Ghoujdam, Mousaab El Khair et al.

Article

Consumer credit risk analysis through artificial intelligence: a comparative study between the classical approach of logistic regression and advanced machine learning techniques

Cogent Economics & Finance

Provided in Cooperation with:
Taylor & Francis Group

Cogent Economics & Finance

ISSN: 2332-2039 (Online) Journal homepage: www.tandfonline.com/journals/oaef20

Consumer credit risk analysis through artificial intelligence: a comparative study between the classical approach of logistic regression and advanced machine learning techniques

Mousaab El Khair Ghoujdam, Rachid Chaabita, Oussama Elkhalfi, Kamal Zehraoui, Hicham Elalaoui & Salwa Idamia

To cite this article: Mousaab El Khair Ghoujdam, Rachid Chaabita, Oussama Elkhalfi, Kamal Zehraoui, Hicham Elalaoui & Salwa Idamia (2024) Consumer credit risk analysis through artificial intelligence: a comparative study between the classical approach of logistic regression and advanced machine learning techniques, Cogent Economics & Finance, 12:1, 2414926, DOI: 10.1080/23322039.2024.2414926

To link to this article: https://doi.org/10.1080/23322039.2024.2414926

Published online: 20 Oct 2024.

COGENT ECONOMICS & FINANCE
2024, VOL. 12, NO. 1, 2414926
https://doi.org/10.1080/23322039.2024.2414926

ECONOMETRICS & DATA ANALYTICS | RESEARCH ARTICLE
OPEN ACCESS
Check for updates

Consumer credit risk analysis through artificial intelligence: a comparative study between the classical approach of logistic regression and advanced machine learning techniques

Mousaab El Khair Ghoujdam[a], Rachid Chaabita[a], Oussama Elkhalfi[b], Kamal Zehraoui[a], Hicham Elalaoui[a] and Salwa Idamia[c]

[a]University Hassan II Casablanca, FSJES, Casablanca, Morocco; [b]Sidi Mohamed Ben Abdallah de Fes, FSJES, Fez, Morocco; [c]University Hassan II Casablanca, ESTC, Casablanca, Morocco

ABSTRACT
This research article presents a comparative analysis between logistic regression as a traditional method, artificial neural networks (ANNs), and decision tree as machine learning techniques for predicting credit risk. It meticulously examines and evaluates these three methods, elucidating their contextual nuances and practical implications. The study utilizes consumer credit data comprising 9766 credit applications. The objective is to explore and evaluate the three models using various performance metrics, including accuracy, sensitivity, F1 score, and area under the ROC curve. Results demonstrate the superior performance of ANNs and decision trees over logistic regression across all metrics evaluated. This study provides compelling evidence endorsing ANNs and decision tree as more effective methods for credit risk prediction, thereby opening avenues for further exploration and application in this domain. However, a limitation of this study lies in its focus solely on three prediction methods, whereas considering additional approaches could have offered a more comprehensive perspective.

ARTICLE HISTORY
Received 17 April 2024
Revised 30 September 2024
Accepted 5 October 2024

KEYWORDS
Artificial neural network; decision tree; logistic regression; credit risk; machine learning

SUBJECTS
Finance; Econometrics; Artificial Intelligence

IMPACT STATEMENT
This study provides valuable insights into the comparative performance of logistic regression, artificial neural networks (ANNs), and decision trees for credit risk prediction, based on a dataset from a Moroccan bank. The results clearly demonstrate the superiority of machine learning techniques, such as ANNs and decision trees, in terms of predictive accuracy and robustness, compared to traditional methods. By proving that these models outperform logistic regression across various performance metrics, this research contributes to improving credit risk assessment practices in the Moroccan financial sector. The implications of this study extend to risk management strategies, where the integration of advanced machine learning techniques can significantly enhance the reliability of forecasting tools.

1. Introduction

Credits play a fundamental role in driving economic activity by serving various critical functions. They facilitate access to financing for businesses, thereby fostering innovation and promoting economic growth (Gerken et al., 2015). Additionally, credits such as auto loans and mortgages contribute to maintaining demand stability and provide individuals with the flexibility to manage their expenses over time. Moreover, credits play a pivotal role in supporting the growth of startups and small businesses by providing them with essential funds for development (Kerr & Nanda, 2009). Undoubtedly, credits are indispensable for development and progress. However, it is incumbent upon banks to rigorously assess credit applications to mitigate the risk of imprudent risk assessment or credit distribution.

This comprehensive evaluation commences with the utilization of credit risk prediction techniques aimed at gauging the likelihood of repayment for each borrower. The paramount significance of our

CONTACT Mousaab El Khair Ghoujdam moussabgh1@gmail.com University Hassan II Casablanca, FSJES, Casablanca, Morocco.

research lies in its potential to enhance financial risk management practices. By furnishing a comprehensive comparative analysis of the performance of logistic regression artificial neural networks (ANNs) and decision tree in credit risk prediction.

The choice of this topic is justified by several motivations. Firstly, the growing importance of the financial industry and banking institutions in the global economy underscores the need to develop robust tools for effectively assessing and managing financial risks, particularly credit risk (Kimiagari & Baei, 2022). Additionally, with the advent of artificial intelligence and machine learning, there is increasing interest in applying these advanced technologies in the field of credit risk assessment to test their performance in this domain. Lastly our study seeks to inform the decision-making processes of financial institutions and fortify their capacity to evaluate and manage risk effectively. By offering valuable insights into the most suitable methodology for predicting credit risk, we aspire to positively influence risk management practices and policies, thereby contributing to the stability and sustainability of the financial sector.

A thorough review of the literature reveals an ongoing debate regarding the relative efficacy of approaches in predicting credit risk. Logistic regression has traditionally been a popular choice for risk assessment. However, artificial neural networks have emerged as a cutting-edge approach with the potential to revolutionize credit scoring models. A study conducted by Hu and Su (2022), utilized a neural network model with 5114 nodes was employed for credit risk assessment in banks, resulting in improved prediction accuracy.

The objective of this research is not only to explore but also to ascertain the most effective approach for credit risk prediction by examining tree methods. Logistic regression, considered a traditional approach, artificial neural networks and decision trees, acknowledged as an innovative method in artificial intelligence (AI). We shall not confine our assessment solely to performance evaluation but also consider other characteristics, particularly in terms of fairness, explainability, and robustness. The 2008 financial crisis underscored banks inability to accurately anticipate risk, resulting in bankruptcies, layoffs, and global instability. In light of this scenario and from this standpoint, we formulate our problem statement: To what extent can artificial intelligence methods such as Artificial Neural Networks and decision tree serve as a viable alternative to traditional approaches in anticipating credit risk?

To achieve the objective of our study, we will commence by reviewing existing research on logistic regression, decision tree and artificial neural networks. Subsequently, we will meticulously examine our database comprising 9766 consumer credit applications through comprehensive univariate and bivariate analyses Additionally we will provide insights into the methodology employed in constructing the scoring model. Finally, we will present the outcomes achieved using the three methods, enabling an assessment of their performances.

2. Literature review and methodology

2.1. Literature review

Artificial Neural Networks are sophisticated models crafted to emulate the functionality of the human brain, processing information through interconnected nodes organized in layers. Their versatility extends across various sectors, including manufacturing, where they excel in tasks requiring pattern recognition (Abiodun et al., 2019). ANNs mimic the behavior of neurons in the brain, endowing them with resilience and enabling them to address challenges such as speech recognition, medical diagnosis, financial forecasting, and more (Ali et al., 2019). The efficacy of ANNs relies heavily on their training data, which allows them to continuously enhance their accuracy over time. As these learning algorithms evolve, they have become indispensable tools in the realms of computer science and artificial intelligence. They excel in classifying and grouping data, finding practical applications in diverse fields such as biology, physics, and finance. ANNs are now recognized alongside traditional statistical approaches, with successful implementations ranging from intrusion detection to modeling energy systems and identifying lung cancer. Their remarkable accuracy in data categorization has been demonstrated in various studies (Hodo et al., 2016; Elsheikh et al., 2019; Nakano et al., 2019; Angelini et al., 2008).

The essence of ANNs lies in their autonomous capacity to learn and discern the relationships between variables by analyzing sample data, akin to human reasoning. They establish connections between

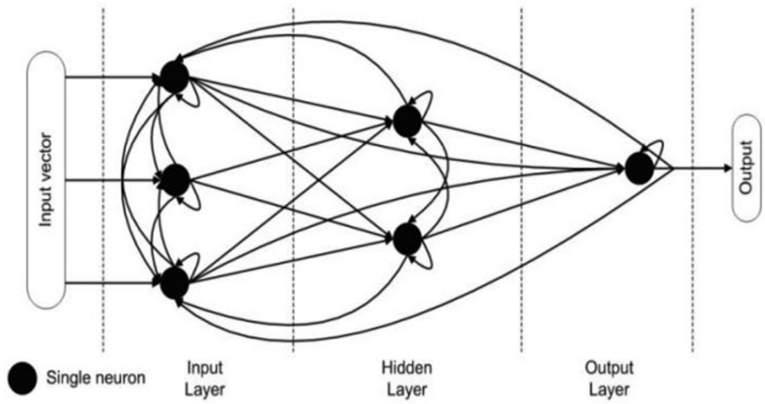

Figure 1. Neural network architecture (Krenker et al., 2011).

inputs (data) and outputs (results) based on the assumption of a linear relationship, akin to the functioning of credit risk detection. Constructing a network capable of emulating reasoning involves three stages: the input phase, the transfer function phase, and the learning phase (see Figure 1).

Artificial neural networks typically consist of an input layer for receiving data, hidden layers for computations and transformations, and an output layer for presenting the network's predictions or decisions (El-Jerjawi & Abu-Naser, 2018). These models are comprised of interconnected nodes or functions with adjustable weights that are iteratively tuned during training to minimize the error between predicted and actual outputs (Su & Wang, 2019).

Hu and Su explored credit risk assessment in banks using a neural network model. They developed a set of credit risk evaluation indicators by selecting 14 metrics and applying cluster analysis and factor analysis to determine the credit rating of sample data. Moreover, they compared the prediction accuracy of two models and three ANNs models. Their findings revealed that the three-layer perceptron neural network model with 5114 nodes achieved a prediction accuracy of 95.8% for the test dataset (Hu & Su, 2022). Gupta and Goyal (2018) explores credit risk prediction through the application of ANNs algorithms. Their study aims to demonstrate the effectiveness of neural network models in accurately predicting creditworthiness. By comparing the performance of neural network models with linear regression models, they highlight the suitability of both approaches for different applications, emphasizing their ability to provide precise credit risk assessments. Mahbobi et al. (2023) propose an effective classification algorithm to address the issue of imbalanced data, particularly in the context of default payments. By combining resampling techniques with machine learning models such as SVM and neural networks, the study significantly enhances the accuracy of predictions, achieving a rate of 98.6%. Similarly, Dissananayake et al. (2007) focus on credit risk assessment within the hotel industry of Sri Lanka using artificial neural network technology. Their research aims to develop a novel approach to credit risk evaluation specifically tailored to the hotel sector. By achieving high accuracy levels for credit risk assessment in five-star hotels, their study showcases the potential of neural network technology in non-traditional industries. Additionally, Islam et al. (2009) investigates the application of artificial intelligence, specifically artificial neural networks, for credit risk assessment in the context of credit card scoring. Their research emphasizes the importance of accurate credit scoring models in minimizing financial losses from defaulters. By highlighting the superiority of credit scoring models over judgmental decisions in credit assessment, they advocate for the adoption of advanced modeling techniques in credit risk management. The advantage of neural networks lies in their adaptability and capacity to discern intricate patterns within extensive datasets. This attribute proves particularly beneficial when confronted with large volumes of data, where traditional models often struggle to yield significant performance improvements. Unlike conventional approaches, artificial neural networks excel at uncovering deeper correlations between input features and output variables as the dataset size expands. However, a notable drawback is their opaque nature, rendering artificial neural networks as black-box models, as the patterns they identify are not easily interpretable by humans. Essentially, ANNs learn by memorizing patterns within the data through the estimation of a vast array of parameters. Moreover, ANNs leverage nonlinear structures via activation functions, thereby enhancing the model's flexibility (Bazarbash, 2019).

Logistic regression is a modeling technique used to predict outcomes by estimating the probability of specific events or behaviors. It finds widespread application in various fields such as health, medicine, business, and finance, particularly when dealing with variables having two categories (Hosmer et al., 2013).

The logistic regression model establishes a function linking a variable Y with one or more variables X_1, $X_2 \ldots$ X_n, represented as:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_n X_{n,i} + \varepsilon_i, \tag{1}$$

for $i = 1,2, \ldots, N$, where in this equation, Y_i is the dependent variable for each observation(i). The β_j coefficients (where $j = 0,1,\ldots,n$) quantify the effect of the explanatory variables $X_{j,i}$ on Y_i. The term ε_i represents the random error for each observation, which follows a logistic distribution.

The coefficients β are determined using the likelihood method and help explain variations in the target variable, with a negative coefficient indicating that an increase in X leads to a decrease in Y. Unlike in linear relationships, the logistic regression model does not directly depict a connection between Y and X. The logit formula is employed to convert values into probabilities to measure variation:

$$P(n) = \frac{e^{\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_n X_{n,i}}}{1 + e^{\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_n X_{n,i}}}, \tag{2}$$

for $i = 1,2, \ldots, N$, where The equation represents the probability P(n) of an event, X(1,i), X(2,i), … ,X(n,i) are the explanatory variables, and β_0, $\beta_1, \ldots, \beta_n$ are their respective coefficients.

In regression analysis, the model predicts the likelihood of an event occurring based on one or more variables, ensuring that the predicted probabilities fall within the range of 0 to 1 (Schober & Vetter, 2021). Logistic regression is widely utilized in research analysis due to its ability to interpret relationships between categorical outcomes and explanatory variables (Christodoulou et al., 2019). One notable application of logistic regression in credit risk prediction is exemplified in Ohlson's seminal work in 1980, where a regression model was developed using data from defaulted and financially stable companies. the assumption of linearity in the regression model may be constraining when studying complex phenomena. To address this limitation, Zhang and Lin (2003) introduced the "Parametric Logistic Regression Model" in 2003, which replaced linear relationships with parametric functions to more accurately represent variable effects. Research findings indicate that logistic regression outperforms discriminant analysis in predicting default risk for small and medium-sized enterprises (SMEs), achieving an accuracy rate of 76.7% according to Khemais et al. (2016). Moreover, studies comparing the performance of logistic regression with linear discriminant analysis in credit risk prediction have shown enhanced model predictions when sectoral effects are taken into consideration. Several studies have investigated the application of logistic regression in predicting credit risk for consumer loans. Costa e Silva et al. (2020) identified that loan distribution, loan duration, customer age, and credit card ownership influence default risk. Ni (2010) proposed a model that combines self-organizing maps and logistic regression to assess consumer credit risk. Additionally, Mushunje (2021) utilized logistic regression to identify potential defaulters and manage credit risks, demonstrating its effectiveness with significant precision. Collectively, these studies underscore the potential of logistic regression in predicting credit risk for consumer loans.

A decision tree represents a pivotal instrument within the realm of machine learning, serving the dual purpose of prediction and classification, particularly within the domain of supervised learning paradigms. This sophisticated structure provides either a graphical or textual depiction of data, rendering complex information in an intelligible manner Chen and Li (2010). A decision tree adopts a hierarchical structure like a flowchart that starts from a root node, progresses to lower nodes through possible states or decisions (represented as a branch), and ends at the terminal node that shows the consequence of the entire branch. Decision trees have the advantage of being used for both regression and classification models.

Foundational to the establishment of such trees are venerable algorithms like Classification and Regression Tree (CART) and Iterative Dichotomiser (ID3).

ID3, an algorithm of commendable simplicity in the arena of learning decision trees, was conceived by Quinlan (1986). Its cardinal principal hinges upon the methodical top-down exploration of provided datasets, wherein the construction of a decision tree unfolds, endeavoring to unveil the salience of each attribute in classifying a given dataset. Central to this process is the algorithm's adept discernment of attributes, predicated upon the metric of information gain, as represented by Eqs. (3) and (4)

$$H(S) = -\sum_{i=1}^{C} Pi \ \mathrm{Log2}(Pi), \tag{3}$$

where H(S) measures the uncertainty in dataset S. C represents the number of classes in the dataset. For each class i, Pi is the probability of that class, calculated by dividing the number of instances of class i by the total number of instances. The term log2(Pi) is the logarithm base 2 of Pi, reflecting the amount of information each class contributes.

$$IG(A) = H(s) - \sum_{v \epsilon Values(A)} \frac{\lfloor Sv \rfloor}{\lfloor S \rfloor} H(Sv), \tag{4}$$

where IG(A) measures the reduction in uncertainty about the dataset S due to the attribute A. H(S) represents the entropy of the original dataset S. For each value v of attribute A, $\frac{\lfloor Sv \rfloor}{\lfloor S \rfloor}$ is the proportion of instances in subset Sv compared to the total instances in S. H(Sv) is the entropy of subset Sv.

Conversely, The Classification and Regression Tree (CART) emerges as a statistical innovation pioneered by Breiman (2017). Primarily, CART functions as a classification apparatus, adept at segregating objects into two or more distinct populations. Remarkably versatile, it wields the capability to analyze both categorical and continuous data, leveraging a unified technological approach Lee et al. (2006), as illustrated in Eqs. (5) and (6)

$$Gini(k) = 1 - \sum_{i=1}^{c} (Pi, k)2, \tag{5}$$

where Gini(k) measures the impurity or disorder of a dataset k. cc represents the number of classes in the dataset. Pi, k is the probability of occurrence of class i in dataset k, calculated as the proportion of instances belonging to class i within k.

$$Gini\_weight(A) = \sum_{v \in Value(A)} \frac{\lfloor Sv \rfloor}{\lfloor S \rfloor} Gini(Sv), \tag{6}$$

where Gini_weight(A) calculates the overall impurity of a dataset split by attribute A. Here, $\frac{\lfloor Sv \rfloor}{\lfloor S \rfloor}$ Proportion of instances in subset Sv relative to the total dataset S, and Gini (Sv) is the Gini impurity of Sv.

Decision tree models have been extensively researched for credit risk prediction (Jiaqi, 2023; Wang et al. 2023). These models have demonstrated promising outcomes in credit risk management owing to their capability to handle large datasets and offer interpretable results. Studies have compared decision trees with alternative machine learning algorithms such as support vector machines and logistic regression, showcasing the competitive performance of decision trees (Gang et al., 2022). Moreover, integrating supply chain information into decision tree ensemble models has significantly enhanced credit risk prediction accuracy, surpassing traditional and ensemble models.

Decision-tree models offer the advantages of interpretability, especially for small trees, aiding in clear decision-making guidelines for credit rating training. They can also handle multi-output problems, crucial for assessing liquidity risk in risk management. However, decision trees have practical drawbacks, such as complex and overfitted trees, suboptimal solutions, and bias in cases where certain classes dominate the sample, posing challenges in credit analysis, particularly for underserved populations (Bazarbash, 2019).

Previous research suggests a unanimous agreement regarding the enhanced predictive accuracy of ensemble credit scoring models derived from machine learning. Nevertheless, this heightened accuracy is accompanied by a trade-off, while individual scoring models like logistic regression are transparent and can delineate the impact of each explanatory variable on credit scores, ensemble methods operate as black boxes, lacking the ability to elucidate the factors influencing credit scores to their users. (Bracke et al., 2019; Giudici et al., 2020).

Through a literature review, three hypotheses have been formulated for our research.

H0A: The primary hypothesis suggests that the choice of credit prediction technique significantly impacts prediction performance.

H0B: A secondary hypothesis is that the use of machine learning techniques in credit prediction will lead to a significant improvement in performance compared to classical methods. This hypothesis assumes that AI algorithms can leverage more complex models and non-linear data for more accurate credit risk prediction.

H0C: The third hypothesis proposes that ANNs will decision tree in terms of performance.

The tree hypotheses have been constructed based on a thorough review of existing literature in the field of credit prediction, taking into consideration the recent advancements of Artificial Neural Networks showcasing their remarkable capabilities in credit prediction. These hypotheses are grounded in observed trends and the potential of Artificial Neural Networks within the scope of this study.

2.2. Research methodology and presentation of data

The methodology employed in this study encompasses several crucial steps. Initially, we will construct a credit scoring model drawing insights from Thomas's model and research conducted by Carling et al. Subsequently, we will conduct a comprehensive examination of individual variables within our dataset, including age, income, job tenure, among others. Finally, we will perform a bivariate analysis to evaluate the relationships between explanatory variables and the likelihood of payment default. The primary objective of this article is to compare three methods for assessing the risk associated with consumer credit. In the development of our credit scoring model, we have referenced Thomas's model and integrated findings from Carling and colleagues' research. Thomas's analysis in 2000 revealed that conventional credit scoring methodologies primarily relied on assessments involving factors such as the borrower's reputation (reflecting their credit history), capital (representing collateral for the loan amount and tenure of employment), and capacity (evaluated through factors like income, debt-to-income ratio, and age) (Thomas et al., 2001). Additionally, a study conducted by Carling, and Roszbach explored the influence of status on the risk of non-repayment within the consumer credit domain (Carling et al., 1998).

2.2.1. Data

The database utilized in this study, sourced from a Moroccan bank, spans a two-year period, covering data from 2017 to 2018. Initially comprising 10,230 credit files, meticulous data cleaning procedures were conducted to address any missing variables, resulting in a refined sample of 9766 credit files. Within this dataset, 5325 files (54.53%) exhibit no payment defaults, while 4441 files (45.47%) indicate instances of payment defaults. For the construction of scoring models, we will use Stata version 15 for logistic regression and Python for artificial neural network construction.

2.2.2. Variable selection

The variable target is the prediction of default risk evaluated by the bank, indicating whether an individual is likely to default, leading to their credit application being rejected by the bank. Conversely, if there is no default risk, the bank will approve the credit application. Table 1 presents all the variables used to explain our variable along with their data types.

Utilizing the Stata software, we have proceeded with the conversion of the qualitative variables. This step involves transforming variables into numerical representations, which are necessary for the logistic regression model to utilize them effectively. This process is crucial for integrating these variables into our prediction model. For the dependent variable, default risk, the coding is structured as follows: 44.41% of the data points are labeled as 0, signifying the presence of default risk, whereas 53.25% of the data points are labeled as 1, denoting the absence of default risk.

Additionally, qualitative variables will be encoded according to the structure provided in Table 2. This encoding allows us to represent categorical variables in a numerical format, enabling the logistic regression model to analyze them effectively (see Table 2).

Table 1. Explanatory variables, data types, and descriptions.

Variable	Descriptions	Type de la variable
Age	Age of the individual	Quantitative
Income	Individual's income in MAD	Quantitative
Collaterals	Presentation of collateral	Qualitative
Seniority	Seniority in the last job	Quantitative
Credit Type	Credit allocation	Qualitative
Marital Status	Marital status (married or unmarried)	Qualitative
Debt Ratio	Debt-to-income ratio	Quantitative
Credit History	Credit history	Quanitative
Credit Amount	Credit amount	Quantitative

Table 2. The coding of qualitative variables.

	Label	Numeric	Freq
Marital Status	Single	1	2976
	Married	2	6790
Guaranteed	No	1	3773
	Yes	2	5993

2.2.3. Univariate and bivariate analysis

Upon individual analysis of the variables, it is discerned that the average age of the individuals stands at 28 years and approximately two months and nine days. However, it is noteworthy that half of these individuals fall below the age of twenty-seven. The average income is recorded at 44,682 MAD, with a median salary of 4000 MAD. Intriguingly, half of the individuals in our study exhibit an income below 36,000 MAD. Regarding work experience, the average tenure is approximately 3 years and 10 months, with the longest tenure reaching 20 years. The average credit debt ratio is approximately 17.66%, with 50% of applicants displaying a debt ratio below 15.71%. Analysis of the number of credit applications reveals a range from none to up to 5 applications, with half of the participants having made no more than three credit requests. Regarding marital status, 69.52% of participants are married, while approximately 30.47% are single. In terms of guarantees, 61.37% of the sample possess collateral that can be utilized by the bank in the event of payment default, whereas 38.63% do not report having any guarantees.

For our study's bivariate analysis, two methods are employed, the chi-square independence test for qualitative variables and a mean comparison test to assess relationships between a qualitative variable and a quantitative variable (see Table 3). Our primary objective is to evaluate any associations between the variables and the understanding of "default risk". The explanatory variables, whether qualitative or quantitative, are retained for analysis, except for the "credit type" variable, as it was deemed to have no impact on explaining the dependent variable.

3. Results and discussion

3.1. Logistic regression results

3.1.1. Model presentation

The chosen approach to assess credit risk revolves around a logistic regression model aimed at predicting the probability of default. The model is formulated as follows:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_8 X_{8,i} + \varepsilon_i, \tag{7}$$

for $i = 1,2, \ldots, N$, where Y_i represents the dichotomous dependent variable , which takes the value 0 in the case of credit default and 1 otherwise. The terms $X_{1,i}, X_{2,i}, \ldots, \beta_8 X_{8,i}$ represent the eight explanatory variables: collateral, marital status, credit amount, debt ratio, credit history, Seniority, income, and age. The associated coefficients, $\beta_1, \beta_2, \ldots, \beta_8$, measure the impact of these variables on Y_i. The error ε_i follows a logistic distribution.

The objective of this logistic regression analysis is to scrutinize a qualitative variable termed "decision", which encompasses two possibilities, "credit approval" and "credit rejection". This variable is influenced by eight types of variables, both qualitative and quantitative, as delineated in the model. It is noteworthy that the requisite condition for applying the Logit model to qualitative variables is met in this analysis.

3.1.2. Preliminary tests and statistical results analysis

Before delving into logistic regression analysis, it is crucial to conduct a correlation test to identify and address any potential correlation issues between the variables. The results of the correlation test reveal no significant correlation among the variables, indicating that all the variables listed in the table can be included in the analysis. The objective of this logistic regression analysis is to analyze a qualitative variable called "decision", which comprises two possibilities, "credit approval" and "credit rejection". This variable is influenced by eight types of variables, both qualitative and quantitative. It is worth noting

Table 3. Statistical tests for bivariate analysis.

Variable	*p*-value	The selected hypothesis
Collaterals	0.000	H1
Credit Type	0.201	H0
Marital Status	0.000	H1
Credit Amount	0.000	H1
Debt Ratio	0.0001	H1
Credit History	0.008	H1
Seniority	0.000	H1
Income	0.000	H1
Age	0.0002	H1

Table 4. Logistic regression results.

Variable	Coef	Robust std. Err	P>|z|
Age	−0.0310	0.0022	0.000
Income	2.11×10^{-5}	1.28×10^{-6}	0.000
Seniority	0.0958	0.0043	0.000
Credit Amount	6.43×10^{-5}	5.97×10^{-6}	0.000
Debt Rate	−4.718	0.2301	0.262
Credit History	−0.0153	0.0136	0.000
Guarantee	0.1133	0.0291	0.000
Marital Status	−0.1488	0.0298	0.000
const	2.134	0.125	0.000
Statistical Metrics	Number of obs	9766	
	Wald chi2 (8)	1207.49	
	Prob > chi2	0.0000	
	Pseudo R2	0.1146	

that the necessary condition for applying the Logit model to qualitative variables is satisfied in this analysis.

From a statistical perspective, the *p*-value associated with our model is less than 5%, confirming the overall relevance of our model and indicating that all selected independent variables play a significant role in explaining the dependent variable. Furthermore, the *p*-value corresponding to the z-statistic test (which assesses the significance of each variable individually) is statistically significant at a 5% threshold, except for the credit history variable (see Table 4).

3.1.3. Interpretation of coefficients

In the realm of logistic regression analysis, the discernment of coefficient signs provides nuanced insights into the determinants shaping the likelihood of default. Age, bearing a negative coefficient, intimates that advancing age correlates with a diminished propensity for default occurrences. Conversely, the positive coefficient associated with income underscores the protective effect of higher income levels against default risk. Conversely, a deleterious impact on default likelihood emerges with the negative coefficient attributed to the debt ratio, signifying that escalating debt ratios heighten the susceptibility to default events. Similarly, the positive coefficient adorning credit amount underscores a positive relationship between larger credit amounts and the probability of default. Marital status, epitomized by a negative coefficient, conveys that marital bonds mitigate the likelihood of default. Conversely, the positive coefficient ascribed to collateral illuminates its association with an augmented probability of default. Credit history, with its negative coefficient, signifies the mitigating effect of a favorable credit history on default risks. Lastly, the positive coefficient characterizing job seniority suggests that prolonged job tenure mitigates the likelihood of default occurrences.

3.1.4. Assessment of model performance

Now, let us assess the performance of logistic regression in classifying borrowers based on their risk of payment default. The results indicate that the model is 64.23% accurate, meaning it predicts outcomes correctly for 64.23% of the total dataset. The sensitivity, or recall, is at 61.6%, demonstrating how well the model can accurately identify around 61.6% of cases. The F1 score, standing at 64.5%, provides an assessment between precision and recall, with values indicating better overall performance. Lastly, the

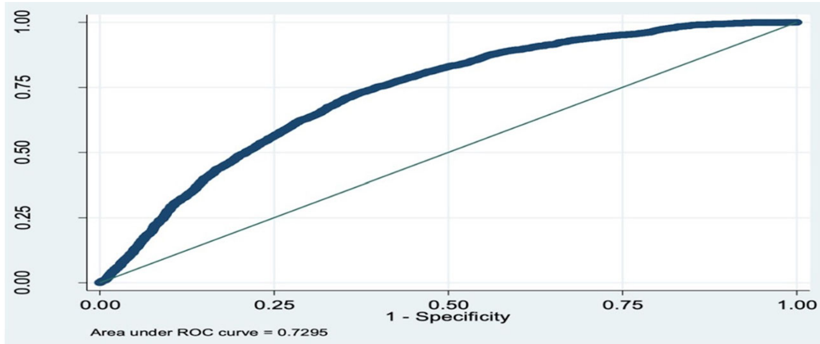

Figure 2. ROC curve of logistic regression.

Area Under the ROC Curve (AUC) is noted at 72.9%, demonstrating the model's ability to differentiate between negative and positive cases (see Figure 2).

3.2. The results of the neural network

3.2.1. Establishing a Python neural network: methodical approach and architectural design

To establish a neural network in Python, we must embark on several steps. Initially, it is imperative to import and prepare the data. This phase entails importing the data into Python using the NumPy library, as well as handling categorical variables. Additionally, data normalization is conducted to ensure all variables have a uniform scale, thereby facilitating the model's learning process. Subsequently, we proceed with partitioning the dataset into training and testing sets. The training set is utilized to adjust the weights of the neural network during the learning phase, while the testing set is reserved for evaluating the model's performance on unseen data. In our scenario, the neural network learns from 80% of the data and subsequently evaluates its performance on the remaining 20%. Lastly, the determination of the neural network's architecture, including the number of hidden layers, the number of neurons in each layer, activation functions, and optimization algorithms.

In our case, the architecture of our neural network comprises an input layer, two hidden layers, and an output layer. The input layer is designed to accommodate eight variables influencing our target variable in the database. The first hidden layer consists of 128 neurons, each activated by a Rectified Linear Unit (ReLU) activation function. Similarly, the second hidden layer comprises 64 neurons, also activated by the ReLU function. These layers facilitate the learning of crucial features for binary classification, where we predict either credit risk (0) or the absence of credit risk (1). Finally, the output layer consists of a single neuron activated by the 'sigmoid' function (see Table 5).

3.2.2. Assessment of model performance

The thorough evaluation of this model reveals impressive and remarkable results. The accuracy of 87.1% indicates the model's ability to accurately predict the credit risk class for the majority of examples. Furthermore, a recall rate of 91.5% means that the model accurately identifies the vast majority of default cases, which is crucial in domains where classification errors can have significant consequences. The F1 score, which combines precision and recall, reaching around 89.3%, highlights a balanced performance of the model in capturing both positive and negative cases effectively. This demonstrates the algorithm's robustness in managing positive cases while maintaining a low rate of false positives. Moreover, the AUC of 95% is a significant indicator of the model's ability to discriminate between credit risk classes (see Figure 3). A high AUC value suggests that the model is capable of effectively separating positive examples from negative ones, thereby bolstering confidence in its predictive capabilities.

Overall, these results testify to the efficacy of the model in accurately predicting credit risk, making it a valuable tool for financial institutions and decision-makers to assess and manage credit risks.

Table 5. Summary of model architecture.

Layer (type)	Output shape	Param#
Dense	(None,128)	1152
Dense_1	(None,64)	8256
Dense_2	(None,64)	65

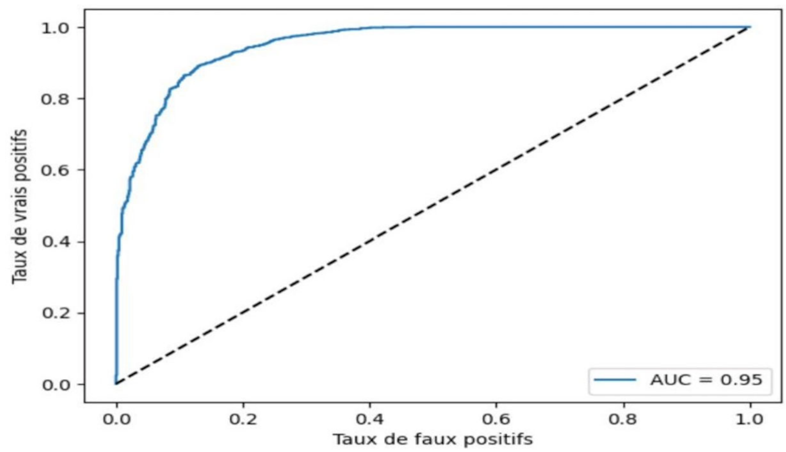

Figure 3. ROC curve of ANN.

3.3. The results of decision tree

3.3.1. Model presentation

The model utilized is the decision tree, a technique commonly employed in supervised machine learning. In this approach, data features are represented by nodes of the tree, subjected to tests, while branches lead to the different values these features can take. The leaves of the tree are associated with the various final categories. The primary objective is to develop a model capable of accurately predicting the category to which an individual belongs, whether they present a credit risk or not, even if it hasn't been observed previously. The construction and evaluation of this model are carried out using the Python programming language, with specific libraries such as Pandas for importing the database and Scikit-learn for building the decision trees. The CART algorithm (Classification and Regression Trees) is employed in this model to minimize error and classify individuals correctly. It relies on the Gini criterion to measure node purity and guide data division at each stage of tree creation. The necessary steps for building and evaluating the tree include importing the required libraries, data preparation by converting categorical variables into numerical ones, selecting explanatory variables, dividing the training and test samples, and creating the decision tree using the Scikit-learn library.

3.3.2. Assessment of model performance

The evaluation results of the decision tree model are promising, demonstrating a precision of 89.4%. This indicates that approximately 90% of the model's predictions are correct. Furthermore, the recall rate reaches 89.96%, signifying the model's ability to correctly identify the vast majority of positive instances. The F1-score, a combined measure of precision and recall, also shows satisfaction at 89%. Finally, the AUC of the ROC is assessed at 88% (see Figure 4). Indicating a strong capacity of the model to distinguish between positive and negative classes. These results suggest that the decision tree is a promising choice for accurately predicting credit risk, offering a balance between precision, recall, and discriminative ability.

3.4. Discussion of the results

In examining the performance of the various evaluated models, notable distinctions emerge. Firstly, in terms of accuracy, the decision tree surpasses the others with an accuracy of 89.4%, closely followed by the neural network at 87.76%, while logistic regression exhibits the lowest accuracy at 64.23%. Transitioning to recall, the decision tree remains in the lead with a recall of 89.96%, followed closely by

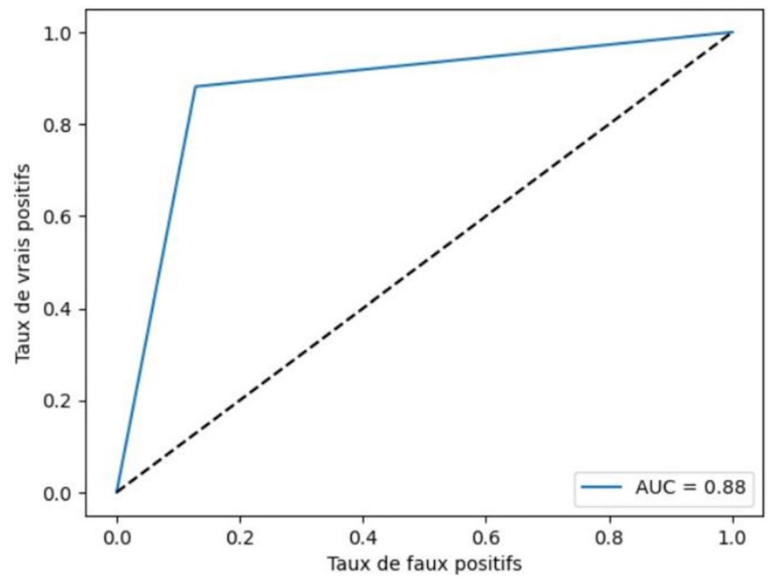

Figure 4. ROC curve of decision tree.

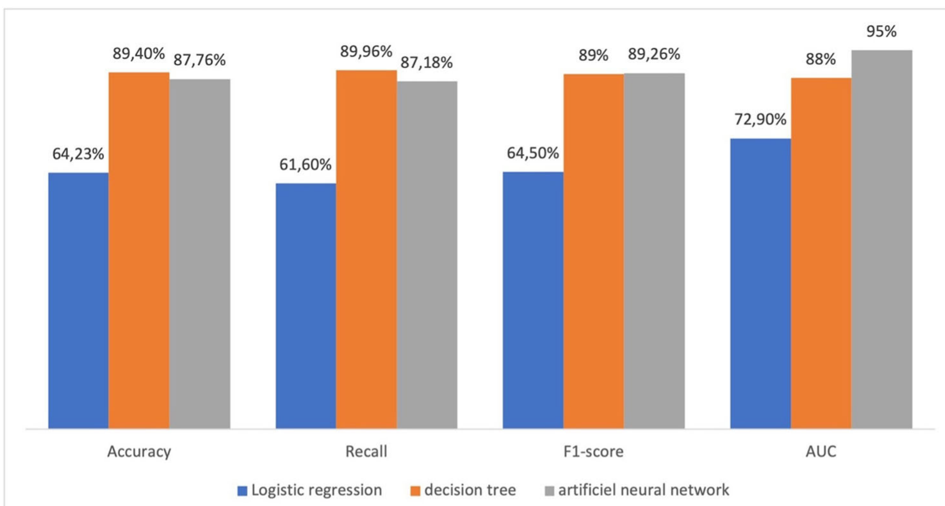

Figure 5. Performance of credit risk prediction models: logistic regression, and artificial neural network.

the neural network at 87.18%, whereas logistic regression displays the lowest recall at 61.62%. Regarding the F1-Score, the neural network distinguishes itself with the highest score at 89.26%, indicating a robust balance between precision and recall. The decision tree closely follows with a score of 89%, while logistic regression exhibits the lowest score at 64.5%. Finally, in terms of (AUC), the neural network outperforms the other models with an AUC of 95%, showcasing excellent discriminatory capability between classes. The decision tree follows with an AUC of 88%, whereas logistic regression displays the lowest value at 72.9%. In summary, the neural network emerges as the most proficient model in terms of overall accuracy, F1-Score and discriminatory capability (AUC). However, the decision tree also demonstrates robust performance in terms of precision and recall. Conversely, logistic regression shows generally weaker performance compared to the other models (see Figure 5).

This comparative analysis highlights the advantages of artificial intelligence-based techniques, particularly decision trees and neural networks, in credit risk prediction. However, logistic regression remains a viable classical method. The results obtained in this section contribute to enlightening our empirical approach. These findings underscore the significance of machine learning techniques, such as artificial neural networks, as a breakthrough in credit risk prediction and emphasize the importance of considering them in practical applications.

Table 6. Model assessment: fairness, explainability, and robustness comparison.

Model	Fairness	Explainability	Robustness
Logistic regression	Moderate	High	Moderate
Decision tree	Moderate	Interpretable but complex	Low
Neural network	Moderate	Low (black box)	High

Although the overall results demonstrate the superiority of machine learning techniques over classical methods, this comparison is not limited solely to predictive performance. Indeed, Table 6 summarizes other characteristics in terms of fairness, explainability, and robustness.

The logistic regression offers moderate fairness, high explainability due to the transparency of its coefficients, but moderate robustness when faced with complex data. On the other hand, the decision tree exhibits variable fairness depending on settings, medium explainability as it can become complex with depth, and low robustness due to its ability to handle noisy data. Neural networks, while performing well, demonstrate low to moderate fairness due to the risk of biases, low explainability as they are often considered black boxes, but high robustness as developed by Giudici and Raffinetti (2023) and Babaei et al. (2023).

4. Conclusion

The main objective of our study was to evaluate the performance of three different approaches for credit risk prediction, logistic regression as a classical traditional method, and artificial neural networks and decision tree based on artificial intelligence, considered significant advancements in this field. We commenced our investigation with a thorough exploration of the existing literature, following this, we embarked on a rigorous evaluation of the performance of the three models. This process began with an exhaustive presentation of our database, accompanied by both univariate and bivariate analyses, which laid the groundwork for constructing the three scoring models.

Through an extensive literature review, we found consistent evidence supporting the superiority of machine learning in credit risk prediction, as highlighted by Gupta and Goyal (2018), Dissananayake et al. (2007) and Islam et al. (2009). However, it is noteworthy that logistic regression has also shown significant performance, as discussed in the works of Khemais et al. (2016), Costa e Silva et al. (2020) and Ni (2010). The aim was to assess the performance of each model and determine the most effective approach for predicting credit risk. The results obtained have provided us with essential information to address our research problem. Our findings reveal that both the decision tree and the artificial neural network emerge as highly effective approaches, whereas logistic regression demonstrates more modest performance. Our discoveries confirm the two hypotheses formulated based on our literature review. However, the results we have obtained do not entirely validate Hypothesis H0C. Instead, the performance of ANN and Decision Tree demonstrates notable similarities.

Our study emphasizes the importance of adopting more sophisticated methods to capture complex relationships within the data. This finding paves the way for improving risk assessment tools in the financial sector, where predictive accuracy is essential for making informed decisions and managing risks effectively. By integrating these advanced techniques, financial institutions can achieve greater robustness in their analyses, thereby contributing to better risk management in an increasingly complex and dynamic environment. The empirical evidence obtained supports our perspective and underscores the importance of considering advances in artificial intelligence methods to enhance credit risk prediction. However, it is essential to acknowledge certain limitations in this study. Firstly, we only compared one conventional approach with two artificial intelligence approach, exploring additional methods would be beneficial for further comparisons and insights. Secondly, our focus was solely on consumer credit, including other types of credit would provide a more comprehensive and robust understanding of credit risk prediction.

Authors' contributions

All authors made significant contributions to this study. Mousaab Ghoujdam was responsible for validation, methodology, and conceptualization. Rachid Chaabita oversaw the work, focusing on formal analysis and conceptualization. Oussama Elkhalfi provided supervision and was in charge of software development. Kamal Zehraoui contributed to software development and data visualization. Hicham Elalaoui also handled visualization and supervision. Finally, Salwa Idamia played a key role in data visualization and was instrumental in writing, and editing the manuscript. All authors have accepted the publication of this version of the article and have provided final approval for the version to be published.

Disclosure statement

The authors affirm that they have no conflicts of interest to disclose.

Funding

This research received no external funding.

About the authors

Mousaab El Khair Ghoujdam Doctor in economic sciences, University Hassan II Casablanca, Maroc.

Rachid Chaabita Professor of Higher Education at the Faculty of Legal, Economic and Social Sciences and Director of the Laboratory Business Intelligence Governance of Organizations, Finance and Financial Crime. University Hassan II Casablanca.

Oussama Elkhalfi Research professor, university Sidi Mohamed Ben Abdallah de Fes, FSJES, Maroc.

Kamal Zehraoui Research professor, university Hassan II Casablanca.

Hicham Elalaoui Doctor in economic sciences, University Hassan II Casablanca.

Salwa Idamia PhD. Student, ESTC, university Hassan II Casablanca.

ORCID

Mousaab El Khair Ghoujdam http://orcid.org/0009-0009-3192-2781

Data availability statement

The data that support the findings of this study are available on request from the corresponding author.

References

Abiodun, O. I., Kiru, M. U., Jantan, A., Omolara, A. E., Dada, K. V., Umar, A. M., Linus, O. U., Arshad, H., Kazaure, A. L. L. A., & Gana, U. (2019). Comprehensive review of artificial neural network applications to pattern recognition. *IEEE Access*, *7*, 158820–158846. https://doi.org/10.1109/ACCESS.2019.2945545

Ali, U., Muhammad, W., Brahme, A., Skiba, O., & Inal, K. (2019). Application of artificial neural networks in micromechanics for polycrystalline metals. *International Journal of Plasticity*, *120*, 205–219. https://doi.org/10.1016/j.ijplas.2019.05.001

Angelini, E., Di Tollo, G., & Roli, A. (2008). A neural network approach for credit risk evaluation. *The Quarterly Review of Economics and Finance*, *48*(4), 733–755. https://doi.org/10.1016/j.qref.2007.04.001

Babaei, G., Giudici, P., & Raffinetti, E. (2023). Explainable fintech lending. *Journal of Economics and Business*, *125–126*, 106126. https://doi.org/10.1016/j.jeconbus.2023.106126

Bazarbash, M. (2019). Fintech in financial inclusion: machine learning applications in assessing credit risk. IMF Working Papers, 2019(109), 1. https://doi.org/10.5089/9781498314428.001

Bracke, P., Datta, A., Jung, C., & Sen, S. (2019). *Machine learning explainability in finance: an application to default risk analysis. Bank of England*. Working Paper No. 816.

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Carling, K., Jacobson, T., & Roszbach, K. (1998). "*Duration of consumer loans and bank lending policy: dormancy versus default risk*". No. 70. Sveriges Riksbank Working Paper Series.

Chen, F. L., & Li, F. C. (2010). Combination of feature selection approaches with SVM in credit scoring. *Expert Systems with Applications*, *37*(7), 4902–4909. https://doi.org/10.1016/j.eswa.2009.12.025

Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*, *110*, 12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004

Costa e Silva, E., Lopes, I. C., Correia, A., & Faria, S. (2020). A logistic regression model for consumer default risk. *Journal of Applied Statistics*, *47*(13-15), 2879–2894. https://doi.org/10.1080/02664763.2020.1759030

Dissananayake, B., Hendahewa, C. H., & Karunananda, A. S. (2007). Artificial Neural Network approach to credit risk assessment. In *2007 International Conference on Industrial and Information Systems*, 301–306.

El-Jerjawi, N. S., & Abu-Naser, S. S. (2018). Diabetes prediction using artificial neural network. *International Journal of Advanced Science and Technology*, *121*, 54–64.

Elsheikh, A. H., Sharshir, S. W., Abd Elaziz, M., Kabeel, A. E., Guilan, W., & Haiou, Z. (2019). Modeling of solar energy systems using artificial neural network: A comprehensive review. *Solar Energy*, *180*, 622–639. https://doi.org/10.1016/j.solener.2019.01.037

Gang, Y., Hu, X-j., & Zhou, T. (2022). Enterprise credit risk prediction using supply chain information: A decision tree ensemble model based on the differential sampling rate. Synthetic Minority Oversampling Technique and AdaBoost. *Expert Systems*, *39*(6), 10–20. https://doi.org/10.1111/exsy.12953

Gerken, W., Mishkin, F. S., & Eakins, S. G. (2015). *Study guide for financial markets and institutions* (8th ed.). F. S. Mishkin, & S. G. Eakins (Eds.). Pearson Education.

Giudici, P., & Raffinetti, E. (2023). SAFE Artificial Intelligence in finance. *Finance Research Letters*, *56*, 104088. https://doi.org/10.1016/j.frl.2023.104088

Giudici, P., Hadji-Misheva, B., & Spelta, A. (2020). Network based credit risk models. *Quality Engineering*, *32*(2), 199–211. https://doi.org/10.1080/08982112.2019.1655159

Gupta, D. K., & Goyal, S. (2018). Credit risk prediction using artificial neural network algorithm. *International Journal of Modern Education and Computer Science*, *10*(5), 9–16. https://doi.org/10.5815/ijmecs.2018.05.02

Hodo, E., Bellekens, X., Hamilton, A., Dubouilh, P. L., Iorkyase, E., Tachtatzis, C., & Atkinson, R. (2016). Threat analysis of IoT networks using artificial neural network intrusion detection system [Paper presentation]. In: *2016 International Symposium on Networks, Computers and Communications (ISNCC)* (pp. 1–6). IEEE. https://doi.org/10.1109/ISNCC.2016.7746067

Hosmer, Jr, D. W., Lemeshow, S., & Rodney, X. (2013). *Sturdivant. Applied logistic regression*. Vol. 398. John Wiley & Sons.

Hu, Y, & Su, J. (2022). Research on credit risk evaluation of commercial banks based on artificial neural network model. *Procedia Computer Science*, *199*, 1168–1176. https://doi.org/10.1016/j.procs.2022.01.148

Islam, M. S., Zhou, L., & Li, F. (2009). Application of Artificial Intelligence (Artificial Neural Network) to Assess Credit Risk: A Predictive Model For Credit Card Scoring.

Jiaqi, F. (2023). *Predicting of credit default by SVM and decision tree model based on credit card data*. BCP Business & Management. https://doi.org/10.54691/bcpbm.v38i.3666

Kerr, W. R., & Nanda, R. (2009). Democratizing entry: Banking deregulations, financing constraints, and entrepreneurship. *Journal of Financial Economics*, *94*(1), 124–149. https://doi.org/10.1016/j.jfineco.2008.12.003

Khemais, Z., Nesrine, D., & Mohamed, M. (2016). Credit scoring and default risk prediction: A comparative study between discriminant analysis & logistic regression. *International Journal of Economics and Finance*, *8*(4), 39. https://doi.org/10.5539/ijef.v8n4p39

Kimiagari, S., & Baei, F. (2022). Promoting e-banking actual usage: Mix of technology acceptance model and technology-organisation-environment framework. *Enterprise Information Systems*, *16*(8–9), 1894356. https://doi.org/10.1080/17517575.2021.1894356

Krenker, A., Bešter, J., & Kos, A. (2011). Introduction to the artificial neural networks. In *Artificial neural networks: Methodological advances and biomedical applications* (pp. 1–18). InTech.

Lee, T. S., Chiu, C. C., Chou, Y. C., & Lu, C. J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis*, *50*(4), 1113–1130. https://doi.org/10.1016/j.csda.2004.11.006

Mahbobi, M., Kimiagari, S., & Vasudevan, M. (2023). Credit risk classification: An integrated predictive accuracy algorithm using artificial and deep neural networks. *Annals of Operations Research*, *330*(1–2), 609–637. https://doi.org/10.1007/s10479-021-04114-z

Mushunje, L. (2021). *Predicting and managing credit risks using RIDGE and logistic LASSO regression*. Social Science Research Network.

Nakano, M., Sugiyama, D., Hori, T., Kuwatani, T., & Tsuboi, S. (2019). Discrimination of seismic signals from earthquakes and tectonic tremor by applying a convolutional neural network to running spectral images. *Seismological Research Letters*, *90*(2A), 530–538. https://doi.org/10.1785/0220180279

Ni, H. (2010). Consumer credit risk evaluation by logistic regression with self-organizing map. In *2010 Sixth International Conference on Natural Computation* (vol. 1, pp. 205–209).

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106. https://doi.org/10.1007/BF00116251

Schober, P. & Vetter, T. R. (2021). Logistic regression in medical research. *Anesthesia and Analgesia*, *132*(2), 365–366. https://doi.org/10.1213/ANE.0000000000005247

Sun, G., & Wang, S. (2019). A review of the artificial neural network surrogate modeling in aerodynamic design. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, *233*(16), 5863–5872. https://doi.org/10.1177/0954410019864485

Thomas, L. C., Ho, J., & Scherer, W. T. (2001). Time will tell: Behavioural scoring and the dynamics of consumer credit assessment. *IMA Journal of Management Mathematics*, *12*(1), 89–103. https://doi.org/10.1093/imaman/12.1.89

Wang, S., You, S., & Zhou, S. (2023). Loan prediction using machine learning methods. *Advances in Economics, Management and Political Sciences*, *5*(1), 210–215. https://doi.org/10.54254/2754-1169/5/20220081

Zhang, D., & Lin, X. (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics (Oxford, England)*, *4*(1), 57–74. https://doi.org/10.1093/biostatistics/4.1.57