

Article

Credit Card Default Prediction: An Empirical Analysis on Predictive Performance Using Statistical and Machine Learning Methods

Rakshith Bhandary *  and Bidyut Kumar Ghosh

Department of Commerce, Manipal Academy of Higher Education, Manipal 567104, Karnataka, India; bidyut.ghosh@manipal.edu

* Correspondence: rakshith.bhandary@manipal.edu

Abstract: This article compares the predictive capabilities of six models, namely, linear discriminant analysis (LDA), logistic regression (LR), support vector machine (SVM), XGBoost, random forest (RF), and deep neural network (DNN), to predict the default behavior of credit card holders in Taiwan using data from the UCI machine learning database. The Python programming language was used for data analysis. Statistical methods were compared with machine learning algorithms using the confusion matrix measured in metric terms of prediction accuracy, sensitivity, specificity, precision, G-mean, F1 score, ROC, and AUC. The dataset contained 30,000 credit card users' information, with 6636 default observations and 23,364 nondefault cases. The study results found that modern machine learning methods outperformed traditional statistical methods in terms of predictive performance measured by the F1 score, G-mean, and AUC. Traditional methods like logistic regression were marginally better than linear discriminant analysis and support vector machines in terms of the predictive performance measured by the area under the receiver operating characteristic curve. In the modern machine learning methods, deep neural network was better in the predictive performance metrics when compared with XGBoost and random forest methods.



Academic Editor: Thanasis Stengos

Received: 18 November 2024

Revised: 22 December 2024

Accepted: 26 December 2024

Published: 9 January 2025

Citation: Bhandary, R., & Ghosh, B. K. (2025). Credit Card Default Prediction: An Empirical Analysis on Predictive Performance Using Statistical and Machine Learning Methods. *Journal of Risk and Financial Management*, 18(1), 23. <https://doi.org/10.3390/jrfm18010023>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: credit card default; confusion matrix; deep neural network; default prediction; linear discriminant analysis; logistic regression; machine learning; random forest; support vector machine; XGBoost

JEL Classification: C44; C45; C52; C55; C63

1. Introduction

Consumer credit in the U.S. rose by USD 8.93 billion in June 2024, following an upwardly revised USD 13.94 billion increase in the prior month and below market expectations of a USD 10 billion gain (Saraiva, 2024). Also, revolving credit, including credit cards, fell by nearly USD 1.7 billion, marking the largest drop since early 2021. Meanwhile, nonrevolving credit, such as loans for cars and education, climbed by USD 10.6 billion, the highest in 2024 (Saraiva, 2024). Americans have reduced their credit card debt and even reduced their credit card spending mainly because banks are carefully selecting the customers to whom they extend credit (Dicon, 2024). The credit card debt default crisis took place in several countries like South Korea in 2001 and Hong Kong in 2002, followed by Taiwan in the year 2005 (Chang, 2022). The same study highlighted adverse borrower selection and information asymmetry as the main reasons for default. The Taiwan cards and payments

market size was USD 139.5 billion in 2023, and the market is expected to grow at 8% until 2027 as per the [Globaldata \(2023\)](#) report. Taiwan has 58.12 million credit cards in circulation as of 2023, with 37.64 million active credit cards as per the 2024 press report released by the Financial Supervisory Commission of Republic of China (Taiwan).

Loan default prediction has always been a risk management strategy for financial institutions because a minor increase in the prediction accuracy results in better risk management strategies ([Hubbard, 2020](#)). Traditionally, statistical and econometric models have been used in prediction and decision-making processes. These prediction methods are not just for the purpose of assessing credit risk and scoring the individuals; they also have wider scope, from risk management strategies to optimizing the capital reserve requirements, forecasting the nonperforming assets and losses, and better recovery management strategies ([Gauthier et al., 2012](#)). Better customer profiling has always been the primary objective of banks to differentiate clients as prompt or irregular payers. Also, a minor improvement in the prediction accuracy of loan default will result in an increased profitability of the institution. In addition, early identification of the default potential of customers will help lending organizations prevent slippage into bad loans and encourage clients to repay by enforcing recovery management strategies.

Risk profiling of customers is an important aspect of credit risk management, and it is done based on the assessment of creditworthiness of the borrower that is measured in terms of the customers' ability and willingness to repay the loan ([Bhandary et al., 2023a](#)). Ability can be evaluated by income, age, number of dependents, marital status, and expenses, whereas willingness can be evaluated by attitudinal and psychological factors including the credit track record. The disadvantage associated with traditional credit scoring models in risk profiling is their inability to capture the details of people without a credit history ([DeVaney, 1999](#)). The past track record of the borrower is an important predictor of future payments to be made by the potential borrower ([Schreiner, 2000](#)). Hence, an analysis of the customers' probability of default based on the input parameters, model selection, and validation, with an objective to predict the default, will not only help the financial institutions with better risk management strategies and decision making but also help them to plan their loan portfolio and diversify accordingly.

The International Financial Reporting Standards (IFRS) 9 prescribes a structure to predict the expected credit loss (ECL) measured by the probability of default (PD), exposure at default (EAD), and loss given default (LGD). The PD is the major determinant of the ECL. It is measured by credit ratings based on historical repayment track records with default statistics ([Bandyopadhyay, 2023](#)). The LGD is defined as the estimated amount that is impossible to recover from an asset in the event of a default, equationally represented as one minus the recovery rate. Increased prediction accuracy of loss given default helps the banks accurately calculate the economic capital requirement of the bank ([Bastos, 2010](#)). EAD is the loan outstanding amount on which PD and LGD are applied. IFRS 9 classifies assets into three categories based on credit risk assessment. The three classes are assets with low credit risk, assets with high credit risk, and credit-impaired assets ([IASB, 2014](#)).

The biggest challenge in loan default prediction is the calculation of probability of default of the borrower ([Baesens et al., 2003](#); [Coşer et al., 2019](#); [Hand & Henley, 1997](#); [Yeh & Lien, 2009](#)). To overcome this problem, there is a need to evaluate models on the pretext of loan default prediction with an increased predictive performance for better approximation of the probability of default. The aim of this study was to analyze credit card users' data from the University of California Irvine (UCI) machine learning platform, which contained details of credit card clients with default and nondefault data. The study evaluated a series of statistical and machine learning algorithms using six predictive models that could explain the studied event through classifiers, such as linear discriminant analysis (LDA),

logistic regression (LR), support vector machines (SVMs), XGboost, random forest (RF), and deep neural networks (DNNs).

2. Literature Review

A study on predicting credit defaults among microfinance institution borrowers found age, income, marital status, gender, number of dependents, residential status, tenure, and the amount of loan borrowed as the significant determinants of loan repayment (Ofori et al., 2014). Loan repayment depends on ability and willingness of the borrower to repay. Salary determines the ability of the borrower, and it affects the repayment behavior of the borrower (Bhandary et al., 2023b). Willingness is measured by attitude toward repayment that influences repayment behavior (Bhandary et al., 2023a). Also, the attained level of education increases loan repayment performance (Acquah & Addo, 2011; Addisu, 2006). To summarize, age, income, expenses, education attainment level, marital status, gender, number of dependents, psychological factors, attitude, and credit history are the factors that affect loan repayment.

Traditionally, lenders would use their human intuition to judge the repayment behavior of the applicant. Statistical methods use input variables from the customers' loan application forms and other sources that correlate with the output variable (default case) to calculate the probability of default (Hand & Henley, 1997). The statistical models used include the Bayes classifier, linear discriminant analysis, logistic regression technique, and k-nearest neighbor-based prediction models. Linear discriminant analysis is a very common technique for classification problems and dimensionality reduction (Tharwat et al., 2017). In addition, the performance of linear discriminant analysis is usually compared with a logistic regression model, which is another most widely used method in binary data classification problems (Maalouf, 2011). The advantage of using the discriminant analysis and logistic regression technique is its simplicity in application for classification problems, but the disadvantage is that it cannot capture the interaction effect of input variables, and neither can it be applied for nonlinear relations (Yeh & Lien, 2009).

Support vector machines is another solution for default prediction problems that is an extremely robust and powerful method in classification and regression applications (Cervantes et al., 2020). Also, the gradient boosting algorithm was developed for high predictive performance but with a compromised speed of execution. Hence, extreme gradient boosting was introduced with a high speed of execution and good predictive ability for classification problems (Ramraj et al., 2016). The random forest classifier is an ensemble approach that combines several classifiers. This combination provides high accuracy and is a superior technique for class-imbalanced datasets (More & Rana, 2017). The interaction effect is confirmed when the combined effect of two or more independent variables/features produces a significantly higher effect on the dependent variable when compared with the addition of their individual contributions. Neural network models have the unique ability to capture the presence of interactions in nonlinear relationships as mentioned in the findings of Shao et al. (2023). In addition, the deep neural network is a fast-growing machine learning method because of its superior performance (Sze et al., 2017).

The logistic regression credit scoring model was compared with the neural network model for predicting credit defaults in the Indian microfinance industry; the results showed that the neural network model outperformed the logistic regression model in calculating the prediction accuracy, and the major determinants of probability of default for the model included the income of the borrower, the quantum of the requested loan amount, the total expense, age, family size, and the length of stay at current residence (Viswanathan & Shanthi, 2017). A text mining technique from the loan application forms of an online crowdfunding platform in the U.S. was analyzed using machine learning to predict po-

tential loan defaults, and the findings of the study included patterns of words written by defaulting borrowers indicative of their personality traits and personality states, which seem to disclose their true nature, grounded in human behavior (Netzer et al., 2019).

Hamdi et al. (2024) compared the predictive performance of six models, namely, linear discriminant analysis, logistic regression, decision trees, support vector machines, random forest, and deep neural networks, for the bankruptcy prediction of Tunisian companies and found that the deep neural network performed with better accuracy, F1 score, and area under the curve in comparison with conventional models. A study on predicting the banking crisis in India using statistical methods like logistic regression; artificial intelligence; and machine learning methods like random forest, naïve Bayes, gradient boosting, support vector machines, neural networks, K-nearest neighbors, and decision trees found that neural networks and random forest models were effective models in banking crisis prediction (Puli et al., 2024).

A study on financial risk assessment using big data analysis and ten algorithms found that ensemble models using boosting algorithms outperformed traditional models like logistic regression and decision trees in terms of prediction accuracy (Suhadolnik et al., 2023). The study on predicting financial inclusion in Peru using machine learning methods like decision trees, random forests, artificial neural networks, XGBoost, and support vector machines found that these methods could be a valuable complement to standard models like generalized linear models and logistic regression models for assessing financial inclusion in Peru (Maehara et al., 2024). The study also found that neural network was the most effective method to predict account access. For account usage prediction, the random forest method and support vector machines employing the radial basis function were the most effective prediction methods.

Traditional methods like discriminant analysis and logistic regressions can capture linear relationships only, and this is a major limitation for big data analysis, whereas machine learning models can measure the interaction effect among explanatory variables, as well as predicting complex linear and nonlinear relationships (Chiang et al., 2006). Decision trees, support vector machines, and deep neural networks are more effective to capture complex nonlinear relationships than simple linear models (Varian, 2014). Also, relying exclusively on statistical models for decision making may lead to questionable findings. Hence, it is recommended to use a wide range of tools for reaching conclusions on data (Breiman, 2001b).

Machine learning models have emerged as a significant tool in default prediction methods mainly because they are scalable and agile. These models are widely used in decision making for classification, regression, and clustering problems. The predictive performances of these models are analyzed using the confusion matrix and the receiver operating characteristic curve (ROC). A confusion matrix classifies the actual values with the values predicted by the model (Luque et al., 2019). The ROC graphically characterizes the positive and negative events exploring the trade-offs between competing losses at different threshold levels (Brown & Davis, 2006). It is free from parametric assumptions and the value is independent from the ratio of positive and negative events. The area under the ROC (AUC) is also independent of the decision thresholds and hence suitable for classification problems in addition to confusion matrix analysis in machine learning applications. Machine learning models have been applied in a variety of fields like predicting asset pricing (S. Gu et al., 2020), option return predictions (Bali et al., 2023), stock market asset pricing (Drobotz & Otto, 2021), bankruptcy prediction (Hamdi et al., 2024), predicting banking crises (Puli et al., 2024), credit risk assessment (Suhadolnik et al., 2023), and predicting financial inclusion (Maehara et al., 2024).

Credit default prediction is not just for calculating the credit score; it has a wide range of applications in terms of devising credit risk management strategies, calculating the capital reserve requirements, loss forecasting, and provisioning purposes. Recent studies highlight a broad range of machine learning applications including credit risk assessment, and this study aimed to predict the credit card default risk with the following research questions:

- RQ1. Which model has the best predictive performance as per the confusion matrix?
- RQ2. Which model has the best area under the receiver operating characteristics?
- RQ3. How can the default score be calculated?

3. Materials and Methods

3.1. Linear Discriminant Analysis

Linear discriminant analysis (LDA) is a statistical technique to find a linear combination of features that best separates two or more classes of events. The primary goal of LDA is to project the data onto a lower-dimensional space with good class separability. The equation for linear discriminant analysis (LDA) introduced by Ronald Fisher is central to dimensionality reduction and classification tasks (Fisher, 1936). The equation is given below as

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (1)$$

where β_n is the weight associated with the input variable x_n .

3.2. Logistic Regression

Logistic regression (LR) is a tool for classifying binary data and making predictions between zero and one. Logistic regression is used to explain the relationship between a binary dependent variable and independent variables. Logistic regression is used to obtain the odds ratio in the presence of more than one explanatory variable (Sperandei, 2014). It generates the coefficients and standard errors for the significance levels of the formula used to predict the probability of event occurrence. It is represented by the equation

$$P = \frac{1}{1 + e^{-(a+bx)}} \quad (2)$$

where P is the probability, and a and b are the parameters of the model.

3.3. Support Vector Machine

Support vector machine (SVM) is a supervised learning model used for classification problems developed by Vapnik (1998). This technique determines the best separating hyperplane between two classes of a dataset. The mathematical formulation is expressed in terms of optimization problem and decision function (Hearst et al., 1998). The optimization problem is written as

$$\text{minimize } \frac{1}{n} \sum_{i=1}^n \zeta_i + \lambda \|w\|^2 \quad (3)$$

Subject to $y_i(w^t x_i - b) \geq 1 - \zeta_i$ and $\zeta_i \geq 0$ for all i , where w^t is the normal vector to a hyperplane, where the y_i values are either 1 or -1 , each indicating the class to which the point x_i belongs. Each x_i is a p -dimensional real vector. The final classifier is given as

$$f(x) = \text{sgn}(w^t x - b) \quad (4)$$

where sgn is the sign function.

3.4. XGBoost

XGBoost is a highly scalable tree boosting machine learning algorithm. It can be scaled for big data with far fewer resources compared with the existing systems (Chen & Guestrin, 2016). In extreme gradient boosting (XGBoost), the objective function (loss function and regularization) \mathcal{L}^t at iteration t that we need to minimize is given by the following equation:

$$\mathcal{L}^t = \sum_{i=1}^n l\left(y_i, \check{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (5)$$

where y_i is the real label from the training dataset, and l is the function of the sum of current and previous additive trees.

3.5. Random Forest Classifier

Random forests are a combination of tree predictors proposed by Breiman (2001a). When performing random forest (RF) based on classification data, we can use the Gini index or entropy to decide how the nodes on a decision tree branch. The Gini formula is given by

$$Gini = 1 - \sum_{i=1}^c (p_i)^2 \quad (6)$$

This formula uses the class and probability to determine the Gini of each branch on a node, determining which of the branches is more likely to occur. Here, p_i represents the relative frequency of the class that is observed in the dataset, and c represents the number of classes.

The formula for entropy is given by

$$Entropy = \sum_{i=1}^c -p_i * \log_2(p_i) \quad (7)$$

Entropy uses the probability of a certain outcome to decide on how the node should branch. Unlike the Gini index, it is mathematically intensive due to the logarithmic function used in the calculation.

3.6. Deep Neural Network

Deep Neural Network Architecture

A deep neural network (DNN), as shown in Figure 1, is a type of artificial neural network (ANN) that has multiple hidden layers (usually more than one) between the input and output layers. Artificial neural networks (ANN), first proposed by McCulloch and Pitts (1943), are broadly classified into two types based on the direction of information flow between the input and output layers. The two types are feedforward neural network (uni-directional flow of information between the layers from input to output) and backpropagation neural network (bi-directional flow of information between the layers) (Krenker et al., 2011). Feedforward neural networks are simpler to design and test, and multilayer perceptron and radial basis functions are the most popular feedforward neural networks. Multilayer perceptrons (MLPs) are limited to performing linear functions only. Hence, neurons are used for nonlinear transformations. Backpropagation models are advanced, complex, and accurate for increased prediction accuracies since they iterate based on the given criteria and adjust the weights and bias accordingly (Krenker et al., 2011). The weight and bias adjustment with activation function is displayed in Figure 2. The activation functions used are tangent function, logistic function, hyperbolic tangent, rectifier function, softplus function, radial basis function, rectified linear unit (ReLU), and leaky ReLU. ReLU is usually used in the hidden layers as activation functions in deep learning (DL) models

to address the vanishing gradient problem. The network learns through an optimization algorithm known as the gradient descent. This algorithm compares the predicted output with the actual output and tunes the parameters (weights and bias) of the network through an optimization technique known as the learning rate (Hochreiter et al., 2001). Other learning functions include stochastic gradient descent and conjugate gradient descent, among many others.

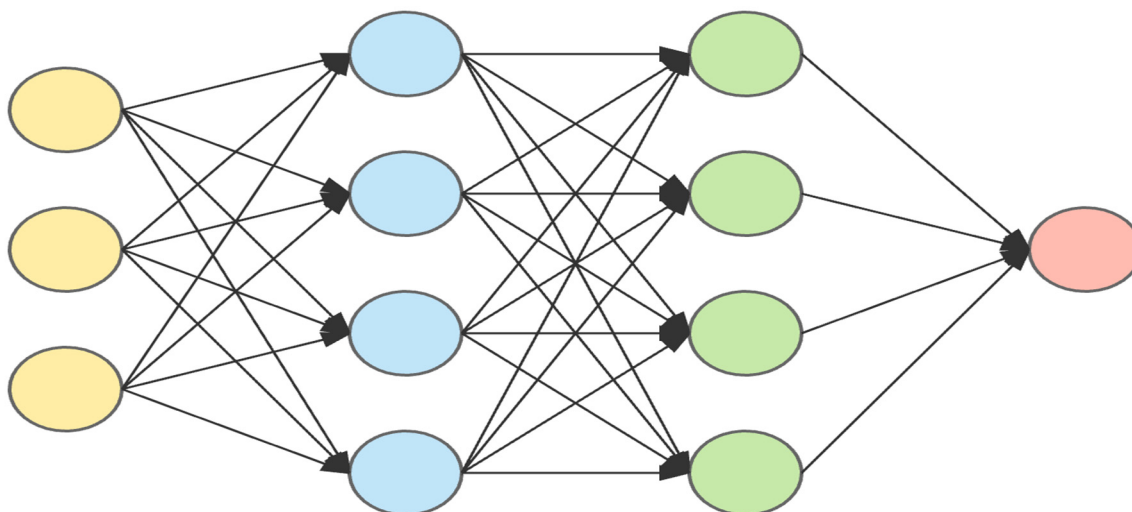


Figure 1. The deep neural network architecture. Source: Authors' own.

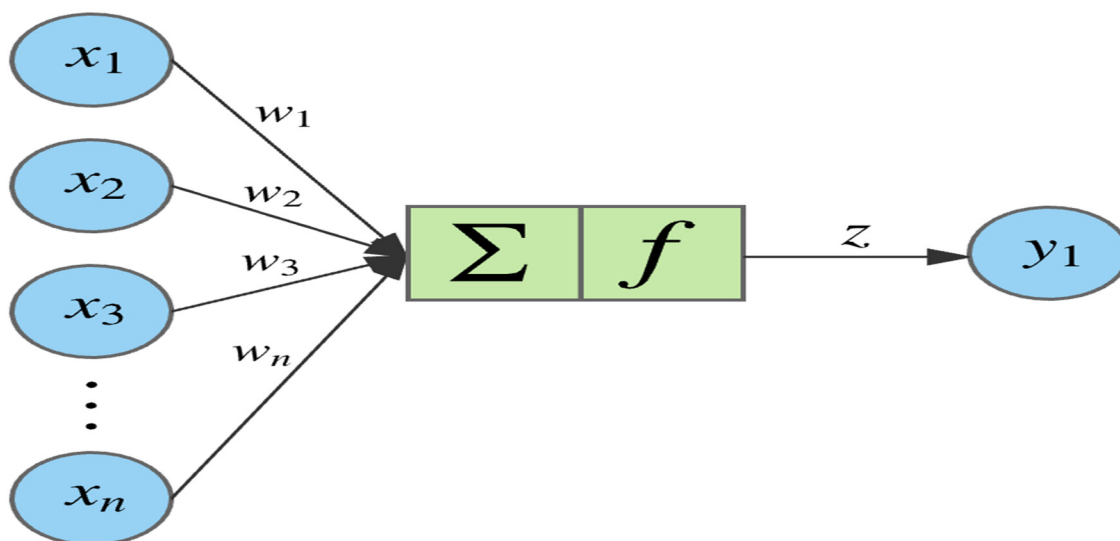


Figure 2. Weight and bias adjustment with activation function. Source: Authors' own.

The z value is calculated as given below:

$$z = f(b + x \cdot w) = f\left(b + \sum_{i=1}^n x_i w_i\right) \tag{8}$$

$$x \in d_{1 \times n}, w \in d_{n \times 1}, b \in d_{1 \times 1}, z \in d_{1 \times 1} \tag{9}$$

where b is the bias, w is the weight corresponding to input x , and n is the number of neurons.

3.7. Methodology

The dataset contains the following attributes as displayed in Table 1. The ID column was removed as it would not add value to analyze the data. Each of the input variables selected was correlated with an output variable (Default_status) and shortlisted if the values were greater than 0.2. The input variables shortlisted include the following: the sanctioned credit limit, coded as Limit_bal; the gender of the borrower, coded as Sex; the schooling of the borrower, coded as Education; the marital status and age of the borrower, coded as the same; and the repayment status, coded as Def_pay. The monthly billed amount was coded as Bill_amt1 for the month of April, Bill_amt2 for the month of May, Bill_amt3 for the month of June, Bill_amt4 for the month of July, Bill_amt5 for the month of August, and Bill_amt6 for the month of September. The completed monthly payment was coded as Pay_amt1 for the month of April, Pay_amt2 for the month of May, Pay_amt3 for the month of June, Pay_amt4 for the month of July, Pay_amt5 for the month of August, and Pay_amt6 for the month of September. The output variable was coded as Default_status (0 for no default and 1 for the presence of default). The income of the borrower was approximated by the credit limit provided by the issuer of the card since high-income customers are provided with a higher credit limit than low-income customers, who are provided with lower credit limits.

Table 1. Attributes of the dataset with feature code and description.

Feature ID	Feature Code	Description
X1	Limit_bal	The amount of credit that the card holder is entitled to avail. It includes individual and family credit.
X2	Sex	(Gender) 1 = male, 2 = female
X3	Education	1 = graduate, 2 = university, 3 = high school, 4 = others
X4	Marital status	1 = married, 2 = single, 3 = others
X5	Age	21 years to 79 years
X6 to X11	Repayment status codes – 1 = paid duly 1 = payment delay for one month 2 = payment delay for two months 9 = payment delay for 9 months and above	History of past payment month-wise X6 = repayment status for the month of September 2005 X7 = repayment status for the month of August 2005 X8 = repayment status for the month of July 2005 X9 = repayment status for the month of June 2005 X10 = repayment status for the month of May 2005 X11 = repayment status for the month of April 2005
X12 to X17	Amount of bill statement	X12 = amount of bill statement for September 2005 X13 = amount of bill statement for August 2005 X14 = amount of bill statement for July 2005 X15 = amount of bill statement for June 2005 X16 = amount of bill statement for May 2005 X17 = amount of bill statement for April 2005
X18 to X23	Amount of previous payment	X18 = amount paid in September 2005 X19 = amount paid in August 2005 X20 = amount paid in July 2005 X21 = amount paid in June 2005 X22 = amount paid in May 2005 X23 = amount paid in April 2005

Source: (Yeh, 2016).

The dataset was taken from the UCI machine learning repository that contained the credit card customers’ payment details with default details in Taiwan released for public usage and analysis by Yeh (2016). The dataset contained 30,000 credit card users’

information with 6636 default observations and 23,364 nondefault cases. This dataset contained only numeric features with no duplicates and missing values. The dataset was imbalanced since the default and nondefault cases were not exactly equal. Hence, the efficiency levels of all the models were lower compared with balanced datasets (Karatas et al., 2020). Since the practical cases were always imbalanced, the dataset was not balanced using undersampling or oversampling techniques.

The data preprocessing was performed to clean the data (García et al., 2015). Feature scaling is a crucial data pre-processing technique, and it was performed to ensure that none of the variables (features) dominated the model because of their magnitude, in addition to reducing the impact of outliers, if any, present in the dataset (Zheng & Casari, 2018). The input variables were rescaled to normalize the training data. Feature scaling was performed using the standard scaler in Python. The dataset was split in the ratio of 80:20 for training and testing, correspondingly. Later, correlation among the explanatory variables was visualized using the heat map analysis (Gu, 2022).

The various models were benchmarked with confusion matrix analysis and receiver operating characteristic curves. Confusion matrix classifies the actual values with the predicted values of the model (Luque et al., 2019). The predictive performance was evaluated using the confusion matrix as established by the literary work of Kuhn and Johnson (2013). The data were analyzed as per the confusion matrix in Table 2 and the metrics mentioned in Table 3. The ROC curve characterized the positive and negative events, exploring the trade-offs between competing losses at different threshold levels (Brown & Davis, 2006). It was free from parametric assumptions, and the values were independent from the ratio of positive and negative events. The area under the ROC (AUC) was also independent of the decision thresholds and hence suitable for classification problems in addition to confusion matrix analysis.

Table 2. Confusion matrix.

Actual	Prediction	
	0 (Negative)	1 (Positive)
0 (negative)	True negative (TN)	False positive (FP)
1 (positive)	False negative (FN)	True positive (TP)

Source: (Bekkar et al., 2013).

Table 3. Metric used to assess the model’s performance.

Metric	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Error rate = 1 – accuracy	$\frac{FP+FN}{TP+TN+FP+FN}$
Sensitivity (or recall, accuracy of positive examples)	$\frac{TP}{TP+FN}$
Specificity (accuracy of negative examples)	$\frac{TN}{TN+FP}$
Precision	$\frac{TP}{TP+FP}$
F1 score	$2 \times \frac{Precision * Recall}{Precision + Recall}$
G-mean	$G = \sqrt{Sensitivity * Specificity}$

Source: (Bekkar et al., 2013).

The dataset was analyzed with the most widely used metric for classifier evaluation, i.e., accuracy. In addition to accuracy, the study used the model assessment method over imbalanced datasets as mentioned by Bekkar et al. (2013). Practical cases were mostly imbalanced in nature, and hence, the models’ accuracy might decrease because of

imbalance in the default and nondefault cases (Karatas et al., 2020). Furthermore, the feature importance was calculated for the best performing model to highlight the importance of the input variables. Also, the default score formula was derived based on the F-score values of the feature importance graph of the best performing model.

The study was analyzed using the Python programming language. The packages Numpy and pandas were used for data analysis and interpretation. LDA, LR, SVM with radial basis function as the kernel, and RF were performed using the scikit learn package (Géron, 2022). Extreme gradient boosting was performed using the XGBoost package. The study used the Keras deep learning framework for deep neural network processing, which is a high-level API on top of Tensorflow (Géron, 2022). The best accuracy of the DNN model in this study was achieved using 4 hidden layers with rectified linear units as activation functions and repetitions of 25 epochs. Increasing the hidden layers above 4 and decreasing below 4 reduced the accuracy, and the case was similar with 25 epochs also. Hence, the model was optimized using 4 hidden layers and running 25 epochs. Furthermore, cross-validation was conducted, which was a resampling procedure to prevent overfitting and to check the model’s ability to perform well on new and unseen data. K-fold cross-validation was performed to prevent overfitting of the model and to avoid normalizing the outliers. It was performed 10-fold as it is the most widely used method for validating the results (Nti et al., 2021). The various models were validated using the k-fold cross-validation technique to minimize the effects of outliers and therefore to prevent overfitting.

4. Results

4.1. Descriptives

The dataset was skewed for limit balance and age, as shown in Figure 3. The dataset had a greater number of clients having a limiting balance between 0 and 200,000 currencies as the limit balance and a greater number of clients in the age bracket of 20 to 40, i.e., clients from mostly young to middle-aged groups.

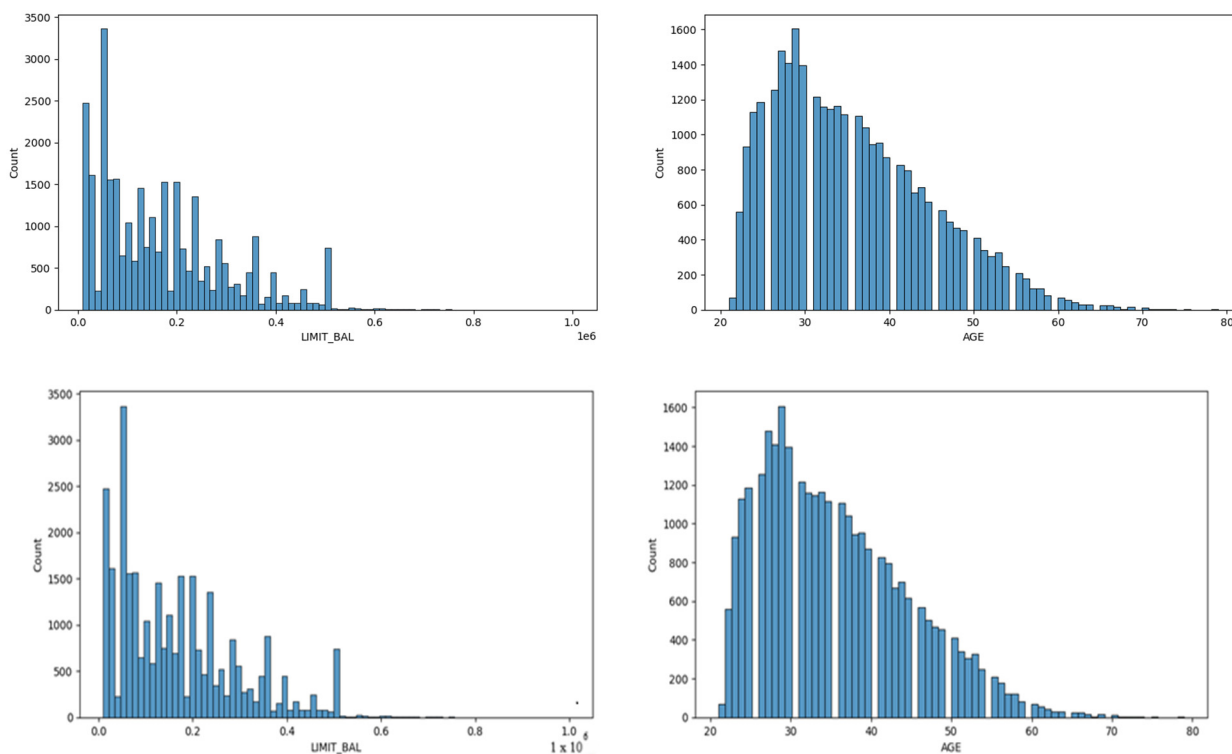


Figure 3. Skewed data for limit balance and age. Source: Authors’ own.

4.2. Heat Map

Heat maps revealed complex patterns and correlations in the dataset (Gu et al., 2016). The heat map for the explanatory variables is represented in Figure 4. A high correlation can be seen between the bill amounts (month-wise) and the payment amounts (month-wise). It was not an issue since it was expected to be high. It could be inferred that a card holder spending a certain amount in the month of April was highly likely to spend a similar amount in the next month (May) also. Similarly, the card holder who made a certain bill payment in the month of April would make the repayment in the next months also, and there were fewer chances of default. A high correlation was found between the generated bill amount and the payment made in the corresponding month. The pair-wise correlation values in Figure 4 indicate that the explanatory variables were range-bound, and they did not display strong interdependence. It was an indication that the explanatory variables were independent of each other and displayed sufficient divergence to be treated as separate variables for further analysis.

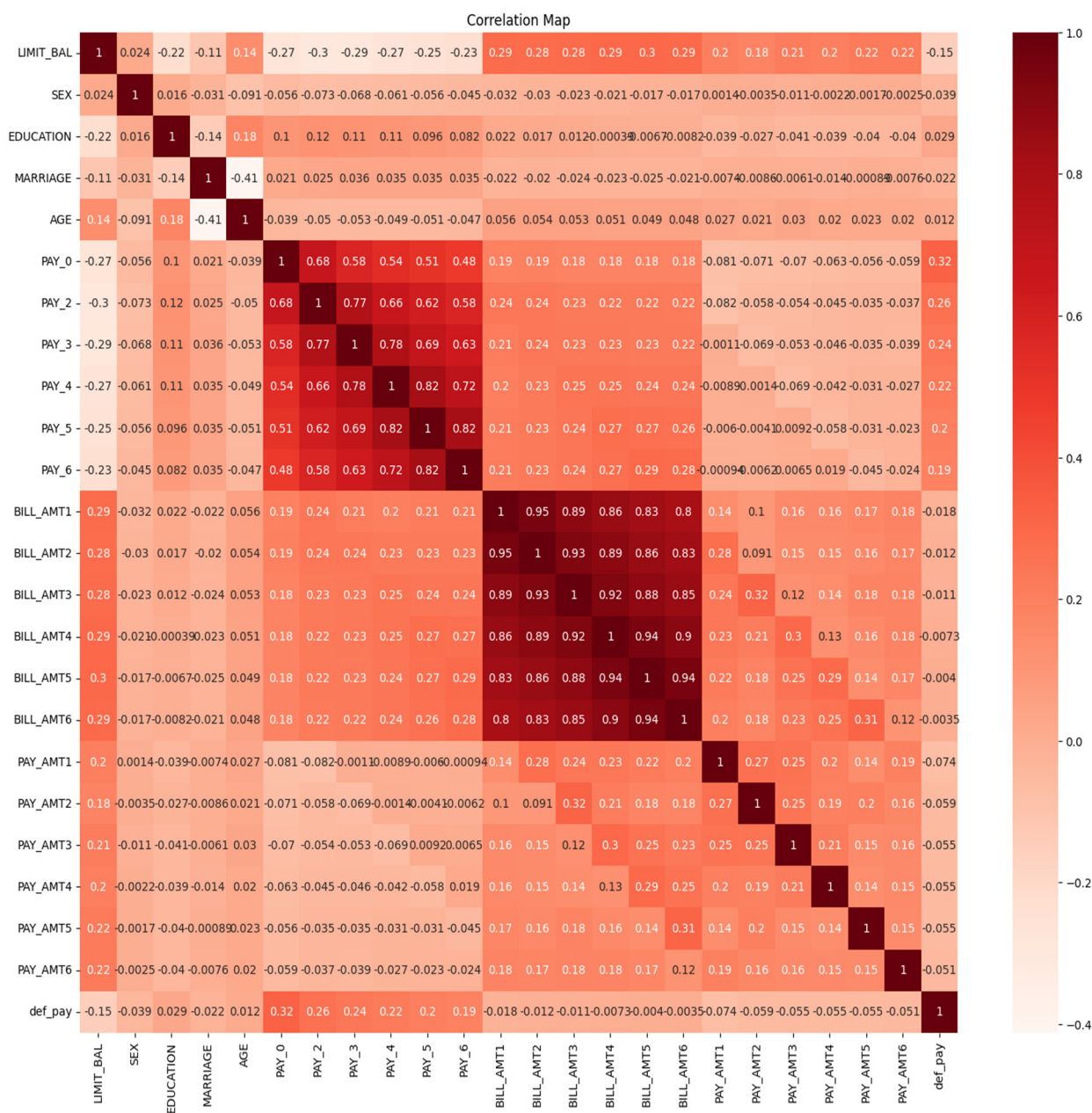


Figure 4. Heat map for the various features of the dataset. Source: Authors' own.

Figure 5 shows the relationship between the generated bill amount (Y-axis) and the corresponding payment made (X-axis) month-wise for 6 months. Since a maximum number of datapoints along the Y-axis are tightly packed near the scale zero of the X-axis in all six plots, it can be inferred that there was a greater proportion of clients for whom the bill amount was high, and at the same time the payment against the bill amount was very low. This indicated that the dataset contained significant details about delayed payments or default cases every month.

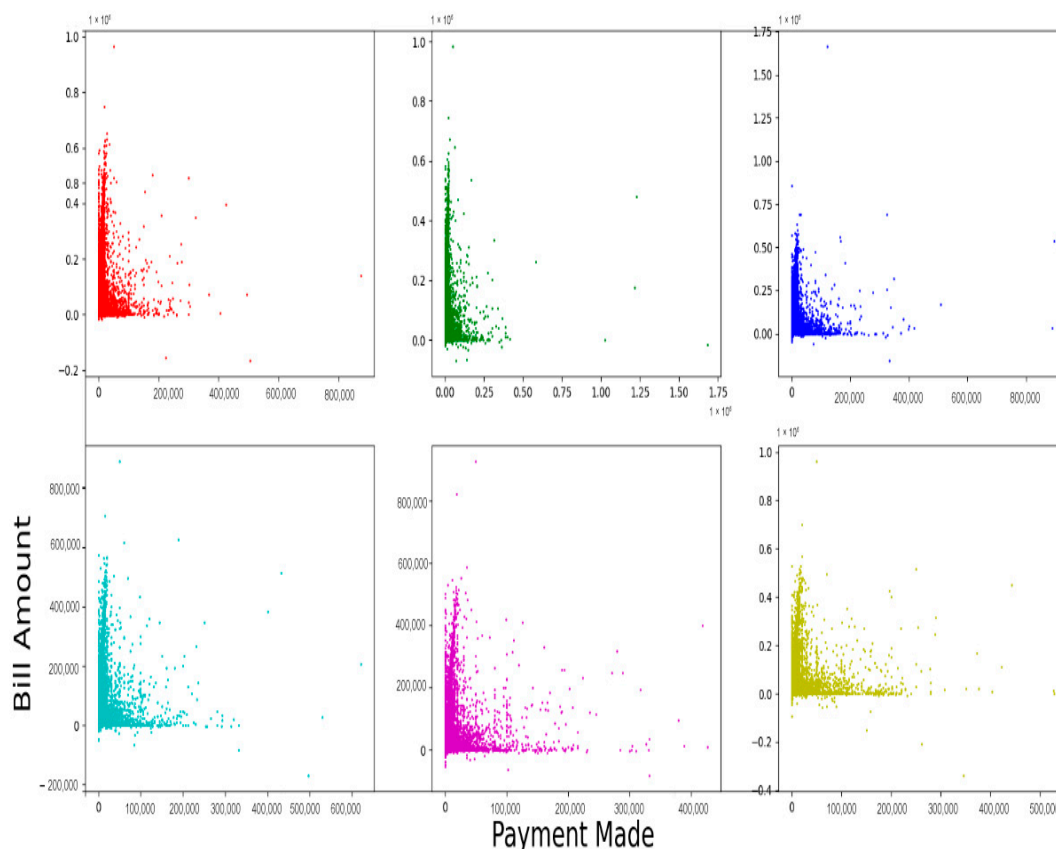


Figure 5. Relationship between the bill amount and the payment made. Source: Authors' own.

4.3. Confusion Matrix Analysis

The confusion matrices for the various models are presented in Table 4, and the metrics for the various models are given in Table 5. The values displayed in Table 5 are the average values of the cross-validated scores performed 10-fold using the K-fold cross-validation technique. The prediction accuracy of the deep neural network was at 81.80%, which was better than all the other models. The accuracy of the modern methods like XGBoost, RF, and DNN was marginally better than the traditional models like LDA, LR, and SVM. The modern methods performed well in predicting the sensitivity, but the traditional methods were better in terms of specificity. The precision score of 0.7047 for the SVM model was better in comparison with all other models. The F1-score value of 0.4820 of the DNN model was the best in comparison with the other models, and the modern methods outperformed the traditional methods in terms of the F1 score. Also, the G-mean value of 0.6027 of the DNN model was the best when compared with all other models, and the modern methods outperformed the traditional methods in terms of the G-mean score also.

Table 4. Confusion matrix for the various models.

LDA			LR			SVM		
Actual	Prediction		Actual	Prediction		Actual	Prediction	
	0	1		0	1		0	1
0	4529	158	0	4549	138	0	4560	127
1	988	325	1	1002	311	1	1010	303

XGBoost			RF			DNN		
Actual	Prediction		Actual	Prediction		Actual	Prediction	
	0	1		0	1		0	1
0	4406	281	0	4417	270	0	4400	287
1	819	494	1	832	481	1	805	508

Source: Authors' own.

Table 5. Performance metrics for the various models.

Metric	LDA	LR	SVM	XGBoost	RF	DNN
Accuracy	0.8090	0.8100	0.8105	0.8167	0.8163	0.8180
Sensitivity or recall	0.2475	0.2369	0.2308	0.3762	0.3663	0.3869
Specificity	0.9663	0.9706	0.9729	0.9400	0.9424	0.9388
Precision	0.6729	0.6927	0.7047	0.6374	0.6405	0.6390
F1 score	0.3619	0.3530	0.3477	0.4732	0.4661	0.4820
G-mean	0.4891	0.4794	0.4738	0.5947	0.5876	0.6027
AUC	0.72	0.73	0.71	0.77	0.76	0.77

Source: Authors' own.

4.4. Receiver Operating Characteristics (ROCs)

The ROC curve illustrates the classification ability of a binary classifier system as its discrimination threshold level is varied. It is plotted with sensitivity in the Y-axis and specificity in the X-axis (Bewick et al., 2004). A good predictive model will have an ROC curve that is close to the top left corner, indicating a high true positive rate with a low false positive rate. The ROC curves for the various models are plotted in Figure 6. The area under the ROC curve (AUC) measures the area underneath the ROC curve (Bewick et al., 2004). A higher value of the AUC is desirable, and a value close to 1 is considered as ideal. The area under the ROC curve was highest for the DNN model and XGBoost model at 77%, and the AUC was 72% for LDA, 73% for LR, 71% for SVM, and 76% for RF.

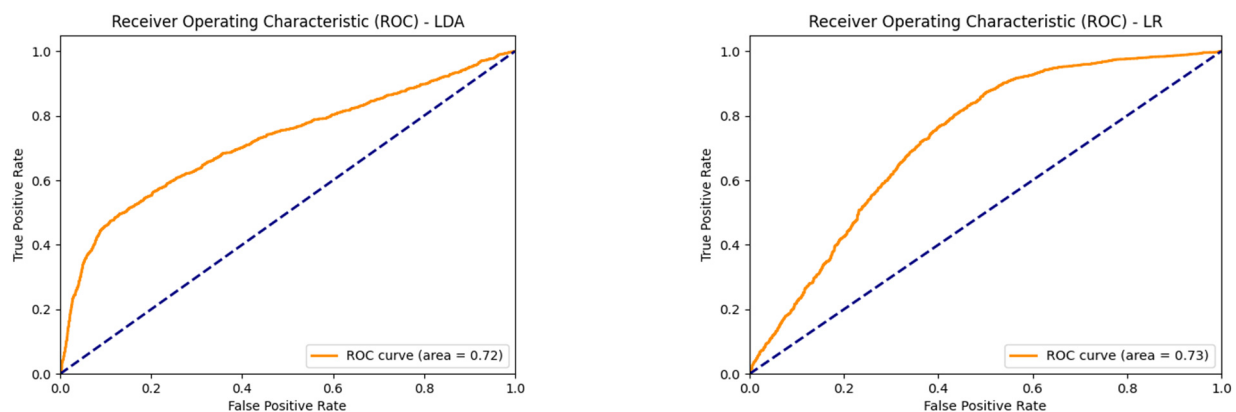


Figure 6. Cont.

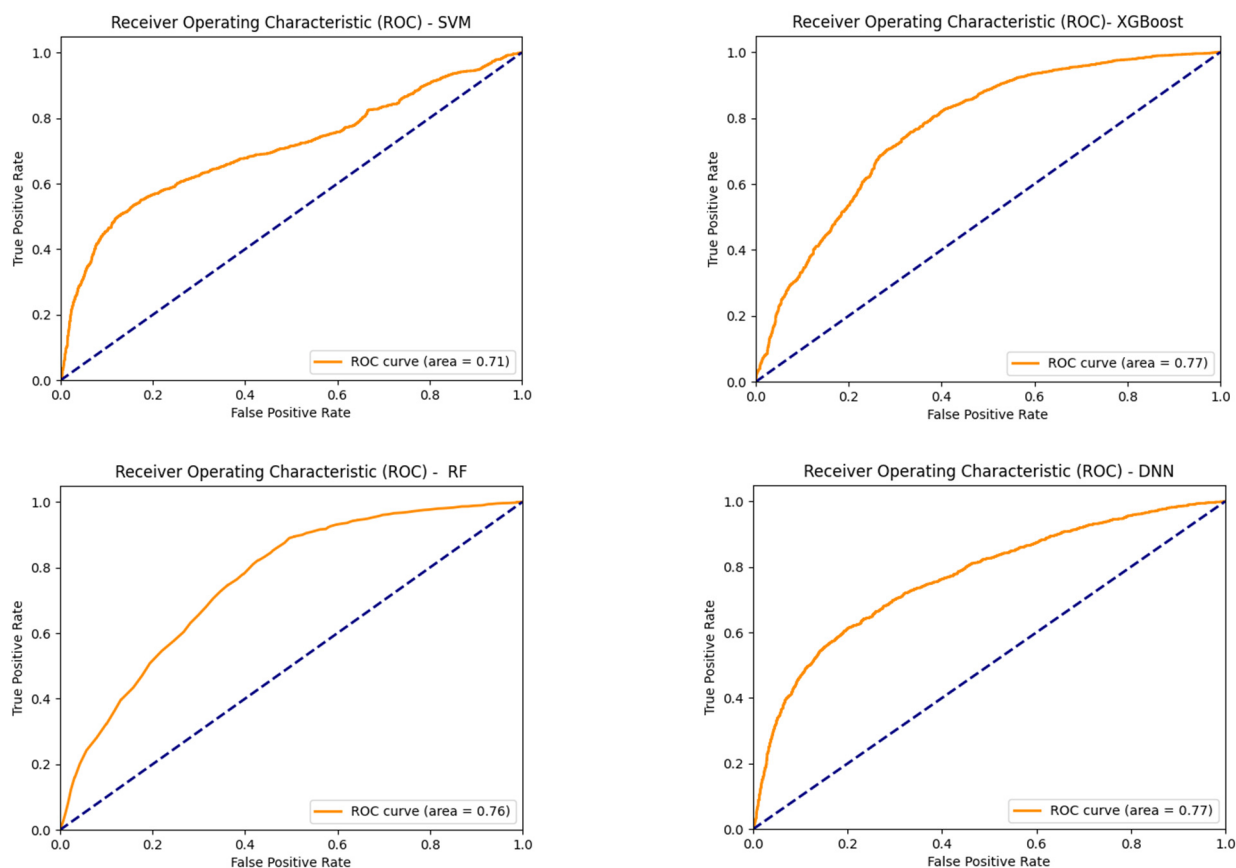


Figure 6. ROC for the various models benchmarked with the random model (blue dotted lines). Source: Authors’ own.

4.5. Feature Importance

The feature importance permits the researcher to peek inside the black box of machine algorithms to see which features are critical in informing a good prediction as per the ranking (Musolf et al., 2022). The DNN model was selected to compute the important features since the accuracy of the DNN model was the highest in comparison with all other models. The feature importance for the DNN model based on the weights in the first layer with the F score on the x-axis and features on the y-axis is depicted in Figure 7. The BillSum (sum total of the bill generated for 6 months) emerged as the most important feature, followed by age, payment made (PAY_AMTX) for various months, limit balance, and the PaySum (total sum of the payment made for 6 months). Sex, marital status, and education attainment level were the least important features in predicting the default behavior of the borrower.

4.6. Default Score Calculation

The DNN model had the best accuracy in comparison with all the other models in this study. Hence, the weightage for each feature based on their corresponding F-score value in the DNN model feature importance was the basis for deriving the default score formula. The total F-score value was 4126, and the weightage for each feature was measured by its corresponding F-score value as shown in Figure 7. Based on these parameters the default scoring formula was developed in which each borrower could be assigned a default score as per the equation given below. A score greater than 1 was considered as a good borrower, and a score less than 1 was considered as a bad credit card borrower. Based on the scores the limit balance could be increased or decreased.

$$\text{DefaultScore} = \frac{(\text{PaySum} * 0.0691) + (\text{PAY}_{\text{AMT1}} * 0.1030 + \text{PAY}_{\text{AMT4}} * 0.0962 + \text{PAY}_{\text{AMT6}} * 0.0875 + \text{PAY}_{\text{AMT3}} * 0.0858 + \text{PAY}_{\text{AMT2}} * 0.0858 + \text{PAY}_{\text{AMT5}} * 0.0826)}{(\text{Billsum} * 0.1299) + (\text{LIMIT}_{\text{BAL}} * 0.0892)} \quad (10)$$

The above equation gives a default score that can be classified under several ranges. A higher score indicated good credit utilization and payments made, whereas a lower score indicated either lower credit utilization or default in payments made.

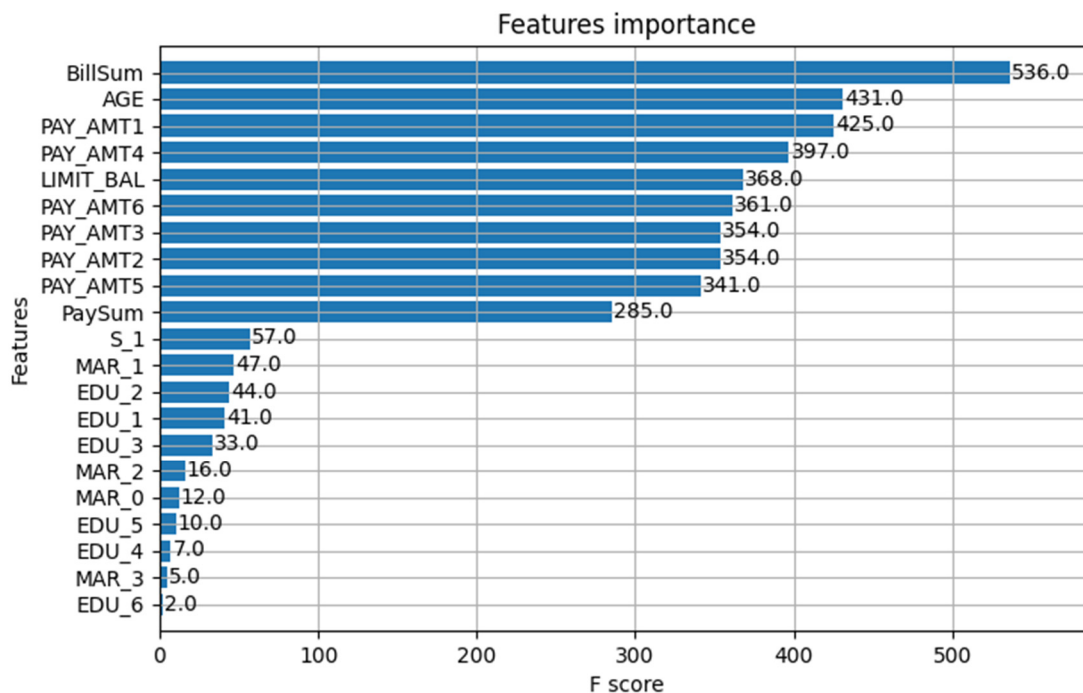


Figure 7. Feature importance for the DNN model. Source: Authors’ own.

5. Discussions and Implications

The study results are discussed with previous research findings in this section. The results of this study indicated the superiority of deep neural networks when compared with all other models. The linear models like discriminant analysis and logistic regression also showed satisfactory predictive performances. In a few cases the machine learning methods were marginally better than statistical methods, and in many other metrics, machine learning methods were superior compared with statistical methods.

As evident from the literature, age, income, marital status, gender, and education attainment level were the significant predictors of credit default behavior among borrowers (Ofori et al., 2014). This study confirmed the influence of age, marital status, gender, education attainment level, and income in predicting credit card default behavior as per the feature importance graph. In this study, income was measured with the credit limit provided by the issuer of the card. As per the literature, the past track record of the borrower was also an important predictor of future payments to be made by the potential borrower (Schreiner, 2000). Similar findings were observed in this study since the monthly payments made emerged as the significant important feature.

The cross-validated predictive accuracy of the deep neural network model was 81.80%, which was better than all the other models in this study. The accuracy of the modern methods like XGBoost, RF, and DNN was marginally better than the traditional models like LDA, LR, and SVM. The study was consistent with the findings of Hamdi et al. (2024), Maehara et al. (2024), and Puli et al. (2024). These studies confirmed the superiority of neural network models for improved predictive performance in classification problems.

The area under the ROC curve was highest (77%) for the DNN model and XGBoost model in this study. The findings of this study were consistent with the study results conducted by Hamdi et al. (2024) that displayed the superiority of DNN models with an AUC of 88%. Additionally, the study by Suhadolnik et al. (2023) compared the ROC performance of ten machine learning algorithms and found that the AUC was highest for XGboost at 71.85%, followed by the artificial neural network at 71.41%. Surprisingly, the study by Maehara et al. (2024) found that the AUC for prediction of usage of bank accounts was highest for the random forest model and support vector machine at 81.28%, and the artificial neural network was the lowest performing one at 75.73%. The reason for the deviation was due to the fact that the study by Maehara et al. (2024) was on predicting the bank account usage and not on credit default prediction.

Machine learning outperformed traditional prediction techniques to predict loan defaults and performed significantly well in forecasting recovery rates on nonperforming assets when compared with regression techniques (Bellotti et al., 2021). Also, machine learning and deep learning models had better loan default prediction accuracies when compared with statistical methods like logistic regression, there were no restrictive assumptions on the input data, and it provided the flexibility to change the input criteria with ease (Jayadev et al., 2019). This study supported the findings of Bellotti et al. (2021) and Jayadev et al. (2019) by concluding that machine learning methods had better predictive performance in comparison with traditional methods.

Fintech companies can reduce the cost of lending by taking full advantage of the advancement in digital technology and big data analytics using machine learning methods (Bazarbash, 2019). The main advantage of using machine learning over traditional methods is the improvement in calculating the out-of-sample prediction accuracy, but at the same time these methods suffer with a black box problem without a proper logic on the decision arrived at (Bazarbash, 2019). Hence, feature importance for the deep neural network model was calculated to solve the black box problem by ranking the input parameters according to their importance in the prediction (Musolf et al., 2022). The feature importance showcased that the bill sum and the age of the borrower emerged as the most important features for credit card default prediction in this study.

A good predictive model not only filters the customers based on predictive ability and reduces the credit default risk but also helps to accurately plan the economic capital requirement of the bank, ultimately increasing the profitability (DeVaney, 1999). Also, loss given default is an important parameter to determine the credit default risk. It is defined as the estimated amount that is impossible to recover from an asset in the event of a default, equationally represented as one minus the recovery rate. The loss given default depends on the borrower's ability and willingness to repay, as well as the lending institute's recovery strategies (Thomas et al., 2016). Better prediction accuracy of the loss given default helps the banks accurately calculate the economic capital requirement of the bank resulting in increased profits (Bastos, 2010). Hence, an improvement in the default prediction accuracy ultimately increases the profitability of the organization.

6. Conclusions

The main objective of this study was to compare the prediction accuracy of credit card default behavior using traditional and modern machine learning methods. Traditional methods like discriminant analysis, logistic regression, and support vector machine were compared with modern methods like XGboost, random forest, and deep neural network methods. As per the results of this study, modern machine learning methods outperformed traditional statistical methods in terms of predictive performance measured by the F1 score, G-mean, and AUC. Also, machine learning methods performed marginally better in terms

of prediction accuracy. Among the traditional methods, logistic regression was slightly better than linear discriminant analysis and support vector machines when benchmarked with the predictive performance measured by the area under the receiver operating characteristic curve. In the modern machine learning methods, deep neural networks were overall better than XGBoost and random forest methods. Also, this study found a novel method to calculate the default score using the F-score values of the DNN model's individual features as per their importance. Although the DNN model had the best accuracy in this study, a combination with other models to calculate the weightage of individual features based on averaging the feature importance F-score values can be a scope for future study to improve the default score values.

This study focused on credit card default prediction using various models, and the findings of this study supported the findings of many other articles that confirmed the superiority of machine learning models in terms of predictive performance in comparison with the traditional methods. Even though the study incorporated random forest, XGBoost, and deep neural network, it fell short on an extensive and comprehensive comparison with other gradient boosting algorithms and long short-term memory neural network methods. The comparison could have also studied the regression methods for various models with the calculated error in terms of the root mean square error and the mean absolute percentage error. Also, comparison with econometric methods like autoregressive integrated moving average could have added more insights to this study. The study was also limited to analyzing the dataset from a single database; future studies can focus on combining different datasets from different databases for training and testing the model for better predictive performance.

Author Contributions: Conceptualization, R.B. and B.K.G.; methodology, R.B. and B.K.G.; software, R.B.; validation, R.B. and B.K.G.; formal analysis, R.B. and B.K.G.; investigation, R.B.; resources, R.B. and B.K.G.; data curation, R.B. and B.K.G.; writing—original draft preparation, R.B. and B.K.G.; writing—review and editing, R.B. and B.K.G.; visualization, R.B.; supervision, R.B. and B.K.G.; project administration, R.B. and B.K.G.; funding acquisition, B.K.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: The data presented in this study are openly available in the UCI Machine Learning Repository (C55S3H, Yeh, (2016), Default of Credit Card Clients [Dataset], UCI Machine Learning Repository, <https://doi.org/10.24432/C55S3H> (accessed on 9 September 2024), UCI Machine Learning Repository, <https://archive.ics.uci.edu/dataset/350/default+of+credit+card+clients> (accessed on 9 September 2024), 10.24432/C55S3H).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Acquah, H. D., & Addo, J. (2011). Determinants of loan repayment performance of fishermen: Empirical evidence from Ghana. *Cercetari Agronomice in Moldova*, 44(4), 89–97.
- Addisu, M. (2006). Micro-finance repayment problems in the informal sector in Addis Ababa. *Ethiopian Journal of Business & Development*, 1(2), 29–50.
- Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003). Using neural network rule extraction and decision tables for credit-risk evaluation. *Management Science*, 49(3), 312–329. [CrossRef]
- Bali, T. G., Beckmeyer, H., Mörke, M., & Weigert, F. (2023). Option return predictability with machine learning and big data. *The Review of Financial Studies*, 36(9), 3548–3602. [CrossRef]

- Bandyopadhyay, A. (2023). Predicting the probability of default for banks' expected credit loss provisions. *Economic & Political Weekly*, 58(10), 15–19.
- Bastos, J. A. (2010). Forecasting bank loans loss-given-default. *Journal of Banking & Finance*, 34(10), 2510–2517. [CrossRef]
- Bazarbash, M. (2019). *Fintech in financial inclusion: Machine learning applications in assessing credit risk*. International Monetary Fund.
- Bekkar, M., Kheliouane Djemaa, H., & Akrouf Alitouche, T. (2013). Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications*, 3(10), 27–38.
- Bellotti, A., Brigo, D., Gambetti, P., & Vrins, F. (2021). Forecasting recovery rates on non-performing loans with machine learning. *International Journal of Forecasting*, 37(1), 428–444. [CrossRef]
- Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 13: Receiver operating characteristic curves. *Critical Care*, 8, 1–5.
- Bhandary, R., Shenoy, S. S., Shetty, A., & Shetty, A. D. (2023a). Attitudes toward educational loan repayment among college students: A qualitative enquiry. *Journal of Financial Counseling and Planning*, 34(2), 281–292. [CrossRef]
- Bhandary, R., Shenoy, S. S., Shetty, A., & Shetty, A. D. (2023b). Education loan repayment: A systematic literature review. *Journal of Financial Services Marketing*, 29, 1365–1376. [CrossRef]
- Breiman, L. (2001a). Random forests. *Machine Learning*, 45, 5–32. [CrossRef]
- Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3), 199–231. [CrossRef]
- Brown, C. D., & Davis, H. T. (2006). Receiver operating characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, 80(1), 24–38. [CrossRef]
- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189–215. [CrossRef]
- Chang, C. -H. (2022). Information asymmetry and card debt crisis in taiwan. *Bulletin of Applied Economics*, 9(2), 123–145. [CrossRef]
- Chen, T., & Guestrin, C. (2016, August 13–17). *XGBoost*. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794), San Francisco, CA, USA. [CrossRef]
- Chiang, W. K., Zhang, D., & Zhou, L. (2006). Predicting and explaining patronage behavior toward web and traditional stores using neural networks: A comparative analysis with logistic regression. *Decision Support Systems*, 41(2), 514–531. [CrossRef]
- Coşer, A., Maer-Matei, M. M., & Albu, C. (2019). Predictive models for loan default risk assessment. *Economic Computation & Economic Cybernetics Studies & Research*, 53(2), 149–165. [CrossRef]
- DeVaney, S. A. (1999). Determinants of consumer's debt repayment patterns. *Consumer Interests Annual*, 45, 65–70.
- Dicon, H. (2024, May 8). *Why are americans cutting back on credit card debt?* Investopedia. Available online: <https://www.investopedia.com/why-americans-are-cutting-back-on-credit-card-debt-8645241> (accessed on 12 September 2024).
- Drobotz, W., & Otto, T. (2021). Empirical asset pricing via machine learning: Evidence from the European stock market. *Journal of Asset Management*, 22(7), 507–538. [CrossRef]
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), 179–188. [CrossRef]
- García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72, pp. 59–139). Springer International Publishing.
- Gauthier, C., Lehar, A., & Souissi, M. (2012). Macroprudential capital requirements and systemic risk. *Journal of Financial Intermediation*, 21(4), 594–618. [CrossRef]
- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc.
- Globaldata. (2023). *Taiwan cards and payments market report overview*. Globaldata Report Store. GDFS0736CI-ST.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273. [CrossRef]
- Gu, Z. (2022). Complex heatmap visualization. *Imeta*, 1(3), e43. [CrossRef]
- Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18), 2847–2849. [CrossRef] [PubMed]
- Hamdi, M., Mestiri, S., & Arbi, A. (2024). Artificial intelligence techniques for bankruptcy prediction of tunisian companies: An application of machine learning and deep learning-based models. *Journal of Risk and Financial Management*, 17(4), 132. [CrossRef]
- Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society, Series A—Statistics in Society*, 160(3), 523–541. [CrossRef]
- Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and Their Applications*, 13(4), 18–28. [CrossRef]
- Hochreiter, S., Younger, A. S., & Conwell, P. R. (2001, August 21–25). *Learning to learn using gradient descent*. Artificial Neural Networks—ICANN 2001: International Conference, Proceedings 11 (pp. 87–94), Vienna, Austria.
- Hubbard, D. W. (2020). *The failure of risk management: Why it's broken and how to fix it*. John Wiley & Sons.
- IASB—International Accounting Standards Board. (2014). *IFRS 9 financial instruments*. Available online: <http://www.ifrs.org> (accessed on 12 September 2024).

- Jayadev, M., Shah, N., & Vadlamani, R. (2019). Predicting educational loan defaults: Application of machine learning and deep learning models. *IIM Bangalore Research Paper*, 601, 1–46.
- Karatas, G., Demir, O., & Sahingoz, O. K. (2020). Increasing the performance of machine learning-based IDSs on an imbalanced and up-to-date dataset. *IEEE Access*, 8, 32150–32162. [CrossRef]
- Krenker, A., Bešter, J., & Kos, A. (2011). Introduction to the artificial neural networks. In *Artificial Neural Networks: Methodological Advances and Biomedical Applications* (Volume 1, pp. 1–18). InTech.
- Kuhn, M., & Johnson, K. (2013). Measuring performance in classification models. In *Applied predictive modeling*. Springer.
- Luque, A., Carrasco, A., Martín, A., & de Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, 216–231. [CrossRef]
- Maalouf, M. (2011). Logistic regression in data analysis: An overview. *International Journal of Data Analysis Techniques and Strategies*, 3(3), 281–299. [CrossRef]
- Maehara, R., Benites, L., Talavera, A., Aybar-Flores, A., & Muñoz, M. (2024). Predicting financial inclusion in peru: Application of machine learning algorithms. *Journal of Risk and Financial Management*, 17(1), 34. [CrossRef]
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5, 115–133. [CrossRef]
- More, A. S., & Rana, D. P. (2017, October 5–6). *Review of random forest classification techniques to resolve data imbalance*. 2017 1st INTERNATIONAL Conference on Intelligent Systems and Information Management (ICISIM) (pp. 72–78), Aurangabad, India.
- Musolf, A. M., Holzinger, E. R., Malley, J. D., & Bailey-Wilson, J. E. (2022). What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. *Human Genetics*, 141(9), 1515–1528. [CrossRef] [PubMed]
- Netzer, O., Lemaire, A., & Herzenstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research*, 56(6), 960–980. [CrossRef]
- Nti, I. K., Nyarko-Boateng, O., & Aning, J. (2021). Performance of machine learning algorithms with different K Values in K-fold crossvalidation. *International Journal of Information Technology and Computer Science*, 13(6), 61–71. [CrossRef]
- Ofori, K. S., Fianu, E., Omoregie, O. K., Odai, N. A., & Oduro-Gyimah, F. (2014). Predicting credit default among micro borrowers in Ghana. *Research Journal of Finance and Accounting ISSN*, 1697–2222.
- Puli, S., Thota, N., & Subrahmanyam, A. C. V. (2024). Assessing machine learning techniques for predicting banking crises in India. *Journal of Risk and Financial Management*, 17(4), 141. [CrossRef]
- Ramraj, S., Uzir, N., Sunil, R., & Banerjee, S. (2016). Experimenting XGBoost algorithm for prediction and classification of different datasets. *International Journal of Control Theory and Applications*, 9(40), 651–662.
- Saraiva, A. (2024, August 8). *US Consumer Borrowing Rises Less Than Forecast on Credit Cards*. Bloomberg. Available online: <https://www.bloomberg.com/news/articles/2024-08-07/us-consumer-credit-misses-forecast-on-lower-card-balances> (accessed on 12 September 2024).
- Schreiner, M. (2000). Credit scoring for microfinance: Can it work. *Journal of Microfinance/ESR Review*, 2(2), 105–118.
- Shao, Y., Ahmed, A., Zamrini, E. Y., Cheng, Y., Goulet, J. L., & Zeng-Treitler, Q. (2023). Enhancing clinical data analysis by explaining interaction effects between covariates in deep neural network models. *Journal of Personalized Medicine*, 13(2), 217. [CrossRef]
- Sperandei, S. (2014). Understanding logistic regression analysis. *Biochimica Medica*, 12–18. [CrossRef] [PubMed]
- Suhadolnik, N., Ueyama, J., & Da Silva, S. (2023). Machine learning for enhanced credit risk assessment: An empirical approach. *Journal of Risk and Financial Management*, 16(12), 496. [CrossRef]
- Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12), 2295–2329. [CrossRef]
- Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI communications*, 30(2), 169–190. [CrossRef]
- Thomas, L. C., Matuszyk, A., So, M. C., Mues, C., & Moore, A. (2016). Modelling repayment patterns in the collections process for unsecured consumer debt: A case study. *European Journal of Operational Research*, 249(2), 476–486. [CrossRef]
- Vapnik, V. (1998). The support vector method of function estimation. In *Nonlinear modeling: Advanced black-box techniques* (pp. 55–85). Springer.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3–28. [CrossRef]
- Viswanathan, P. K., & Shanthi, S. K. (2017). Modelling credit default in microfinance—an indian case study. *Journal of Emerging Market Finance*, 16(3), 246–258. [CrossRef]
- Yeh, I. -C. (2016). Default of credit card clients. *UCI Machine Learning Repository*, 10, C55S3H.

- Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 36(2), 2473–2480. [[CrossRef](#)]
- Zheng, A., & Casari, A. (2018). *Feature engineering for machine learning: Principles and techniques for data scientists*. O'Reilly Media, Inc.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.