Research article

Credit scoring using machine learning and deep Learning-Based models

Sami Mestiri[1,2,*]

[1] Faculty of Management and Economic Sciences of Mahdia, University of Monastir, Tunisia
[2] Sidi Messaoud Hiboun, Mahdia, Tunisia

* **Correspondence:** Email: sami.mestiri@fsegma.u-monastir.tn; Tel: 216-95-504-632.

Abstract: Credit scoring is a useful tool for assessing the capability of customers repayments. The purpose of this paper is to compare the predictive abilities of six credit scoring models: Linear Discriminant Analysis (LDA), Random Forests (RF), Logistic Regression (LR), Decision Trees (DT), Support Vector Machines (SVM) and Deep Neural Network (DNN). To compare these models, an empirical study was conducted using a sample of 688 observations and twelve variables. The performance of this model was analyzed using three measures: Accuracy rate, F1 score, and Area Under Curve (AUC). In summary, machine learning techniques exhibited greater accuracy in predicting loan defaults compared to other traditional statistical models.

Keywords: credit scoring; machine learning; artificial intelligence; model comparison; personal loan

JEL Codes: C530, G210, G33

1. Introduction

For banks and other lending institutions, predicting loan defaults has always been crucial and extremely difficult. Financial analysts and specialists are therefore searching for and identifying the most effective methods to assist them in making decisions. Conventional methods have been extensively applied to credit risk assessment for a considerable amount of time. There are two types of credit risk assessment. In the first category, applicants are classified as having "good" and bad credit risk. Application scoring is the process of grouping data according to financial information. In the second category, the applicant's payment history, payment patterns, and other information are taken into account. This is called behavioral scoring (e.g. Woo and Sohn (2022)). Application scoring is the main topic of this paper. However, there are issues with these models' ability to precisely predict loan default. (Lyn et al. (2002) and Mestiri and Farhat (2021)).

In recent years, artificial intelligence and machine learning models have emerged as powerful tools for forecasting, as they can handle large datasets and capture nonlinear relationships between input variables and the output. We discuss the idea of machine learning, including deep neural network, which are non-linear algorithms, to compare their performances and demonstrate possibilities of the sophisticated modelling in finance. One useful tool for many financial tasks is deep learning, a subfield of machine learning that specializes in handling intricate, non-linear patterns in data (see Tran et al., (2016) and Wang et al., (2018)). Deep learning has been used more and more in the finance industry in recent years. For more details about the deep learning approach, we can refer to the studies of Deng and Yu (2014) and Le Cun et al., (2015).

The research paper is organized as follows: Section 2 provides a pertinent literature review related to forecasting loan defaults. Section 3 presents the different statistical and artificial intelligence techniques used in this study. In Section 4, the data used are described. Section 5 is devoted to the empirical investigation to forecast the loan defaults of Tunisian customers. Finally, Section 6 concludes the paper.

2. Related literature

Statistical methods have been used since the 1950s and are still popularly used today because they enable lenders to use concepts of sample estimators, confidence intervals, and statistical inference in credit scoring. This allows scorecard developers to evaluate the discriminatory power of models and determine which borrower characteristics are more important in explaining borrower behaviour. Linear discriminant analysis was one of the earliest approaches used in credit scoring. Even though the scorecards it produced were very robust, the assumptions needed to ensure satisfactory discriminatory power were restrictive.

Lyn et al. (2002) used logistic regression, a statistical technique for credit scoring that has proven successful and has replaced linear discriminant analysis. According to Mellisa (2020), the methods used for credit scoring have increased in sophistication in recent years. They have evolved from traditional statistical techniques to innovative methods such as artificial intelligence, including machine-learning algorithms such as random forests, gradient boosting, and deep neural networks. With the ongoing discussions in the banking industry, the future of machine learning (ML) models will soon be more prevalent. Several advanced techniques have been used to predict loan default, such as decision trees and support vector machines. As a matter of fact, with the invasion of artificial intelligence modeling algorithms since the 1990s in diverse domains, the artificial neural network (ANN) was the most popular machine learning technique used in finance (Fuster et al., (2022)).

Liu (2018) conducted research on credit card clients and compared Support Vector Machine, k-Nearest Neighbours, Decision Tree, and Random Forest with Feedforward Neural Network and Long Short-Term Memory. Their research was done in order to improve on earlier findings by Lien and Yeh (2009). Liu (2018) proposed to add two important factors, drop-out and long short-term memory, to neural networks in order to find their effect on improving accuracy and also solving the problem of overfitting. They used the same dataset as in Lien and Yeh (2009) with the 30,000 samples randomly shuffled, after which the top 10,000 samples were chosen. The top 8500 samples were used as a training set, with the other 1500 used as a testing set. The data was normalized to a mean of zero and a variance of one.

Stefan et al. (2015) looked at some of the improvements that could be introduced to credit scoring. They used 41 classification methods across 8 credit scoring data sets. Their work was focus on data, classification methods, and indicators used to assess the performance of the classification methods. Support Vector Machine (SVM) emerged as a competing approach to Artificial Neural Network (ANN). Giudici et al. (2020) categorized SVM and ANN as part of the non-parametric approach. SVM is a method used to classify objects without considering multicollinearity among the predictors. This method implements the concept that input data is transformed into a higher feature dimension using the kernel function. Woo and Sohn (2022) created a weighted machine learning model using text-mining techniques and psychometric characteristics. According to the justification given above, the goal of this work is to determine how to use machine learning for credit scoring. Then, the aims of this work is to compare the effectiveness of various parametric statistical and non-parametric machine learning approaches for customer loan classification.

3. Statistical, machine learning and deep learning techniques

3.1. Linear discriminant analysis (LDA)

Fisher's 1933 study laid the foundation for the use of multiple quantitative variables combined in a linear fashion to discriminate between various groups or categories. This linear combination of descriptors is called the discriminant function. The output of LDA is a score that consists of classifying a data observation between the good and bad classes.

$$Score = \sum_{i=0}^{p} a_i X_i \qquad (1)$$

Where a_i are the weights associated with the quantitative input variables X_i.

3.2. Logistic regression (LR)

Logistic regression is a statistical method used for binary classification tasks (e.g., 0 or 1, bad or good, health or default, etc.). The outcome of the LR model can be written as:

$$P(y = 1|X) = sigmoid(z) = \frac{1}{1 + exp(-z)} \qquad (2)$$

where $P(y = 1|X)$ is the probability of y being 1, given the input variables X, z is a linear combination of X: $z = a_0 + a_1 X_1 + a_2 X_2 + .. + a_p X_p$
where a_0 is the intercept term, $a_1, a_2, ..., a_p$ are the weights, and $X_1, X_2, ..., X_p$ are the input variables.

3.3. Decision trees (DT)

Decision trees proceed with a recursive partitioning of the data into subsets based on the values of the input variables, with each partition represented by a branch in the tree Quinlan (1986). The functioning of decision trees aims to train a sequence of binary decisions that can be used to predict the value of the output for a new observation. Each decision node in the tree corresponds to a test of the value of one of the input variables, and the branches correspond to the possible outcomes of the test. The leaves of the tree correspond to the predicted values of the output variable for each combination of input values. The

decision tree algorithm works by recursively partitioning the data into subsets based on the values of the input variables. At each step, the algorithm selects the input variable that provides the best split of the data into two subsets that are as homogeneous as possible with respect to the output variable. An information gain or Gini impurity criterion, which measures the amount of uncertainty reduced about the output variable that the split reduces, is commonly used to assess the quality of a split.

Decision trees are typically not formulated in terms of mathematical equations, but as a sequence of logical rules that describe how the input variables are used to predict the output variable. However, the splitting criterion used to select the best split at each decision node can be expressed mathematically. Suppose i have a dataset with n observations and m input variables, denoted by $X_1, X_2, ..., X_p$, and a binary output variable y that takes values in 0,1. Let S be a subset of the data at a particular decision node, and let p_i be the proportion of observations in S that belong to class i. The Gini impurity of S is defined as:

$$G(S) = 1 - \sum_i (p_i)^2 \tag{3}$$

The Gini impurity measures the probability of misclassifying an observation in S if i randomly assign it to a class based on the proportion of observations in each class. A small value of G(S) indicates that the observations in S are well-separated by the input variables.

To split the data at a decision node, we consider all possible splits of each input variable into two subsets and choose the split that minimizes the weighted sum of the Gini impurities of the resulting subsets. The weighted sum is given by:

$$\Delta G = G(S) - (\frac{|S_1|}{|S|}).G(S_1) - (\frac{|S_2|}{|S|}).G(S_2) \tag{4}$$

where S_1 and S_2 are the subsets of S resulting from the split, and $|S_1|$ and $|S_2|$ are their respective sizes. The split with the smallest value of ΔG chosen as the best split. The decision tree algorithm proceeds recursively, splitting the data at each decision node based on the best split until a stopping criterion is met, such as reaching a maximum depth or minimum number of observations at a leaf node.

3.4. Support vector machine (SVM)

Support vector machine (SVM), developed by Vapnik (1998), is a supervised learning algorithm used for classification, regression, and outlier detection. The basic idea of this technique is to find the best separating hyperplane between the two classes in a given dataset. The mathematical formulation of SVM can be divided into two parts: The optimization problem and the decision function.

Given a training set (x_i, y_i) where x_i is the ith input vector and y_i is the corresponding output label $y_i = (-1, 1)$, SVM seeks to find the best separating hyperplane defined by:

$$w.x + b = 0 \tag{5}$$

where w is the weight vector, b is the bias term, and x is the input vector.

SVM algorithm aims to find the optimal w and b that maximize the margin between the two classes. The margin is defined as the distance between the hyperplane and the closest data point from either class. Then, SVM optimization problem can be formulated as:

minimize $\frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} \xi_i$ subject to $y_i(w^t x_i + b) \geq 1 - \xi_i$ and $\xi_i \geq 0$

where $\|w\|^2$ is the L2-norm of the weight vector, C is a hyperparameter that controls the trade off between maximizing the margin and minimizing the classification error, ξ_i is the slack variable that allows for some misclassification, and the two constraints enforce that all data points lie on the correct side of the hyperplane with a margin of at least $1 - \xi_i$.

The optimization problem can be solved using convex optimization techniques, such as quadratic programming. Once the optimization problem is solved, the decision function can be defined as:

$$f(x) = sign(wx + b) \tag{6}$$

where sign is the sign function that returns +1 or -1 depending on the sign of the argument. The decision function takes an input vector x and returns its predicted class label based on whether the output of the hyperplane is positive or negative. Thereafter, SVM finds the best way to separate hyperplanes by solving an optimization problem that maximizes the margin between the two classes, subject to constraints that ensure all data points are correctly classified with a margin of at least $1 - \xi_i$. The decision function then predicts the class label of new data points based on the output of the hyperplane.

3.5. Random forests (RF)

Random Forest is an ensemble of learning algorithms developed by Breiman in 2001. It is a type of ensemble learning method that combines multiple decision trees for making predictions. The algorithm is called "random" because it uses random subsets of the features and random samples of the data to build the individual decision trees. The data is split into training and testing sets. The training set is used to build the model, and the testing set is used to evaluate its performance. At each node of a decision tree, the algorithm selects a random subset of the features to consider when making a split. This helps to reduce overfitting and increase the diversity of the individual decision trees.

A decision tree is built using the selected features and a subset of the training data. The tree is grown until it reaches a pre-defined depth or until all the data in a node belongs to the same class. Suppose we have a dataset with n observations and p features. Let X be the matrix of predictor variables and Y be the vector of target variables.

To build a Random Forest model, i first create multiple decision trees using a bootstrap sample of the original data. This means that i randomly sample n observations from the dataset with replacement to create a new dataset, and this process is repeated k times to create k bootstrap samples. For each bootstrap sample, we then create a decision tree using a random subset of p features. At each node of the tree, i select the best feature and threshold value to split the data based on a criterion such as information gain or Gini impurity. I repeat the above steps k times to create k decision trees. To make a prediction for a new observation, i pass it through each of the k decision trees and therefore obtain k predictions.

3.6. Deep neural network (DNN)

Deep neural network (DNN) is an enhanced version of the conventional artificial neural network with at least three hidden layers Schmidhuber (2015). Figure 1 illustrates the standard architecture of a deep neural network.
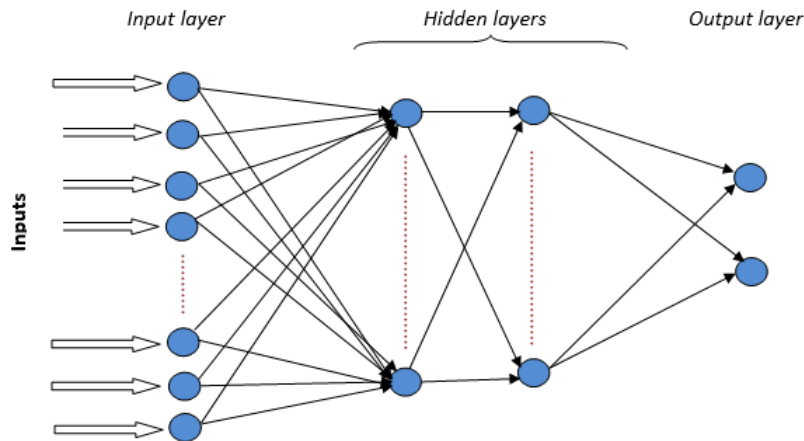
Figure 1. The standard architecture of DNN.

A solid understanding of the fundamentals of artificial neural networks is required to completely comprehend how DNN functions. The following formula determines the DNN output:

$$y(t) = \sum_{k=1}^{L} f(w_k + x_k(t)) + \epsilon(t) \tag{7}$$

where w_k are the weights of the layer trained by backpropagation. $x_k(k = 1, ..., L)$ is the total number of sequence of real values called events, during an epoch. f is the activation function.

4. Data

The aim of this research is to predict the outcome of variable as good or bad using the data set, which is classified based on features or attributes, utilizing different machine learning classification algorithms that are applied to the same data set to compare the accuracy of each of them.
This empirical study used a Tunisian commercial bank's personal loan data set (Available from the author). This data contains both continuous and categorical data. A total of 12 variables are used for this analysis; each instance is characterized by the first 12 variables, and the last attribute is used to classify if a transaction is good or bad. Table 1 presents the different attributes and their classes, which are either numerical or categorical in nature.

The data consists of 688 personal loans, with 577 good loans and 111 bad loans. The proportion of bad loans (default) compared to good loans (non-default) is 19.23. Based on Table 2, the average of the debtor's ages is 33.5 years old. The average of net income is 3.286 Dinars.

According to Table 3, the majority of the debtors are men, and up to 533 of them are already married. Most people's educational backgrounds are acquired through others (high school). This dataset will be divided into 70:30 proportions before each method is examined. To validate, the best model is the one that is used the most.

Table 1. List of variables used in credit score modeling.

ID variables	Description	Type
x_1	Age in years plus twelfths of a year	Numerical
x_2	Yearly income (in Dinars)	Numerical
x_3	Credit length (in months)	Numerical
x_4	Amount of loans (in Dinars)	Numerical
x_5	Length of stay (in years)	Numerical
x_6	Purpose	Categorical
x_7	Employment	Categorical
x_8	Type of house	Categorical
x_{10}	Marital Status	Categorical
x_{11}	Education	Categorical
x_{12}	Number of dependent	Categorical
y	Default: Good-Bad Indicator	Categorical

Table 2. Statistics descriptive for continuous variables.

Variables	Mean	St.Dev	Min	Max
x_1	33.55	8.9	25.66	64.16
x_2	3.286	1.498	2.237	12.000
x_3	87.14	77.89	10.00	240.00
x_4	106.6	145.812	5.0	600.0
x_5	3.55	1.87	1	8

Table 3. Statistics descriptive for categorical variables.

Variables	Category	Mode	Freq.mode
x_6	3	House	233
x_7	2	Private	648
x_8	2	Own	346
x_9	2	Male	357
x_{10}	3	Married	533
x_{11}	4	High school	405
x_{12}	4	0	350

5. Empirical investigation

5.1. Predictive performance measures

The predictive power of the methods used can be compared and assessed using a number of criteria, such as accuracy rate, F1 score, and AUC.

5.1.1. Accuracy rate

The accuracy rate is the most famous performance metric, deduced from the matrix confusion (see. Table 4) and calculated following this formula:

$$Accuracy\ rate = \frac{(T_0 + T_1)}{(T_0 + F_1) + (F_0 + T_1)} \tag{8}$$

Table 4. Confusion matrix.

	Predicted class "0"	Predicted class "1"
Actual class "0"	True positive (T_0)	False positive (F_1)
Actual class "1"	False negative (F_0)	True negative (T_1)

5.1.2. F1 score

The F1 score is also computed from the confusion matrix. The value of F1 score varies between 0 and 1, since 1 is the best possible score. A high F1-score indicates that the model shows both high precision and high recall, meaning it can correctly identify positive and negative cases.

$$F1\ score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{9}$$

Where $Recall = \frac{T_0}{T_0 + F_0}$ and $Precision = \frac{T_0}{T_0 + F_1}$

5.1.3. AUC

Area Under Curve (AUC) is a synthetic indicator derived from the ROC curve. This curve is a graphical indicator used to assess the forecasting accuracy of the model. The ROC curve is based on two relevant indicators that are specificity and sensitivity (see Pepe (2000) for further details). This curve is caracterized by the 1- specificity rate on the x axis and by sensitivity on the y axis. where

$$Sensitivity = True\ positive\ rate = \frac{T_0}{Positives} = \frac{T_0}{T_0 + F_1} \tag{10}$$

and

$$Specificity = True\ negative\ rate = \frac{T_1}{Negatives} = \frac{T_1}{T_1 + F_0} \tag{11}$$

Moreover, the AUC measure reflects the quality of the model classification between heath and default firms. In the ideal case, AUC is equal to 1, i.e. the model makes it possible to completely separate all the positives from the negatives, without false positives or false negatives.

5.2. Results and discussion

The data set is divided using feature extraction into training and testing data in order to efficiently assess the credit risk using machine learning classification algorithms. To the training data, different classification algorithms are applied. These models are implemented in Rstudio (see Mestiri (2024)). This hyper parameter tuning was used to train a number of models in this study:

Decision Tree: max. depth=6

Random Forest: max. depth=10 and ntree=1000

SVM model: Kernel function used is Gaussian Radial Basis Function (RBF), cost = 10 and gamma= 0.076.

Deep Neural Network : Recurrent neural networks (RNNs) used three hidden layer. Nodes per Layer are 200,100,40,1 (using Keras Sequential API). Activation function is ReLU and Loss function is binary cross entropyand. Output unit is Sigmoid. Optimizer use default settings with Epochs 100, Batchsize 100 and Validation size 0.3. Early-Stopping is applied to mitigate overfitting.

The predictive model is built using the test data. The predictive model's output is compared to the model built using trained data. Table 5 presents the empirical results of the accuracy rate, F1 score, and AUC criteria used to evaluate the classifier's performance in the applied models.

Table 5. Prediction results and models accuracy.

Models	Accuracy rate	F1-score	AUC	**Rank**
Linear Discriminant Analysis (LDA)	70.9%	0.790	0.474	5
Logistic Regression (LR)	75.8%	0.822	0.533	3
Decision Trees (DT)	64.3%	0.738	0.575	6
Random Forest (RF)	78.2%	0.833	0.715	2
Support Vector Machine (SVM)	74.8%	0.810	0.563	4
Deep Neural Network (DNN)	83.6%	0.864	0.788	**1**

According to Table 5, the deep neural network outperforms the other techniques in terms of all forecasting performance metrics. DNN shows the highest accuracy rate with 83.6% whereas 78.2% for RF and 75.8% for LR. The lowest rate of prediction accuracy was found using DT (64.3%). For the same objective to assess the predictive ability of the proposed algorithms, F1-score equal to 0.864 proves DNN's ability to identify with a great precision good from bad customers. Since 1 is the most desired F1 score, DNN reaches the highest score while F1 score value was equal to 0.833, 0.822, 0.810, 0.790 and 0.738 for RF, LR, SVM, LDA, and DT, respectively.

Other graphical indicators were also used to evaluate the quality of classification of the models under study. We talk about the ROC curve (see. Figure 2). From this curve, we deduce the AUC measure. The AUC value is nearing unity, and the model shows a high quality of classification between health and default firms. Based on Table 5, the AUC of DNN yields 0.788. In the second rank, we found the RF with an AUC equal to 0.715. The LR and LDA models present the worst classification results, as the AUC is 0.533 and 0.474, respectively, in the testing sample.
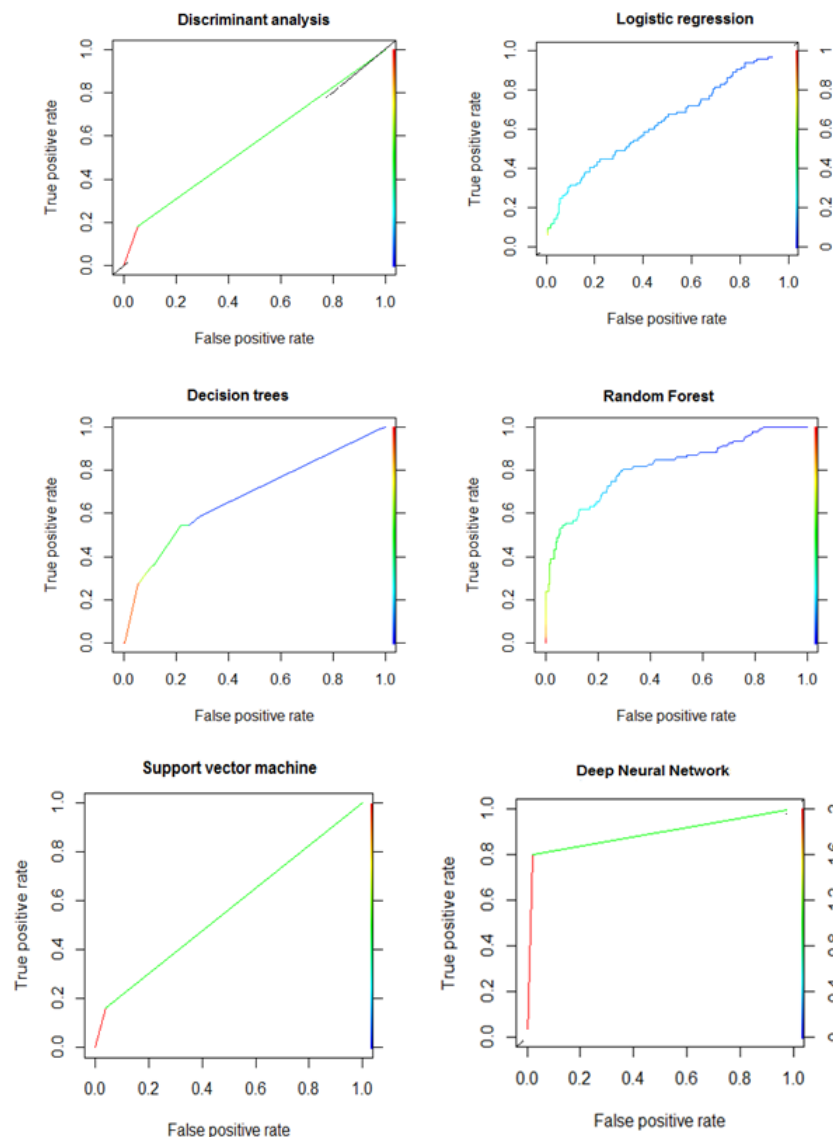
Figure 2. ROC curve for the five machine learning models and DNN.

The results show that when compared to statistical and traditional machine learning techniques, the deep neural network model more especially, recurrent neural networks (RNNs) has better prediction performance. In summary, DNN performs better at predicting loan default than statistical and traditional machine learning models. In the second rank, i found that RF has a significantly higher prediction accuracy compared to other techniques employed. Our empirical research suggests that the deep neural network is the most effective method for identifying loan defaults among customers, which can aid in managerial decision-making.

To note, in this empirical task, i used 20% of the sample to test the forecasting accuracy and classifier's quality of the models. For the training process, a deep feed-forward network with three hidden layers is adopted, with a sigmoid activation function for the hidden layers and a linear activation function for the output layer.

6. Conclusions

Financial credit institutions have always been very concerned about forecasting loan defaults in order to make the right lending decisions. Our objective of this study is to create a useful model for categorizing credit applicants in order to accurately predict their financial difficulties. In this work I compare several machine learning methods for assessing credit risk in the Tunisian credit dataset. Several classification algorithms, including LDA, LR, DT, SVM, RF, and DNN, have been used to implement and evaluate these. From the above analysis, using the DNN methodology provides higher accuracy in credit risk evaluation. The RF model performs better than other machine learning and statistical methods.

As suggested by Lyn et al. (2002), the robustness of classification methods in credit scoring can be examined better with more data, and use more datasets. We mainly focused on popular machine learning methods and explored novel techniques that can incorporate the foundational algorithms. I covered a lot of recurrent neural networks and selective ensemble methods. Evaluation of classification algorithms is another area that has received much attention recently, and best practices suggest the use of different types of evaluation metrics. There are three broad types of evaluation metrics: Threshold, ranking, and probabilistic metrics. Most studies, including our own, have focused on threshold metrics (e.g., accuracy and F-measure), ranking methods and metrics (e.g., ROC analysis and AUROC), and have left out other crucial measures of classifier performance.

The results of the empirical study demonstrated that DNN is an excellent tool for researching financial defaults in credit institutions. Compared to past work, this study incorporates machine learning models used in predicting loan defaults. Unlike traditional methods, these models can learn complex, non-linear relationships between various data points and loan defaults. Machine learning models can continuously learn and improve as they are exposed to more data. They can potentially discover complex relationships between various factors that might not be evident through traditional statistical methods. Further technical aspects could be explored in future research, and the current study's results still require discussion and improvement, particularly in order to create a new model that performs better than earlier models, such as identifying new or counter intuitive insights and finding a significant and meaningful new variable.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The author states that there is no conflicts of interest with the publication of this manuscript.

References

Breiman L (2001) Random forests. *Mach Learn* 45: 5–32. https://doi.org/10.1023/A:1010933404324

Deng L, Yu D (2014) Deep Learning: Methods and Applications. *Found Trends Signal Proc* 7: 197–387. http://dx.doi.org/10.1561/2000000039

Fuster A, Goldsmith Pinkham P, Ramadorai T, et al. (2022) Predictably unequal? The effects of machine learning on credit markets. *J Financ* 77: 5–47. https://doi.org/10.1111/jofi.12915

Giudici P, Hadji-Misheva B, Spelta A (2020) Network based credit risk models. *Qual Eng* 32: 199–211. https://doi.org/10.1080/08982112.2019.1655159

Le Cun Y, Bengio Y, Hinton GE (2015) Deep learning. *Nature* 521: 436–444. https://doi.org/10.1038/nature14539

Lyn T, David Edelman, Jonathan Crook (2002) Credit Scoring and its Applications. *Mathematical Modeling and Computation*. https://doi.org/10.1137/1.9780898718317

Liu RL (2018) Machine learning approaches to predict default of credit card clients. *Modern Econ* 9: 18–28. https://doi.org/10.4236/me.2018.911115

Lien CH, Yeh IC (2009) The Comparisons of Data Mining Techniques for the Predictive Accuracy of Probability of Default of Credit Card Clients. *Expert Syst Appl* 36: 2473–2480. https://doi.org/10.1016/j.eswa.2007.12.020

Mellisa K (2020) Credit Scoring Approaches guidelines. *World Bank Group*, Washington, DC, USA.

Mestiri S (2024) Financial Applications of Machine Learning Using R Software. *SSRN Electronic J*. https://dx.doi.org/10.2139/ssrn.4716425

Mestiri S, Farhat A (2021) Using Non-parametric Count Model for Credit Scoring. *J Quant Econ* 19: 39–49. https://doi.org/10.1007/s40953-020-00208-w

Pepe MS (2000) Receiver operating characteristic methodology. *J Am Stat Assoc* 95: 308–311. https://doi.org/10.2307/2669554

Giudici P (2001) Bayesian data mining, with application to credit scoring and benchmarking. *Appl Stoch Models Bus Ind* 17: 69–81. https://doi.org/10.1002/asmb.425

Quinlan JR (1986) Induction of decision trees. *Mach Learn* 1: 81–106.

Tran K, Duong T, Ho Q (2016) Credit scoring model: A combination of genetic programming and deep learning, In: *2016 future technologies conference (ftc) IEEE*, 145–149.

Schmidhuber J (2015) Deep learning in neural networks: An overview. *Neural Networks* 61: 85–117. https://doi.org/10.48550/arXiv.1404.7828

Stefan Lessmann, Bart Baesens, Hsin-Vonn Seow, et al. (2015) Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *Eur J Oper Res* 247: 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

Vapnik V (1998) The nature of statistical learning theory. *New York: Springer.*

Woo H, Sohn SY (2022) A credit scoring model based on the Myers–Briggs type indicator in online peer-to-peer lending. *Financ Innov* 8: 1–19. https://doi.org/10.1186/s40854-022-00347-4

Wang C, Han D, Liu Q, et al. (2018)   A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism LSTM.   *IEEE Access* 7: 2161–2168. https://doi.org/10.1109/ACCESS.2018.2887138.

AIMS Press