will delve into the proposed Intelligent Model to Assess the Credit Risk in Egyptian Banks. Our research focused on integration of Machine Learning (ML) techniques for enhanced credit risk assessment. Our proposed model encompasses a comprehensive framework that amalgamates key components and algorithms specifically tailored for this task. We employed binary classification algorithms to categorise the target label as "0" for loans that did not default, indicating the borrowers' eligibility for loans, and "1" for defaulted loans, indicating the borrowers' ineligibility for loans. Figure 1 highlights the proposed model. The procedural steps involved are as follows: Dataset collection, Data pre-processing, Splitting the Data, Model Training using five machine learning models which discussed in section 4, and Model evaluation. Through a rigorous evaluation process, we aim to discern the most suitable model among the five ML approaches, with a particular focus on assessing credit risks within Egyptian banks. This proposed model aims to enhance the accuracy and reliability of credit risk assessment, ultimately contributing to more informed decision-making in the banking sector.

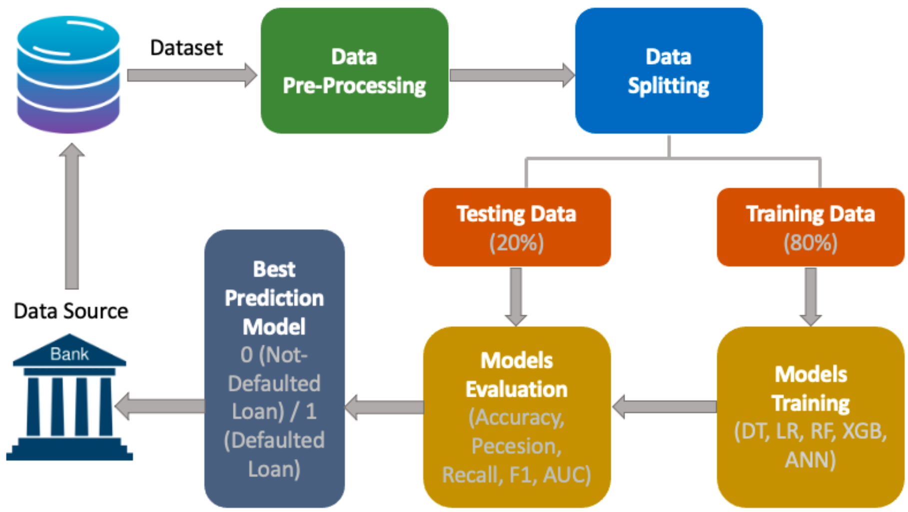

Figure. 1. The proposed intelligent model.

4.1 Data Source

The dataset utilized in this research was derived from the historical records of the Agriculture Bank of Egypt (ABE), one of the country's most prominent and established banks. ABE Bank boasts an extensive network comprising 1,100 branches spread across all governorates of Egypt. This rich internal database serves as the backbone for collecting and analyzing data pertinent to credit risk. The researcher acknowledged that selecting ABE bank's historical customer database as the primary data source offered several key advantages. First and foremost, the use of real-world data from an established financial institution allowed the research to reflect the complexities and dynamics of actual credit risk scenarios faced by banks in their lending operations. Secondly, the extensive timeframe covered by the historical data ensured a sufficiently large sample size, bolstering the statistical robustness of the ensuing credit risk model.

4.2 Dataset Description

Historical loan data, encompassing 166,239 loans, was meticulously selected from various branches to meet the research criteria. The dataset's features were categorized into four main divisions:

1. Target:
   Loan Status: Description of the loan outcome: defaulted or not, (Loan is Defaulted=1/ Loan is not Defaulted=0)
2. Credit Determinants:
   Loan Amount: The total amount of money borrowed by the borrower. Higher loan amounts may increase the default risk, as they represent larger obligations (500-2500000)
   Customer Kind: Specifies whether the client is an individual person or a company (corporate entity), (is Corporate=1/ is Retail=0)
3. Loan Repayment Considerations
   Tenure: The duration of the loan, typically expressed in months. Longer loan terms may be associated with higher default rates (1-120).

Instalments Number: The number of instalments for loan repayment based on the loan term (1-87)

Instalments Value: The value of each instalment to be paid (12-18000)

Pay Period: Payment periodicity, e.g., annual, semi-annual, quarterly, or monthly (1-126).

Allow Month: Indicates if there is a grace period for instalment payments (0-24)

4. Customer Attributes:

   Age: The age of the borrower, as younger individuals may have less stable financial situations and may be at higher risk of default.

   Gender: Gender of the borrower (male=1 / female=2)

   Minor: Indicates whether the customer is a minor (under 18 years old) or an adult (18 years old and above), (is Minor=1/ is not Minor=0)

   Job: Occupation or job_ of the borrower. (1-50): Employee who works in government sector jobs, (51-100): Employees who are working in private sector jobs, and (101-152): Farmer who works in agricultural sector activities.

   Staff: Indicates whether the borrower is an employee at the bank or not, (is Staff=1/ is not Staff=0)

4.3 Data Preprocessing

In pre-processing phase, the dataset undergoes preparation and cleansing. This process entails the identification and rectification (or elimination) of corrupt or inaccurate records within the dataset. Missing data are addressed by imputing them with either mean or median values. Additionally, all categorical data types are converted to numeric form. When not all features influence the outcome of a class, we employ feature selection techniques to identify and retain only the relevant features. This approach enhances the model's interpretability, simplifies its complexity, and shortens its training duration. In the data pre-processing phase, all categorical variables are transformed into numerical formats via encoding, ensuring the dataset's compatibility with the scikit-learn library, for feature selection, the Select K Best class utilizes the chi-squared test to identify crucial features impacting loan approval decisions, optimizing the dataset for predictive modelling.

4.4 Splitting the Data

In the process of data splitting, 80% of the dataset was allocated for training purposes, while the remaining 20% was reserved for testing. The predictors were encapsulated within the array 'x', and the target variable was stored separately in the array 'y'.

4.5 Training ML Models

In this study, ML models were trained using a variety of classification algorithms, specifically Artificial Neural Networks (ANN), Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), and XGBoost (XGB) which were described in section 3. For each algorithm, an object of the respective class was instantiated, and its classification function was utilized to build the model. The models were then fitted with parameters, allowing them to learn from the training data. This process was repeated five times, with predictions made on the testing dataset using the predict function of each model, and the results were compared accordingly.

4.6 Evaluation Metrics

In this study, our assessment of the model's effectiveness primarily hinges on several key metrics: the F1 score, precision, accuracy, recall, as well as the AUC and ROC for evaluation purposes. Initially, we categorize true positives (TP) as the count of positive instances accurately identified by our classification model, false negatives (FN) as positive instances mistakenly labelled as negative by the model, false positives (FP) as negative instances wrongly classified as positive, and true negatives (TN) as negative instances correctly identified. The AUC represents the area beneath the ROC curve, which serves as a prevalent benchmark for binary classification models. The ROC curve visually represents the performance of the classifier, indicating its efficacy in each experiment. However, since the ROC curve alone does not offer a quantitative assessment of the classifier's performance, the AUC value, or the area under the ROC curve, is widely utilized to gauge the model's overall effectiveness, providing a comprehensive measure of its performance (Marzban, 2004). The Evaluation metrics are described as follows:

- Accuracy

Accuracy is defined as the proportion of the number of samples correctly classified by the classifier to the total number of samples for a given test data set (Anand et al., 2022). It can be calculated with the following equation:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \qquad (4)$$

- Precision

Precision is calculated as the ratio of true positive instances out of the total number of instances predicted as positive (Anand et al., 2022).

$$Precision = \frac{TRP}{TRP + FLP} \qquad (5)$$

- Recall

Recall is defined as the proportion of relevant instances that are successfully retrieved (Dasari et al., 2023).

$$Recall = \frac{TRP}{TRP + FLN} \qquad (6)$$

- F1 score

F1 score is a metric that combines both precision and recall calculating their average value (Dasari et al., 2023)

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} \qquad (7)$$

5. Results And Discussion

Following data pre-processing and in accordance with the chosen methodology, 80% of the 166239 collected loans were designated as the training dataset, while the remaining 20% was retained as the test dataset. The study's findings are subsequently presented below.

5.1 Features Analysis

The description of the dataset used in this study is illustrated in Table 1. In figure 2 the loan Default status is represented, value 0 represents the number of not defaulted loans while 1 represents the number of defaulted loans, the figure shows that most involved samples are “not defaulted” loans (70.6%) (118127 Loan) than the “defaulted” loans (29.4%) (48112 Loan).

Table 1: The description of the dataset.

	Loan_ status	Loan_ AMT	Cus_ Kind	Tenure	PAY_ PER	Installme nts_ No	Installme nt_ value	Allow_ Month	Age	Staff	Gender	JOB	Minor
Count	166239	166239	166239	166239	166239	166239	166239	166239	166239	166239	166239	166239	166239
Mean	0.29	8391.13	0	24.41	16.58	2.25	6053.1	0.1	50.53	0.02	1.04	119.26	0
Std	0.45	36775.2	0.07	23.16	4.3	6.65	33837.99	1.09	14.61	0.13	0.19	29.52	0.01
Min	0	500	0	1	1	1	12	0	16	0	1	1	0
25%	0	2300	0	18	18	1	1994	0	39	0	1	127	0
50%	0	4500	0	18	18	1	3750	0	50	0	1	127	0
75%	1	8600	0	18	18	1	7000	0	61	0	1	127	0
Max	1	2500000	1	120	126	87	18000	24	82	1	2	152	1

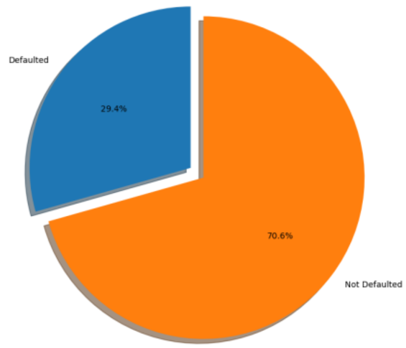

Figure 2. Loan Status distribution (Defaulted/Not Defaulted).

In the following figures, we provide a breakdown of the factors distribution based on the loan status (defaulted and not defaulted). The mean age of participants is 50.53 years, and its standard deviation is 14.61. In figure 3, we show the age distribution of customers, and figure 4 represents age distribution in terms of the loan status. We see that most of the borrowers with age between 21 and 30 are defaulted (85%) while 60% of the age between 30 and 40 are borrowers with defaulted.

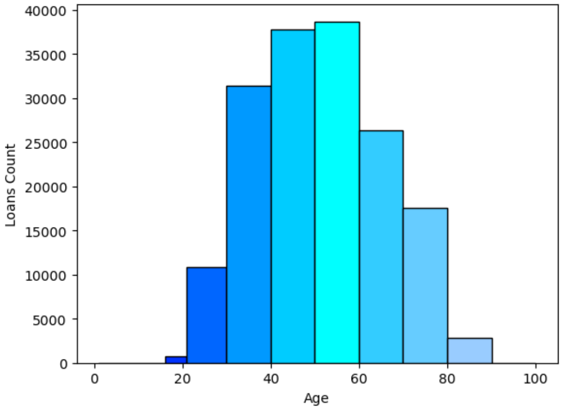

Figure 3. Age distribution of customers

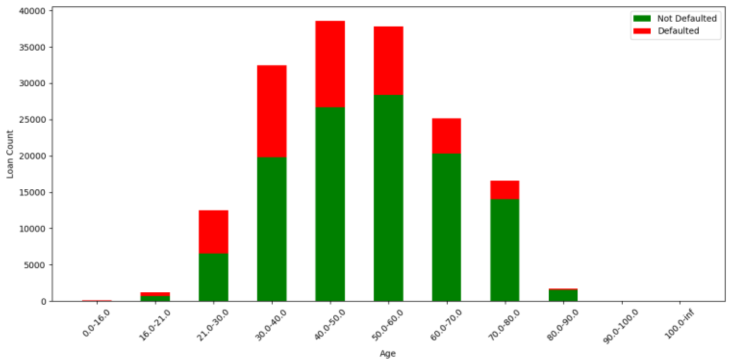

Figure 4. Age distribution based on loan status.

Meanwhile, figure 5 depicts the influence of gender and staff status on default status. In terms of the distribution of gender, 6240 loans attributed to females and 159999 loans to males. Among the loans taken by females, 5,824 defaulted. In contrast, out of the 159999 loans taken by males, 42288 defaulted, leaving 117711 non-defaulted. The distribution concerning staff (yes/no), consists of 2756 loans are belonging to staff and 163483 loans are not. Only 48 staff loans were defaulted, while 2708 were not. Conversely, among the 163483 loans not associated with staff, 48064 defaulted, leaving 115419 non-defaulted.

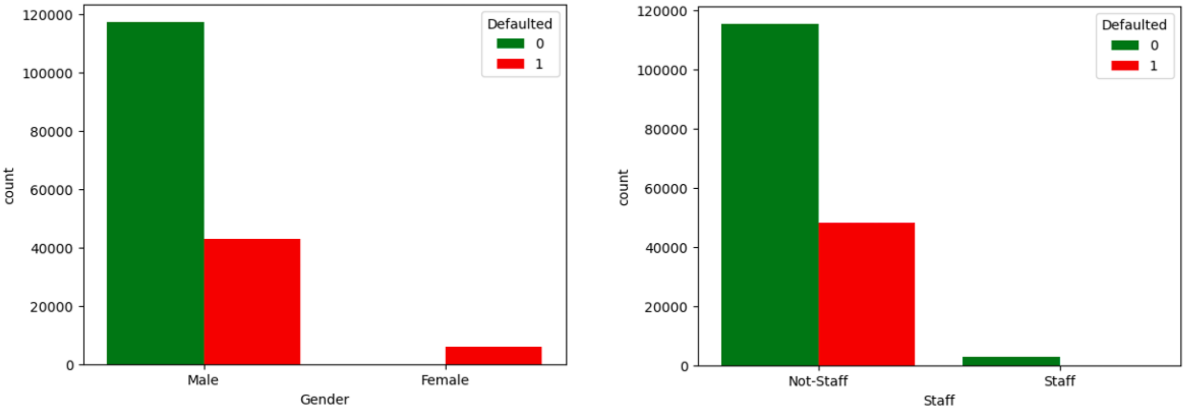

Figure 5. Gender and Staff distribution based on loan status.

Figure 6 illustrates the impact of loan amount on default status. Most of the loan amount is focused on loans ranging from 500 to 10000, which is typical for agricultural loans. Farmers borrow money to develop their agricultural land and repay the loan after selling their crop. Over 70% of the range did not default, while the rest defaulted.

About the Customer Kind (Corporate/Retail) distribution, 713 are corporate loans and 165526 are retail loans. Figure 7 illustrates the impact of Customer Kind (Corporate/Retail) on default status, In the case of corporate loans, 22 were defaulted while 691 were not defaulted. Conversely, among 165526 retail loans, 48090 were defaulted while 117436 were not defaulted.

Figure 8 details the distribution of Instalment Values, and Figure 9 reveals the effect of Instalment Values on loan default status, with the same default and not-default statistics. The mean value of instalments is 6053.1 EGP, and its standard deviation is 33837.99.

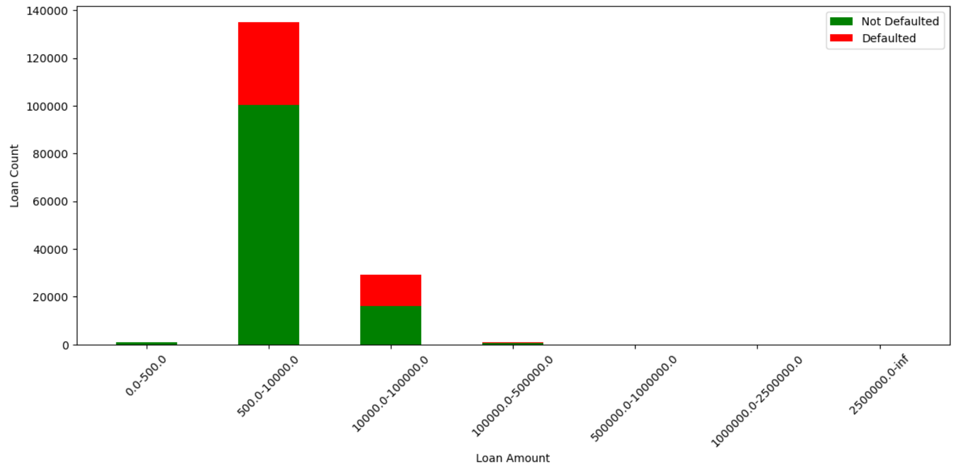

Figure 6. Loan amount distribution based on loan status.

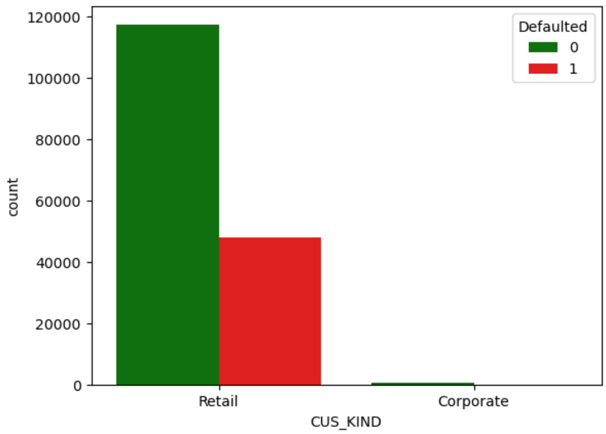

Figure 7. Customer Kind distribution based on loan status.

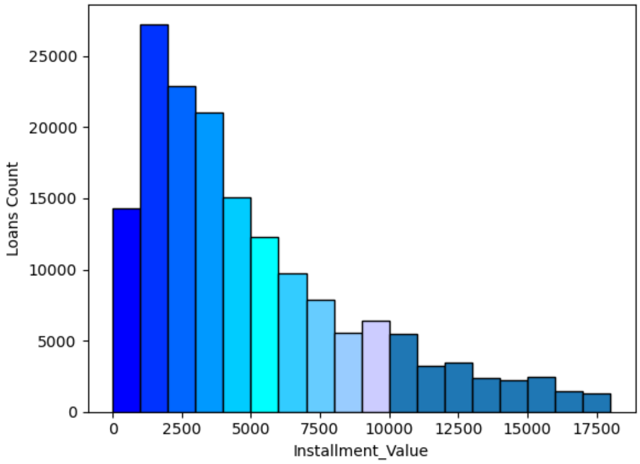

Figure 8. Instalment Value distribution.

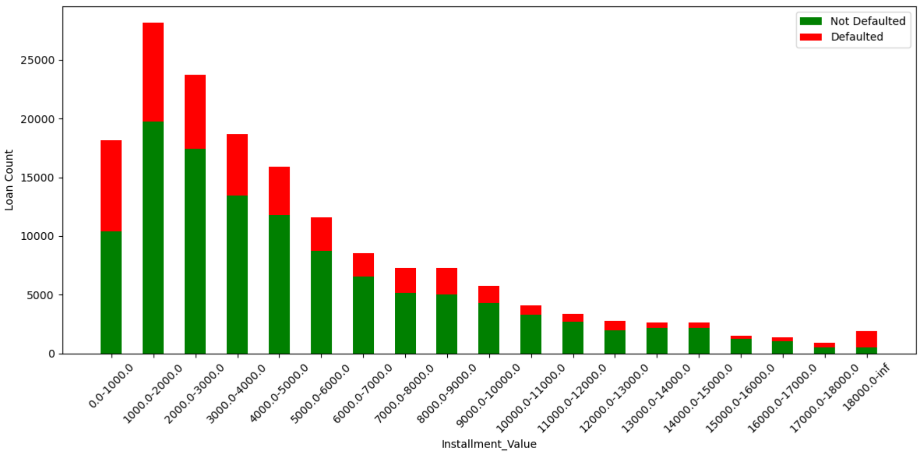

Figure 9. Instalment Value distribution based on loan status.

5.2 ML Models Evaluation

We utilized five different machine learning classification algorithms to predict Loan status. The metrices that have been used to evaluate the ML models help in understanding the performance of each model in different aspects like accuracy, precision, recall, and overall effectiveness in classification tasks. Table 2 presents the confusion matrix for each model.

Table 2. Confusion matrix for the machine learning classification models.

	ANN	RF	LR	DT	XGB
TP	7728	7877	7623	7895	7810
TN	23284	23059	23277	22961	23303
FP	25	250	32	348	6
FN	2211	2062	2316	2044	2129

True Positives (TP) represents the number of instances correctly classified as positive (Defaulted) by the model, True Negatives (TN): represents the number of instances correctly classified as negative (Not Defaulted) by the model, False Positives (FP) represents the number of instances incorrectly classified as positive (Defaulted) by the model, and False Negatives (FN) represents the number of instances incorrectly classified as negative (Not Defaulted) by the model.

Based on the information provided in Table IV, the ANN model exhibits performance characteristics as follows: It accurately predicts 7728 instances of defaulted (TP) and 23284 instances of Not Defaulted (TN), totalling 31012 correct classifications out of 33248 test samples. However, it misclassifies 25 instances as defaulted (FP) and 2211 instances as Not Defaulted (FN), resulting in a total of 2236 misclassifications out of 33248 test samples. The Decision Tree (DT) model identifies 7895 loans classified correctly as defaulted (TP) and 22961 loans classified correctly as Not Defaulted (TN), totalling 30586 accurately classified loans out of 33248 test samples. However, it misclassifies 348 loans as default (FP) and 2044 loans as Not Defaulted (FN), resulting in a total of 2392 misclassified loans out of 33248 test samples. The Random Forest (RF) model exhibits TP = 7877, TN = 23059, FP = 250, and FN = 2062. Demonstrating accuracy in accurately predicting true positives, it successfully identifies 7877 loans classified correctly as defaulted (TP) and 23059 loans classified correctly as Not Defaulted (TN), amounting to 30936 accurately classified loans out of 33248 test samples. It also identifies 250 loans misclassified as defaulted (FP) and 2062 loans misclassified as Not Defaulted (FN), resulting in a total of 2312 misclassified loans out of 33248 test samples. The XGB model accurately identifies 7810 loans classified as defaulted (TP) and 23303 loans classified as Not Defaulted (TN), resulting in a total of 31113 correctly classified loans out of 33248 test samples. However, it incorrectly classifies 6 loans as defaulted (FP) and 2129 loans as Not Defaulted (FN), leading to a total of 2135 misclassified loans out of 33248 test samples.

Finally, the Logistic Regression (LR) model correctly predicted true positives, it successfully identifies 7623 loans classified as defaulted (TP) and 23277 loans classified as Not Defaulted (TN), totalling 30900 correctly classified loans out of 33248 test samples. It also identifies 32 loans misclassified as defaulted (FP) and 2316 loans misclassified as Not Defaulted (FN), resulting in a total of 2348 misclassified loans out of 33248 test samples. Figure 10 provides a summary and comparison of confusion matrix result for the ML models.

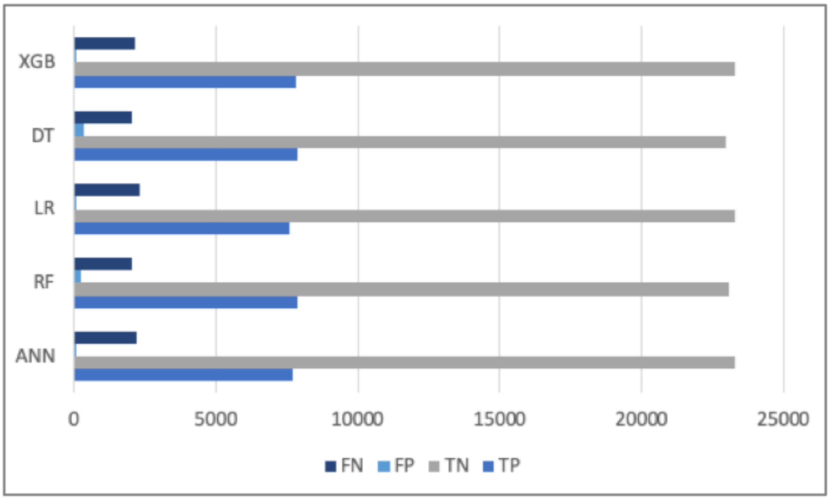

Figure 10. Comparison of confusion matrix result for the ML models.

Figure 11 shows the analysis of loss evolution during the training of the ANN, where epochs (iterations) are depicted on the x-axis and loss values (mean square error) on the y-axis. The behaviour of loss throughout training was observed, providing insights into the optimization process. Trends in loss evolution were monitored to gauge the model's learning progress and convergence.

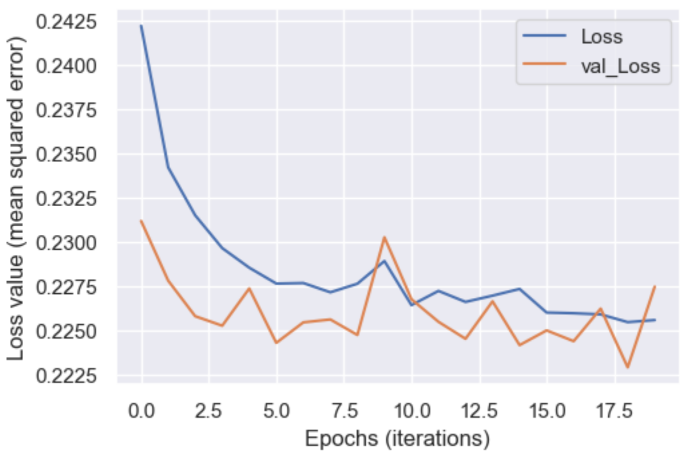

Figure 11. Loss evolution during training ANN model.

Figure 12 depicts the evolution of the AUC score during the training of the ANN model, with epochs (iterations) represented on the x-axis and AUC scores on the y-axis. The analysis of AUC score evolution provides insights into the model's performance dynamics over the course of training. Monitoring the changes in AUC score throughout training allows for the assessment of the model's ability to discriminate between classes and improve predictive accuracy.

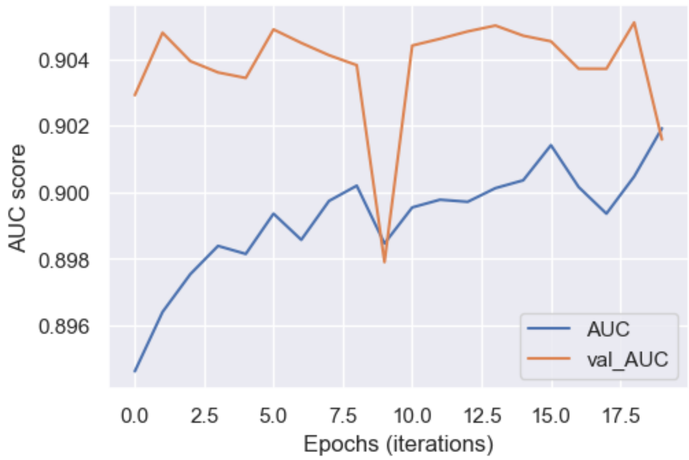

Figure 12. The evolution of the AUC score during the training of the ANN model.

Figure 13 shows that the performance of five models, namely Artificial Neural Networks (ANN), Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), and XGBoost (XGB), was evaluated based on their respective Area Under the ROC curve. AUC values were calculated to assess the probability that a randomly chosen positive instance would have a higher predicted probability than a randomly chosen negative instance. The results indicate that ANN achieved the highest AUC score of 0.9062, followed closely by LR with an

AUC of 0.8962. RF, DT, and XGB also performed well, with AUC values of 0.8909, 0.8897, and 0.8928, respectively. Overall, ANN demonstrated the best performance among the evaluated models based on AUC scores.

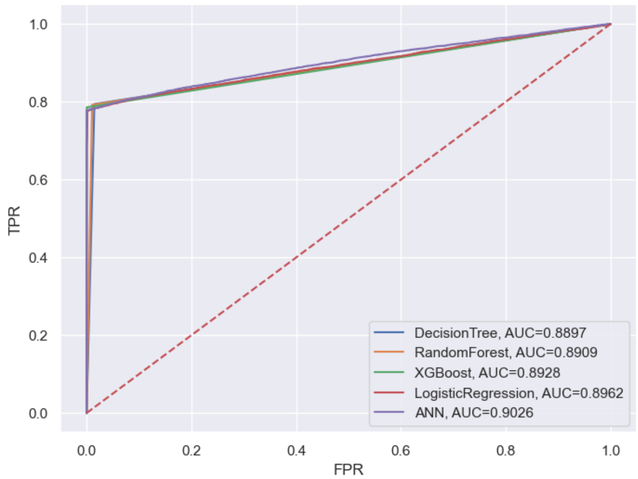

Figure 13. ROC curve and AUC score of DT, RF, XGB, LR, and ANN.

The study's effectiveness was assessed by measuring accuracy, precision, recall, and F1 score. The results of these performance metrics can be found in Figure 14 and Table 3.

Table 3. Performance evaluation of the five machine learning algorithms

	ANN	RF	LR	DT	XGB
Accuracy	0.932748	0.930462	0.929379	0.928056	0.935786
Precision	0.996775	0.969238	0.99582	0.957782	0.999232
Recall	0.777543	0.792534	0.766979	0.794346	0.785793
F1 score	0.873615	0.872025	0.866545	0.868441	0.879752

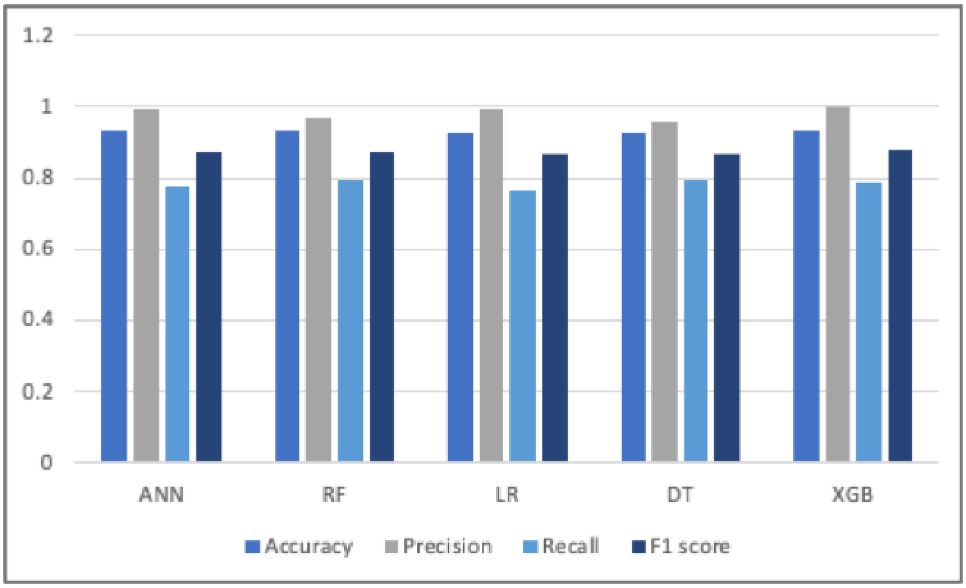

Figure 14. Performance of DT, RF, XGB, LR, and ANN.

The findings demonstrating that XGB surpassed all other machine learning algorithms in terms of accuracy, achieving an accuracy rate of 93.57%. ANN also performed well, securing the second-highest accuracy at 93.27%. After analysing the performance of each model based on accuracy, precision, recall, and F1 score, distinct characteristics emerge that define their suitability for various classification tasks. The Artificial Neural Network (ANN) demonstrates exceptional accuracy and precision, albeit with a slightly lower recall, indicating its proficiency in correctly classifying instances, although it may occasionally miss positive instances. Random Forest (RF) exhibits balanced precision and recall, coupled with a commendable accuracy rate, making it a reliable choice for classification tasks requiring robust performance. Logistic Regression (LR) showcases high precision but relatively lower recall, yet maintains a strong accuracy rate, suggesting its effectiveness in scenarios prioritizing precise predictions. Decision Tree (DT) demonstrates balanced precision and recall, offering a good trade-off between accuracy and reliability in classification tasks. Extreme Gradient Boosting (XGB) achieves high precision and accuracy, with a balanced F1 score, making it a versatile model suitable for a wide range of classification tasks. Ultimately, the choice among these models depends on the specific requirements and trade-offs between precision, recall, and overall accuracy desired for the given task.

6. Conclusion

This study focused on assessing loan default risks within the banking sector by leveraging machine learning algorithms. The selection of an appropriate algorithm is pivotal in managing loan decisions, aiding in the identification of potential client loan defaults. By accurately predicting the likelihood of loan defaults, banks can reduce the risk of financial losses when approving loans. To address this, the paper introduced, examined, and applied five machine learning classification algorithms: Artificial Neural Networks (ANN), Random Forest (RF), Logistic Regression (LR), Decision Tree (DT) and XGBoost (XGB). The research outperformed previous efforts in terms of accuracy, with the highest results being 93.57% for XGBoost (XGB), followed closely by Artificial Neural Networks (ANN) at 93.27%. The Decision Tree (DT) scored the lowest at 92.80%. The main limitation of this study stems from the unavailability of important features in the database. The findings suggest that machine learning methods are effective in predicting loan default with a high degree of accuracy. The proposed model offers potential benefits to financial and banking institutions by enhancing their ability to make informed credit decisions, mitigate loan defaults, and safeguard the financial resources of banks. These findings yield valuable insights into the behavior of loan customers, particularly in the identification of potential defaulters. Experts in the field endorse the generated rules for both defaulting and non-defaulting contracts. In summary, this research suggests that the banking sector can leverage data mining applications, demonstrating the effectiveness of machine learning in comprehending loan customer behavior, particularly in predicting patterns of loan defaults. For future research, it is advisable to expand the scope by collecting more diverse and extensive real-world data and exploring the application of various machine learning algorithms to achieve improved performance.

References

- Abdullah, M., Chowdhury, M. A. F., Uddin, A., & Moudud-Ul-Huq, S. (2023). Forecasting nonperforming loans using machine learning. Journal of Forecasting.
- Alsaleem, Maan Y., and Safwan O. Hasoon. "Predicting Bank Loan Risks Using Machine Learning Algorithms." AL-Rafidain Journal of Computer Sciences and Mathematics 14.1 (2020): 149-158.
- Aslam, U., Tariq Aziz, H. I., Sohail, A., & Batcha, N. K. (2019). An empirical study on loan default prediction models. Journal of Computational and Theoretical Nanoscience, 16(8), 3483-3488.
- Anand, M., Velu, A., & Whig, P. (2022). Prediction of loan behaviour with machine learning models for secure banking. Journal of Computer Science and Engineering (JCSE), 3(1), 1-13.
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

- Dasari, Y., Rishitha, K., & Gandhi, O. (2023). Prediction of bank loan status using machine learning algorithms. International Journal of Computing and Digital Systems, 14(1), 1-1.
- Dosalwar, S., Kinkar, K., Sannat, R., & Pise, N. (2021). Analysis of loan availability using machine learning techniques. International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), 9(1), 15-20.
- Kalyani, S., & Gupta, N. (2023). Is artificial intelligence and machine learning changing the ways of banking: a systematic literature review and meta analysis. Discover Artificial Intelligence, 3(1), 41.
- Krenker, A., Bešter, J., & Kos, A. (2011). Introduction to the artificial neural networks. Artificial Neural Networks: Methodological Advances and Biomedical Applications. InTech, 1-18.
- Lai, L. (2020, August). Loan default prediction with machine learning techniques. In 2020 International Conference on Computer Communication and Network Security (CCNS) (pp. 5-9). IEEE.
- Madaan, M., Kumar, A., Keshri, C., Jain, R., & Nagrath, P. (2021). Loan default prediction using decision trees and random forest: A comparative study. In IOP Conference Series: Materials Science and Engineering (Vol. 1022, No. 1, p. 012042). IOP Publishing.
- Marzban, C. (2004). The ROC curve and the area under it as performance measures. Weather and Forecasting, 19(6), 1106-1114.
- Moradi, S., & Mokhatab Rafiei, F. (2019). A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks. Financial Innovation, 5(1), 1-27.
- Moscatelli, M., Parlapiano, F., Narizzano, S., & Viggiano, G. (2020). Corporate default forecasting with machine learning. Expert Systems with Applications, 161, 113567.
- Rath, G. B., Das, D., & Acharya, B. (2021). Modern approach for loan sanctioning in banks using machine learning. In Advances in Machine Learning and Computational Intelligence: Proceedings of ICMLCI 2019 (pp. 179-188). Springer Singapore.

- Sun, T., & Vasarhelyi, M. A. (2018). Predicting credit card delinquencies: An application of deep neural networks. Intelligent Systems in Accounting, Finance and Management, 25(4), 174-189.
- Wang, Y., Wang, M., Pan, Y., & Chen, J. (2023). Joint loan risk prediction based on deep learning-optimized stacking model. Engineering Reports, e12748.
- Zhu, L., Qiu, D., Ergu, D., Ying, C., & Liu, K. (2019). A study on predicting loan default based on the random forest algorithm. Procedia Computer Science, 162, 503-513.