



Optimization of Scoring for SME Credit Risk Management via Machine Learning

تحسين نماذج التقييم لإدارة مخاطر الائتمان للشركات الصغيرة والمتوسطة عبر تقنيات التعلم الآلي

Tari Mohamed Larbi¹, Bouguerra Amel^{2*}, Selatni Iatidel³

¹ High School of Commerce (Algeria), ✉ Tari.med.larbi@gmail.com.

² High School of Commerce (Algeria), ✉ etd_amelbouguerra@esc-alger.dz.

³ High School of Commerce (Algeria), ✉ etd_iselatni@esc-alger.dz.

Received: 10/12/2024

Accepted: 30/12/2024

Published: 31/12/2024

ABSTRACT

In the contemporary banking sector, managing credit default risk remains a pivotal challenge for financial institutions, especially in light of rapidly advancing technologies and increasingly complex economic conditions. Traditional evaluation methods, though insightful, often fall short in addressing the multifaceted nature of modern risk factors. This study investigates the application of Machine Learning techniques to enhance predictive models for credit risk, with a particular focus on small and medium-sized enterprises (SMEs). Drawing on a dataset of 647 credit applications and 17 explanatory variables, the research adopts an empirical approach. Four supervised learning algorithms—LR, DT, ANN, and RF—were implemented in R Studio software and assessed using confusion matrices and ROC curves. The findings highlight the Random Forest algorithm's superior predictive accuracy, underscoring the transformative role of technological innovations in optimizing risk management practices.

Keywords: Credit Risk, Scoring, Machine Learning, SMEs.

JEL Code : C45, C53, G21, G32.

* Corresponding Author: Amel BOUGUERRA, email: etd_amelbouguerra@esc-alger.dz

في القطاع المصرفي الحديث، تظل إدارة مخاطر عدم السداد تحديًا جوهريًا للمؤسسات المالية، لا سيما في ظل التقدم التكنولوجي السريع والظروف الاقتصادية المتزايدة التعقيد. على الرغم من أن الأساليب التقليدية للتقييم قد قدمت رؤى قيمة، إلا أنها غالبًا ما تعجز عن استيعاب الطبيعة المتعددة الأبعاد لعوامل المخاطر الحديثة. تبحث هذه الدراسة في تطبيق تقنيات التعلم الآلي لتحسين النماذج التنبؤية لمخاطر الائتمان، مع التركيز على الشركات الصغيرة والمتوسطة. استنادًا إلى مجموعة بيانات تتضمن 647 طلب ائتمان و17 متغيرًا تفسيريًا، تعتمد الدراسة على منهج تجريبي. تم تنفيذ أربعة خوارزميات تعلم خاضعة للإشراف—الانحدار اللوجستي، شجرة القرار، الشبكات العصبية الاصطناعية، وغابة القرارات العشوائية—باستخدام برنامج R Studio وتقييمها باستخدام مصفوفات الالتباس ومنحنيات ROC. أظهرت النتائج تفوق خوارزمية غابة القرارات العشوائية من حيث الدقة التنبؤية، مما يبرز الدور التحويلي للتطورات التكنولوجية في تحسين ممارسات إدارة المخاطر.

الكلمات المفتاحية: مخاطر الائتمان، التقييم الائتماني، التعلم الآلي، المؤسسات الصغيرة والمتوسط.

تصنيفات JEL: G32, G21, C53, C45.

1. Introduction

In an economy driven by debt, the banking sector plays a pivotal role in fostering economic growth, acting as a financial intermediary and supporting the broader economy's financing. However, this intermediary function presents significant challenges for lending institutions, as it represents both a source of profitability and a conduit for risk. Among these risks, credit risk—also known as counterparty risk—remains the most prevalent and detrimental for banks, manifesting when a borrower fails to meet their financial obligations within the agreed-upon timeframe.

In Algeria, small and medium-sized enterprises (SMEs) play a strategic role at the heart of the national economy, driving both economic and social development. However, their financing predominantly relies on bank loans, mostly granted by public institutions. While this dependency is essential for sustaining their activities, it may jeopardize the very existence of banks due to the high correlation between SMEs and default probability.

This issue is further exacerbated by several challenges inherent in the structure of SMEs. Given their size and organizational setup, SMEs face significant

information asymmetry. Banks are often unable to obtain accurate and reliable financial data about SMEs, complicating risk assessment. This asymmetry, coupled with insufficient collateral, intensifies the risk, making SME financing even more perilous for financial institutions.

Currently, Algerian public banks rely on traditional financial analysis methods and require collateral to assess credit risk. However, these conventional approaches, often outdated, may lead to poor borrower selection, increasing the likelihood of unpaid debts and threatening the sustainability of banking institutions. Therefore, banks must ensure effective credit risk management by adopting advanced techniques.

The advancement of technology and decision-support tools, driven by progress in computing and data utilization, has led to the development of increasingly sophisticated evaluation methods, such as credit scoring.

Credit scoring is a statistical method employed by financial institutions to evaluate the credit risk of borrowers by examining different financial metrics and historical credit behavior. Initially introduced by Beaver (1966) and Altman (1968), credit scoring has rapidly become an essential tool for borrower classification and risk assessment in the financial sector.

The incorporation of artificial intelligence (AI) in this domain represents a major advancement. AI scoring offers an innovative solution by enabling the analysis of larger datasets to create individualized credit scores. In this context, Machine Learning, a branch of AI, has emerged as a powerful tool, providing more accurate and adaptive predictive models than traditional methods. By analyzing complex and large datasets, algorithms such as random forests, neural networks, and decision trees enable a more reliable and relevant assessment of default risk.

In this era of technological revolution, the application of machine learning classification algorithms in credit scoring raises the following question:

How do machine learning algorithms optimize credit risk modeling for SMEs ?

This research is structured around the central issue, which forms the foundation of our inquiry. To address the study effectively, this article is organized into three main sections. The first section provides a comprehensive literature review and a theoretical framework of the models employed. The second section focuses on the methodology employed, offering an overview

of the dataset, the selected variables. The third section is dedicated to credit scoring modeling through the application of Logistic Regression (LR), Decision Trees (DT), Artificial Neural Networks (ANN), and Random Forests (RF). It includes the presentation of the obtained results, a comparative analysis of each model's performance, and a discussion of the findings.

2. Literature Review: Machine Learning and Automated Scoring in Credit Risk Management

Conventional credit risk management methods are based on statistical modeling and financial analysis. However, with the advent of machine learning, banks and financial institutions are increasingly adopting these techniques to enhance credit scoring accuracy and reduce loan-related risks, as highlighted in McKinsey's 2020 report.

Machine learning has revolutionized credit scoring by enabling the analysis of large datasets and identifying complex patterns that traditional methods, based on fixed criteria, may overlook.

Jordan & Mitchell (2015) emphasize that ML offers increased flexibility and accuracy due to its ability to learn new relationships without constant human intervention. This evolution, driven by the digitalization of financial services, enables faster and more precise credit scoring, based on complex behavioral data.

Supervised learning, which relies on labeled datasets, is a key lever in automating credit scoring. In this context, algorithms such as decision trees (CART), support vector machines (SVM), and random forests (RF) have proven highly effective in classifying borrowers according to their likelihood of repayment. According to Bishop (2006), neural networks and SVMs are particularly powerful in identifying complex relationships between input variables and default probability, making them well-suited for credit datasets where relationships may be nonlinear.

Regarding performance evaluation:

A series of studies, including those by Baesens and al. (2003), have investigated the effectiveness of classification algorithms in evaluating credit risk.. They analyzed 17 algorithms across eight datasets from international bank credits. Support vector machines (SVM) and neural networks delivered

the best predictive performances, with area under the curve (AUC) values of 66% and 91%, respectively.

In a subsequent analysis, Baesens and al. (2005) compared modern machine learning models, such as random forests and SVM, with traditional methods like logistic regression. They showed that these modern models outperform traditional approaches in predictive accuracy due to their ability to better capture nonlinearities in the data.

Lessmann and al. (2015), in a meta-analysis covering 41 studies, conclude that ensemble techniques, such as boosting and random forests, are particularly effective in enhancing the stability and robustness of scoring models.

A study by Zhou and Jiang (2020) evaluates deep neural networks (deep learning) applied to unstructured credit data. The results show a 10% improvement in classification rates over traditional methods, due to their ability to extract complex features from textual and transactional data.

Huang and al. (2004) demonstrate that artificial neural networks (ANNs) outperform logistic regression in contexts where the data exhibits significant nonlinear relationships.

Hand and Henley (1997) investigate the performance of models in real-world credit management environments. They find that, while logistic regression remains a benchmark, modern ML methods offer substantial advantages, particularly in terms of discriminating between high-risk borrowers.

In the Asian banking context, He and Xu (2019) compared ML methods (XGBoost, SVM) with parametric models, revealing a 15% increase in AUC for ML approaches.

In North Africa, Bensaid and al. (2022) show that applying XGBoost to credit scoring in local banks resulted in a significant reduction in default rates, with predictive accuracy increased by 18%.

Moscatelli and al. (2020), in an European study, conclude that ML approaches reduce biases introduced by subjective criteria, thereby making the credit granting process fairer.

3. Overview of the Classification Techniques Employed

3.1. Artificial Neural Networks (ANN)

An artificial neural network is a sophisticated computational architecture, inspired by the mechanisms of biological neurons, designed to model complex relationships in data by learning to identify patterns and nonlinear dependencies from examples. Structured in interconnected layers (input, hidden, and output), it relies on computational units, known as neurons, which interact through weighted connections that are adjusted during the learning process.

3.2. Decision Tree (DT)

Decision trees, as outlined by Hastie, Tibshirani, and Friedman in *The Elements of Statistical Learning* (2009), are predictive models that leverage a tree-like structure to partition data into homogeneous subgroups. The process involves recursively splitting the dataset based on impurity criteria, such as the Gini index or entropy, ultimately leading to clear decisions at the leaf nodes.

The CART algorithm constructs binary decision trees and employs the Gini index, also known as squared entropy, to identify relevant explanatory variables of various types.

Note: The Gini index serves as a metric for evaluating data impurity or heterogeneity. It assesses the probability of a randomly selected element being incorrectly classified based on the class distribution within a specific data subset. The index ranges from 0 to 1, where 0 indicates perfect purity (all observations belong to the same class) and 1 signifies a completely random distribution of classes. A low Gini index suggests that the data is relatively homogeneous, which is desirable for clear decision-making.

3.3. Logistic Regression (RL)

The objective of this approach is to identify a statistical relationship between the most discriminative explanatory variables and the firm's default status. It is frequently applied when the dependent variable is categorical, often taking binary or dichotomous values. The explanatory variables, in turn, can be either qualitative or quantitative.

3.4. Random Forest (RF)

Random Forest is a collective machine learning approach designed to address tasks such as classification, regression and other tasks. It operates by

generating a multitude of independent decision trees through the training phase. The final prediction is made by either averaging the individual tree predictions or selecting the most frequent class label. In this process, the Gini index serves as the criterion for determining the best splits in the decision trees.

4. Methodology

Our empirical study is grounded in both descriptive and comparative approaches. It relies on a dataset obtained from the Credit Department for Industries and Services of the "Crédit Populaire d'Algérie". Initially, the dataset comprised 667 companies and 30 explanatory variables. After a thorough data cleaning process, which included the removal of missing values, treatment of outliers, and statistical tests, the dataset was reduced to 647 companies with 17 explanatory variables, along with the target variable "Default."

Once the data were prepared, a crucial step involved comparing the performance of various Machine Learning (ML) algorithms. To this end, four credit scoring modeling methods were applied: logistic regression, decision trees, artificial neural networks, and random forests. Each algorithm underwent a validation phase using two performance evaluation tools: the confusion matrix and the ROC curve. These methods were employed to assess the ability of each model to accurately predict credit defaults, ultimately identifying the best-performing algorithm.

5. Analytical Description of the Sample

5.1. Exploration of the Study Sample

This study focuses on an initial sample of 667 small and medium-sized enterprises (SMEs) operating in the commercial and industrial sectors, registered with the "Crédit Populaire d'Algérie", which had subscribed to investment and operating loans over the period 2015 to 2022.

As a determining criterion, we defined any payment delay exceeding 90 days from the incident date as a default, in accordance with the Basel Committee guidelines and Regulation 2014-03 concerning the classification and provisioning of receivables.

5.2. Data Processing

In this context, the first step involved addressing the missing values in the dataset. After identifying 20 missing values, we removed them to maintain the

integrity and consistency of the data, reducing the sample to 647 companies. Subsequently, after identifying the outliers using the z-score method, we opted for the Winsorization technique to handle them. This method involves replacing extreme values with appropriate boundaries from the data distribution. To address this, thresholds were set at the 95th and 5th percentiles, with upper outliers replaced by the 95th percentile value and lower outliers substituted by the 5th percentile value. The application of this approach significantly reduced the proportion of outliers, from 2.16% to 0.75% across the 647 observations (see Appendix 1).

Note: An outlier percentage of 0.75% is generally considered acceptable for Machine Learning models, as most algorithms can handle this proportion effectively with proper data preprocessing.

5.3. Presentation of the Study Variables

The preselected variables are divided into two categories. The first category includes quantitative variables, represented by financial ratios, deemed relevant for explaining the financial situation of SMEs. These ratios, derived from the balance sheets, income statements, and cash flow statements of each company, are grouped into four categories: cash flow ratios, profitability ratios, liquidity ratios, and financial structure ratios. A total of 26 ratios were coded from R1 to R26.

Table 1: Exposure of initial ratios

	Ratios	Ratio measurement
<i>Cash flow ratios</i>	R1	Added Value / Revenue
	R2	Working capital in days of revenue ($WC \times 360$) / Revenue
	R3	Working capital requirement in days of revenue ($WCR \times 360$) / Revenue
	R4	Customer Payment Period ($\text{Accounts Receivable} \times 360$) / Revenue
	R5	Supplier Payment Period ($\text{Accounts Payable} \times 360$) / Revenue
<i>Liquidity ratios</i>	R6	Current Assets - Inventory / Short-Term Debt
	R7	Net availability / Total Assets
	R8	Availability (net) / Short-term debts
	R9	Financial expenses / EBITDA
	R10	Short-term debts / Total debt
	R11	Inventory turnover
<i>Profitability ratios</i>	R12	Net income / Total assets
	R13	Net income / Equity
	R14	SFC / Revenue
	R15	Operating added value / Revenue
	R16	EBITDA / Revenue
	R17	Operating profit / Revenue
	R18	Financial Result / Revenue
	R19	Net income before tax / Revenue)
	R20	Net income / Revenue
	R21	Total Debt / Total assets
<i>Financial structure ratios</i>	R22	Financial debt / Equity
	R23	Equity / non-current liabilities
	R24	Equity / Total assets
	R25	Financial debt / Net income
	R26	Long-term debt / SFC

The second category consists of qualitative variables, which pertain to the structure and specific situation of each company. These include the legal form

of the companies, with limited liability companies (SARL) predominating at 70% of our sample, followed by single-member limited liability companies (EURL) at 29%, and simplified joint-stock companies (SPA), general partnerships (SNC) at 8% and 3%, respectively. We also considered variables related to movements entrusted and the risk center statuts.

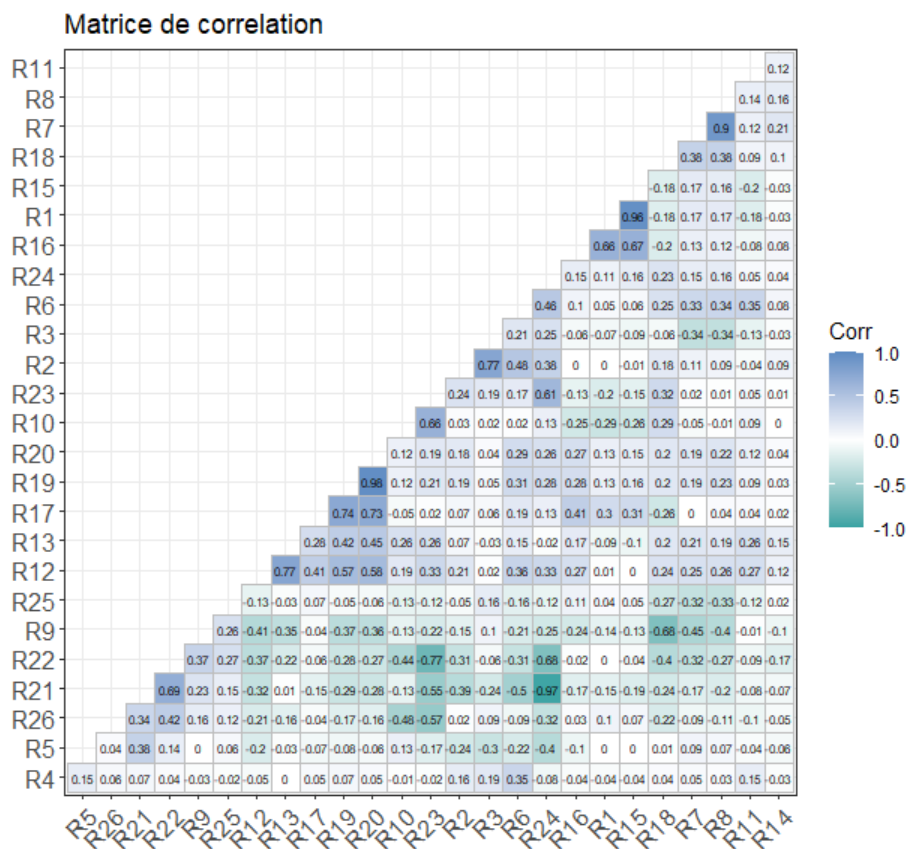
5.4. Variable Selection

The development of the scoring function is based on the assumption that the variables are dependent and exhibit minimal correlation with each other. The objective is to eliminate variables that do not explain default risk while retaining those that are significant.

Initially, a Pearson correlation test between the quantitative variables is essential to detect any potential high correlations that could bias the estimation of the coefficients in the scoring function due to redundant information. The correlation is deemed significant when the coefficient exceeds 0.5 in absolute value.

Subsequently, to identify the variables that significantly explain default, we applied the student's t-test, which assesses the presence of a dependency relationship between a quantitative variable and default, under the following hypothesis:

Figure 1: Correlation Matrix



$$\begin{cases} H0: \text{The means are equal } \alpha \geq 5\% \\ H1: \text{les moyennes ne sont pas égales } \alpha < 5\% \end{cases}$$

Source: R studio Software

Table 2: Quantitative variables retained according to Student's test

	Variables	t	Significance
<i>Cash flow ratios</i>	R1: Added Value / Revenue	3.8766	0.0001113
	R4: Customer payment period / Revenue	12.573	< 2.2e-16
	R5: Supplier payment period	6.3769	3.443e-10
<i>Liquidity ratios</i>	R6: Current Assets - Inventory / Short-Term Debt	7.6792	5.784e-14
	R7: Net availability / Total Assets	-12.797	< 2.2e-16
	R9: Financial expenses / EBITDA	2.7427	0.006256
	R10: Short-term debts /Total debt	25.753	< 2.2e-16
<i>Profitability ratio</i>	R13: Net income / Equity	-2.1254	0.03383
	R14: SFC / Revenue	-6.9521	6.298e-12
	R18: Financial result / Revenue	-1.8996	0.05793
	R19: Net income after taxes / Revenue	-2.1869	0.02911
	R22: Financial debt / Equity	2.0828	0.03766
<i>Financial structure ratios</i>	R23: Equity / non-current liabilities	5.5003	5.286e-08
	R24: Equity / Total assets	7.0909	2.382e-12
	R25: Financial debt / Net income	7.2268	1.4e-12

Finally, we conducted the Chi-square test to assess the dependency between the qualitative variables and default, under the following hypotheses:

$$\begin{cases} H0: \text{The variables are dependent } \alpha < 5\% \\ H1: \text{The variables are independent } \alpha \geq 5\% \end{cases}$$

Table 3: Chi-square test results

Variables	Value	Significance
<i>Legal form</i>	9.0995	0.1052
<i>Entrusted movement</i>	25.351	4.778e-07
<i>Risk center</i>	10.79	0.00102
<i>Unpaid credit from peers</i>	0.18718	0.6653

In conclusion, the results of the statistical tests led to the selection of **17 explanatory variables**, including the "**Risk Center**" variable, "**Entrusted Movements**," and 15 quantitative variables.

6. Data Modeling - Empirical Results

6.1. Data Preparation

This phase begins with the creation of a dataframe recipe, within which the target variable (Default) is identified and the numerical predictors are normalized to ensure consistency in scales. It also includes a data rebalancing step, aimed at correcting class imbalance and ensuring more accurate modeling results.

6.2. Sample Distribution

As part of the model development, the base sample is split into two subsamples using an 80/20 ratio:

- **Training Sample (80%):** Used to design and fit the scoring model based on the training data.
- **Validation Sample (20%):** Used to validate the effectiveness of the model developed from the training sample by testing its performance and predictive capability on future credit data.

The table below provides an overview of the distribution results:

Table 4: Sample Composition

	Initial sample	Training sample	Validation sample
Healthy companies	498	404	97
Defaulting companies	149	113	33
Σ	647	517	130

Source: developed by the authors using **R Studio software**

6.3. Credit Scoring Modeling

The next steps involve modeling, measuring, and discussing four Machine Learning models: Logistic Regression, Decision Tree (CART), Artificial Neural Networks, and Random Forest.

6.3.1. Logistic Regression (LR)

The use of binary logistic regression (binomial) allowed for the identification of the key explanatory variables in estimating the likelihood of default. The analysis highlights that the majority of the selected variables are not statistically significant (see Appendix No. 2), with the exception of three variables: Entrusted Movements (EM), which shows the highest statistical significance at a 0.1% level, followed by the Risk Center Status (RC) at 1%, and finally, the Short-term debt/Total Debt ratio (R10) at 5%.

The scoring function obtained in this model links the score to the probability of company default. A higher score reflects an increased probability of default. The classification of groups is carried out as follows:

$$\begin{cases} Z < 0: \text{The company is considered to be healthy} \\ Z > 0: \text{The company is considered to be defaulting} \end{cases}$$

The application of binary logistic regression resulted in the following function:

$$Z = -0.416 + 0.744 RC - 0.981 EM - 1.119 R10$$

▪ Interpretation of the Model

The analysis of the scoring function reveals an overall consistency between the signs of the coefficients of the variables and the theoretical expectations, with the exception of the variable R10. To validate this consistency, each coefficient

is individually examined to ensure alignment with the underlying financial logic.

The variable "**Risk Center Status**" (RC) plays a predominant role in our analysis, with a positive coefficient consistent with theoretical expectations. This variable reflects the borrower's credit history with other financial institutions. Increased reliance on credit, particularly from multiple banks, exposes the borrower to a higher risk of default due to the accumulation of excessive debt that may exceed their repayment capacity. Specifically, an increase of one unit in this variable corresponds to a **0.7444283** increment in the log-odds of default.

In contrast, **Entrusted Movements (EM)** stand out with a negative coefficient, in line with economic expectations. High entrusted transactions indicate resilient business activity and active account management, reflecting better financial health. An increase of one unit in Entrusted Transactions thus reduces the log(odds) of default by **0.9810156**. These results confirm the findings of Touati-Tliba (2003), who had already identified this variable as a key indicator of default in the Algerian context.

Regarding the **Short-term debt /Total Debt ratio**, a high value is generally associated with lower repayment capacity, which should theoretically increase the probability of default. However, the obtained coefficient suggests an inverse relationship: an increase of one unit in this ratio reduces the log(odds) of default by **1.1196825**, indicating a divergence from the usual financial logic.

To evaluate the model fit, the Log-Likelihood (LR) statistic is calculated (see Appendix No. 3), assuming that the reduced model incorporates only the following variables: **Entrusted Movements, Risk Center Status, and the Short-term debt /Total Debt ratio (R10), in addition to the intercept.**

$$\left\{ \begin{array}{l} H0: \text{The reduced model is more significant than the full model} \\ H1: \text{Otherwise} \end{array} \right.$$

This statistic adheres to a chi-square distribution with 14 degrees of freedom. A p-value of 0.008, significantly low, rejects the null hypothesis that the reduced model fits the data as well as the full model. This suggests that the full model provides a better fit and should, therefore, be retained.

The **McFadden statistic**, or **Pseudo- R^2** , assesses the quality of the model's fit. With a value of 0.12985 (see Appendix No. 4), it indicates that the logistic regression model explains approximately 12.99% of the variance in the data. This indicates a moderate concordance between the model's predictions and the observed data. However, a significant portion of the variance remains unexplained, highlighting potential areas for improvement to enhance the model's overall fit.

6.3.2. Decision Tree (DT)

The CART algorithm (Classification and Regression Trees) employed in this study constructs binary decision trees, leveraging the Gini index—also referred to as the quadratic entropy—to identify explanatory variables with the highest discriminative power, regardless of whether they are continuous or categorical.

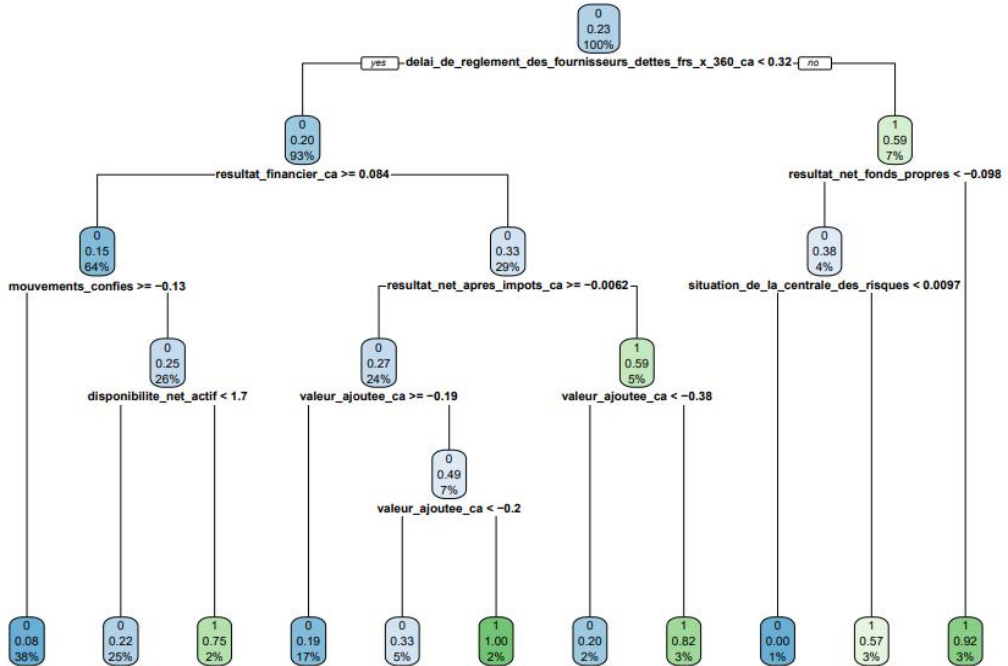
To improve the model's reliability, extensive hyperparameter optimization was conducted. Utilizing the *tune* function, various parameter configurations were explored (Appendix No. 5), with the area under the ROC curve serving as the key metric to assess model performance. The selected hyperparameters are as follows:

- **Maximum tree depth:** Set to 6, enabling the model to capture complex interactions within the data while minimizing the risk of overfitting.
- **Minimum number of observations per terminal node (min_n):** Defined as 2, representing the minimum sample size required to split a node. Higher values result in less complex trees.
- **Cost complexity:** Regularized at 0.01, this parameter penalizes structural complexity in the tree to balance model fit and generalizability.

This systematic methodology ensures optimal calibration of the model, thereby enhancing its capacity to accurately discriminate between classes.

The decision tree's graphical representation, provided below, delineates the segmentation rules and the resulting clusters:

Figure 2: Decision Tree Model



Source: R studio Software

Interpretation: This decision tree highlights the significance of financial ratios in classifying companies based on their financial stability. The ratio of supplier payment delays to revenue (Sales) is identified as the primary discriminating criterion (reflecting financial practices and the ability of firms to meet their obligations), with a critical threshold set at 0.32. If this ratio falls below the threshold, 483 cases are observed, 93% of which belong to Class 0. Subsequently, other criteria, such as the financial result-to-sales ratio (threshold: 0.084), come into play, leading to further subdivisions based on indicators such as entrusted movements, net profit after tax-to-sales, net cash relative to total assets, and value added-to-sales.

The terminal nodes reveal significant variations in classification probabilities. For instance, a 38% probability for Class 0 is observed when entrusted movements exceed -0.1345. Conversely, a 17% probability for Class 0 is noted when value added-to-sales surpasses -0.19. If the initial ratio is greater than or equal to 0.32, 34 cases cluster with a higher probability of belonging to Class 1 (58.82%).

These successive splits, influenced by indicators such as net profit-to-equity and Risk Center Status, illustrate complex financial relationships. This granular structure provides precise insights for analyzing and understanding the factors driving classifications within the model.

6.3.3. Artificial Neural Networks (ANN)

To conduct this analysis, the use of Artificial Neural Networks (ANN) through the "**Multilayer Perceptron**" procedure was favored.

The architecture chosen for the neural network consists of an **input layer** made up of **17 neurons** representing the selected explanatory variables. **Two hidden layers**, containing **8 and 4 neurons** respectively, perform complex data transformations. Finally, a **single decision neuron** in the **output layer** indicates the probability of a company being healthy or in default.

To optimize the network's performance, a series of tests were conducted to determine the most suitable hyperparameter configurations. The goal was to minimize the Mean Squared Error (MSE) while maximizing the correct classification rate. The experimental outcomes are concisely summarized in the following table:

Table 5: ANN Architecture Selection Results

Number of Layers	Batch size	Epochs	Verbos	Validation	MSE of the test sample
Input layer = 17 Hidden layers = (8,4) Output layer =1	30	100	1	0.30	0.2230769
	35	100	1	0.20	0.2384615
	36	100	1	0.15	0.2076923
	36	100	1	0.16	0.2234715
	36	100	1	0.17	0.2 ¹

Source: Developed by the authors using **R Studio software**.

A comprehensive analysis of these configurations led to the establishment of an optimal architecture defined by the following hyperparameters: a batch size of 36, striking an effective balance between convergence speed and stability; a total of 100 epochs, ensuring proper model convergence; a verbose

¹ see Appendix No. 6

setting of 1, providing detailed insights into the training process; and a validation split of 0.17, allocating 17% of the training data for intermediate performance evaluation to mitigate the risk of overfitting.

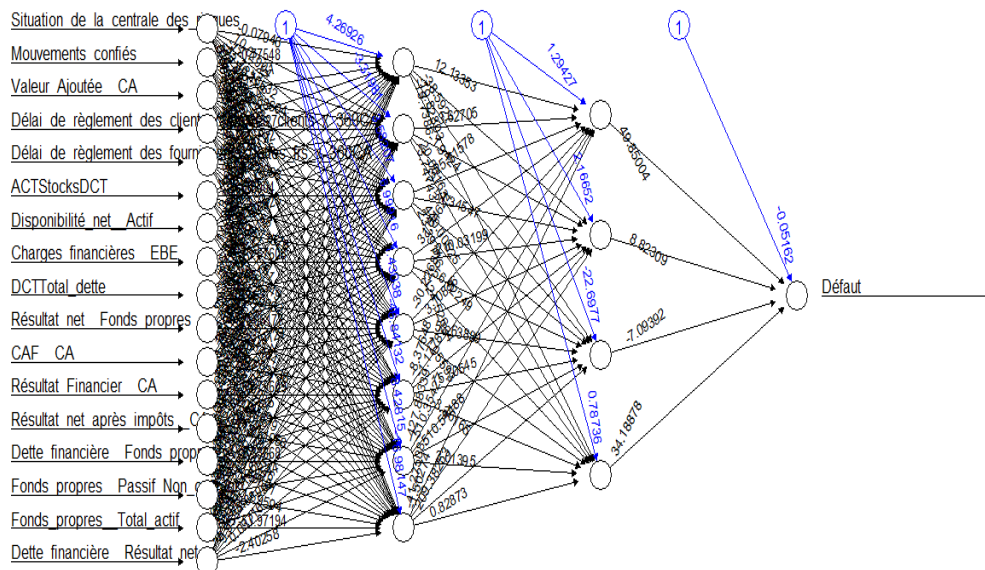
With this optimized configuration, the model achieved a **Mean Squared Error (MSE)** of **0.2** on the test set. Furthermore, the "**ReLU**" (Rectified Linear Unit) activation function was applied to both the input and hidden layers, while the "**Sigmoid**" activation function was utilized in the output layer, further strengthening the model's robustness in the classification task.

The model exhibits an **overall error rate of 21.86%** on the test sample, indicating that approximately one in five observations is misclassified. While this result highlights a significant margin for improvement, it also demonstrates the model's ability to capture a substantial portion of the underlying patterns in the data, while maintaining a degree of generalization to new, unseen data.

▪ The Model Architecture

The diagram below illustrates the architecture of the neural network model:

Figure 3: Architecture of the ANN Model



Source: R studio Software

The key variables in this model reveal their influence on the predictive power of the model and their strategic significance in credit risk assessment. Payment delays, both with suppliers (0.298) and clients (0.294), reflect cash-flow management and short-term solvency, which are essential for SME stability. The debt relative to earnings (0.197) illustrates the financial vulnerability of businesses in relation to their profitability capacity, while the status with the risk center (0.120) emphasizes the importance of credit history in debtor classification. Finally, although less dominant, operational profitability (SFC /Revenue, 0.056) provides critical insight into the economic viability of businesses. These variables, integrated within the model, represent important levers for refined credit risk prediction (see Appendix No. 7).

6.3.4. The Random Forest (RF)

As the final model in our predictive analysis, we employed Random Forest, a non-parametric technique, following prior explorations of neural networks and decision trees. This model computes variable importance based on node impurity, commonly known as Gini importance, which quantifies the contribution of each variable to reducing node impurity. To ensure optimal performance, both in terms of computational efficiency and memory management, the '**ranger**' algorithm was employed for this classification task, facilitating the handling of large datasets and minimizes the risk of overfitting (Wright, Ziegler, 2017).

The construction strategy involves creating a forest of **500 trees**, each selecting a random subset of 4 variables at each split (see Appendix No. 8). This approach reduces tree correlation, improves generalization, and mitigates overfitting.

In this context, the Random Forest model points out the variables with the greatest impact. Leading the way, the ratio of net income after tax to revenue indicates that higher profitability reduces the risk of default. This is followed by financial debt to net income, which measures the pressure of debt on earning capacity, and the supplier payment delay to revenue, which reveals potential cash-flow tensions. Finally, net liquidity to assets and net income to equity reflect liquidity and profitability of equity, respectively (see Appendix No. 9).

7. Performance Assessment and Comparative Analysis of Machine Learning Models for Default Prediction

The assessment of predictive model performance constitutes an essential phase in any analytical framework, facilitating the validation of a model's relevance and its ability to distinguish between different observation classes.

In this regard, key metrics such as accuracy and the area under the ROC curve (AUC) are regarded as robust and widely accepted standards in the evaluation of statistical and machine learning models (Hosmer, Lemeshow & Sturdivant, 2013). Accuracy quantifies the proportion of correctly classified observations, while AUC, which measures the likelihood of a model correctly ranking two randomly chosen instances, provides a holistic evaluation of its discriminatory power (Hanley & McNeil, 1982). These metrics are particularly valuable for comparing the predictive performance of alternative models, as demonstrated by several authors in the field (Kuhn & Johnson, 2013; Fawcett, 2006).

Logistic regression (LR), despite being a classical and extensively applied method in default analysis, demonstrates significant limitations in addressing complex scenarios. While it successfully identifies key explanatory variables, such as *Entrusted Movements* (EM) and *Risk Center Status* (RC), its ability to accurately differentiate high-risk firms remains limited. With an **accuracy of 72%** on the test sample and an **area under the ROC curve (AUC) of 0.59**, the model exhibits weak discriminative performance, barely exceeding random chance. This underperformance can be attributed to the parametric nature of LR, which is inherently ill-suited to capturing nonlinear interactions and intricate relationships between explanatory variables. These observations align with the findings of Kleinbaum and Klein (2010), who highlight the challenges faced by logistic regression when applied to complex and high-dimensional datasets.

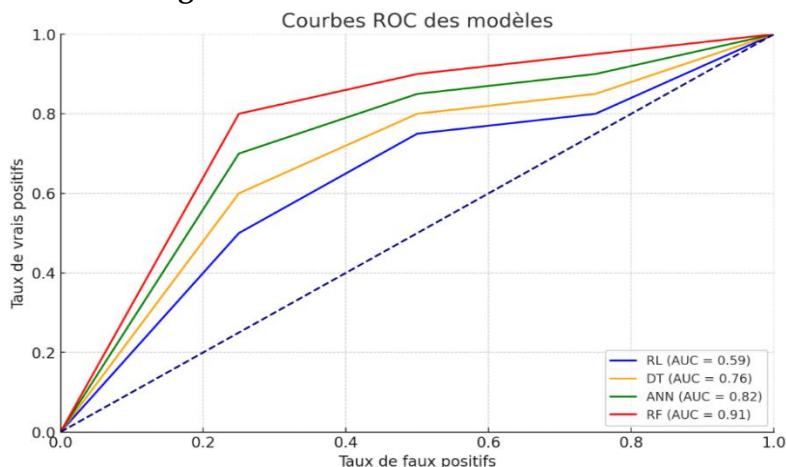
Conversely, **the decision tree (CART)**, as a non-parametric methodology, provides a more flexible and adaptive approach. Its capacity to capture nonlinear relationships and deliver an interpretable framework for decision-makers constitutes a notable strength. Achieving an **accuracy of 82.3%** and an **AUC of 0.76**, this model exhibited superior alignment with the data compared to logistic regression. Notably, the segmentation criteria based on specific financial ratios supported a more nuanced identification of high-risk groups. However, despite its enhanced precision, decision trees are sensitive to

threshold determination and variable interpretation, which can introduce biases when faced with significant data variability. This intuitive modeling approach, offering a clear visualization of the conditions underlying classification decisions, aligns with the findings of Breiman et al. (1984), who highlighted the effectiveness of decision trees in capturing complex data structures.

Artificial Neural Networks (ANN), also renowned for their ability to model nonlinear and multidimensional data structures, have demonstrated superior performance in this context. Achieving an **accuracy of 80%** and an **AUC of 0.82**, this model exemplify its capacity to capture intricate relationships between explanatory variables and to process complex datasets with remarkable precision. The meticulous optimization of hyperparameters—specifically, the configuration of hidden layers and the adoption of the ReLU activation function—has significantly enhanced model convergence, empowering a robust distinction between healthy firms and those at risk. These findings are in agreement with those of Rumelhart and al. (1986), which underscored the efficacy of ANNs in navigating complex datasets characterized by subtle and non-linear variable interactions.

Finally, the **Random Forest (RF)** model, attaining an **accuracy of 89%** and an **AUC of 0.91**, demonstrated the highest performance in this study. By aggregating a multitude of decision trees, RF mitigates the risk of overfitting while enhancing the model's discriminative power. The random selection of variables and the diversity of trees contribute to near-optimal performance, consistent with Breiman's (2001) findings, which underscore the flexibility and reliability of Random Forests in handling complex and noisy datasets. This method proves particularly well-suited for financial analyses, where models must adapt to data variability and the dynamic nature of economic conditions.

In summary, this study highlights a clear hierarchy in model performance, with advanced techniques such as neural networks and random forests significantly outperforming traditional methods like logistic regression. This observation fits into a larger theoretical framework that emphasizes the value of nonparametric methods for solving complex classification problems, as noted by Kuhn and Johnson (2013). These results support the notion that the choice of model should be based not only on its predictive abilities but also on how well it adapts to the specific characteristics of the data being analyzed.

Figure 4: ROC Curves of the 4 Models

Source : R studio Software

8. Conclusion

In a highly competitive banking sector, financial institutions must continuously innovate to sustain their market position. In Algeria, where the economy heavily relies on small and medium-sized enterprises (SMEs), these businesses predominantly depend on loans granted by public banks such as "Crédit Populaire d'Algérie (CPA)". This concentration of lending exposes banks to considerable risks, as their solvency is closely tied to the financial performance of their borrowers.

In this context, the adoption of modern solutions, such as machine learning algorithms, emerges as a strategic breakthrough for proactive credit risk management. This study has demonstrated that these tools significantly enhance scoring accuracy, providing banking advisors with a stronger decision-making foundation. By analyzing 647 credit applications and leveraging 17 carefully selected economic and financial variables, four supervised models were compared: logistic regression, decision tree, artificial neural networks, and random forest.

The results reveal the clear superiority of the random forest model, achieving a performance rate of 91%, followed by neural networks (82%), decision tree (76%), and logistic regression (59%). These findings confirm the potential of machine learning in optimizing credit scoring processes and accentuate the importance of transitioning to approaches that align with the dynamic realities of the modern banking sector.

Beyond technical performance, this study highlights the strategic importance of data quality. The effectiveness of the models is closely dependent on the richness and structure of the datasets used. The research also calls for exploring hybrid approaches that combine machine learning with econometric methods to push the boundaries of predictive analytics further.

In conclusion, this research represents a pivotal step toward the digital transformation of the banking sector in Algeria. The integration of artificial intelligence and machine learning technologies not only addresses current challenges but also positions financial institutions at the forefront of innovation.

9. Bibliography

- Altman, E. I. (1968). Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.
- Anderson, R. (2007). *Credit Scoring and Its Applications*. Springer, USA.
- Baesens, B., Van Gestel, T., Viaene, S., & Mues, C. (2003). Benchmarking classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6), 627-633.
- Baesens, B., Van Gestel, T., Viaene, S., & Mues, C. (2005). Modern classification techniques in credit scoring. *Journal of the Operational Research Society*, 56(8), 975-986.
- Batta, M. (2018). Machine learning algorithms - A review. *International Journal of Science and Research (IJSR)*, 7(11).
- Bensaid, A., Khoukhi, L., & Feki, M. (2022). XGBoost-based credit scoring: Evidence from North African banks. *Journal of Financial Technology*, 11(3), 214-227.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, USA.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. CRC Press, USA.

- GLIZ, A., & TOUATI-TLIBA, M. (2012). Estimation du risque de crédit et qualité de l'information comptable en Algérie. *Les cahiers du CREAD*, 98-99.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press, USA.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29-36.
- 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), 245-275.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, USA.
- He, H., & Xu, L. (2019). Comparison of machine learning methods in credit risk assessment: Evidence from the banking sector in China. *Journal of Financial Services Research*, 56(2), 109-130.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). Wiley, USA.
- Huang, J., Chen, H., & Huang, S. (2004). Credit scoring with artificial neural networks and logistic regression. *Applied Artificial Intelligence*, 18(6), 607-617.
- Jacob, H., & Sardi, A. (2001). *Management des risques bancaires*. AFGES, France.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- Kharoubi, C., & Thomas, P. (2016). *Analyse du risque de crédit : Banque et marché* (2e éd.). Paris, France.
- Kleinbaum, D. G., & Klein, M. (2010). *Logistic Regression: A Self-Learning Text* (3rd ed.). Springer, USA.
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer, USA.

- McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105-142). Academy Press.
- McKinsey & Company. (2020). Report on Credit Risk Management. Seminar on Credit Risk Management, London, UK.
- Moscatelli, M., Rizzi, S., & Fracassi, A. (2020). Reducing biases in credit scoring with machine learning: A European perspective. *European Journal of Financial Economics*, 27(4), 367-380.
- Omar, C. (2018). Étude de différentes méthodes d'analyse de risque crédit : Revue de littérature. *Revue du Contrôle de la Comptabilité et de l'Audit*, 2(4).
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.
- Thomas, L. C., Edelman, D. B., & Crook, J. N. (2002). *Credit Scoring and Credit Control*. Oxford University Press, UK.
- Touati-Tliba, M. (2003). Modèles d'aide à la décision d'octroi de crédits et performance du portefeuille crédit des banques publiques algériennes. *Revue des sciences commerciales et de gestion*, 1(1), 73-100.
- Zhou, Z., & Jiang, L. (2020). Deep learning for credit scoring: An empirical study on unstructured data. *Expert Systems with Applications*, 142, 113003.
- Zhou, Z. H. (2021). *Machine Learning* (1st ed.).

Appendices

Appendix 1: Percentage of outliers before and after transformation

```
> print(percentage_outliers)
[1] 2.163833

> print(percentage_outliers)
[1] 0.7575113
```

Appendix 2 : LR Model

```
> summary(model)
```

Call:

```
glm(formula = Défaut ~ ., family = binomial, data = train_data)
```

Coefficients:

	Estimate	Std. Error
(Intercept)	-0.4161267	0.5126504
'Situation de \\r\\nla centrale des risques'	0.7444283	0.2369570
'Mouvements\\r\\n confiés'	-0.9810156	0.2363007
'valeur Ajoutée / \\r\\nCA'	-1.0503989	0.6501576
'Délai de règlement des clients \\r\\n(créances clients x 360)/CA'	-0.0006494	0.0007181
'Délai de règlement\\r\\n des fournisseurs \\r\\n(dettes frs x 360)/CA'	0.0012226	0.0006265
'ACT-Stocks/DCT'	-0.1112318	0.0584044
'Disponibilité (net) / Actif'	-0.1963977	0.7208925
'Charges financières / EBE'	0.0595026	0.0681285
'DCT/Total dette'	-1.1196825	0.4855677
'Résultat net / Fonds propres'	0.0577618	0.1021567
'CAF / CA'	0.3741230	0.1985866
'Résultat Financier / CA'	-1.6160798	1.9993125
'Résultat net après impôts / CA'	0.4001452	0.5659430
'Dette financière /\\r\\n Fonds propres'	-0.0054930	0.0081085
'Fonds propres / \\r\\nPassif 0 courant'	-0.0690036	0.0996172
'Fonds propres / \\r\\nTotal actif'	0.6806752	0.5669265
'Dette financière / \\r\\nRésultat net'	0.0016240	0.0009896

	z value	Pr(> z)
(Intercept)	-0.812	0.41695
'Situation de \\r\\nla centrale des risques'	3.142	0.00168 **
'Mouvements\\r\\n confiés'	-4.152	3.3e-05 ***
'valeur Ajoutée / \\r\\nCA'	-1.616	0.10618
'Délai de règlement des clients \\r\\n(créances clients x 360)/CA'	-0.904	0.36580
'Délai de règlement\\r\\n des fournisseurs \\r\\n(dettes frs x 360)/CA'	1.951	0.05101 .
'ACT-Stocks/DCT'	-1.905	0.05684 .
'Disponibilité (net) / Actif'	-0.272	0.78529
'Charges financières / EBE'	0.873	0.38245
'DCT/Total dette'	-2.306	0.02111 *
'Résultat net / Fonds propres'	0.565	0.57179
'CAF / CA'	1.884	0.05957 .
'Résultat Financier / CA'	-0.808	0.41891
'Résultat net après impôts / CA'	0.707	0.47954
'Dette financière /\\r\\n Fonds propres'	-0.677	0.49813
'Fonds propres / \\r\\nPassif 0 courant'	-0.693	0.48851
'Fonds propres / \\r\\nTotal actif'	1.201	0.22989
'Dette financière / \\r\\nRésultat net'	1.641	0.10078

 signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 542.94 on 516 degrees of freedom
 Residual deviance: 472.44 on 499 degrees of freedom
 AIC: 508.44

Number of Fisher Scoring iterations: 6

Appendix 3: Log likelihood statistic results

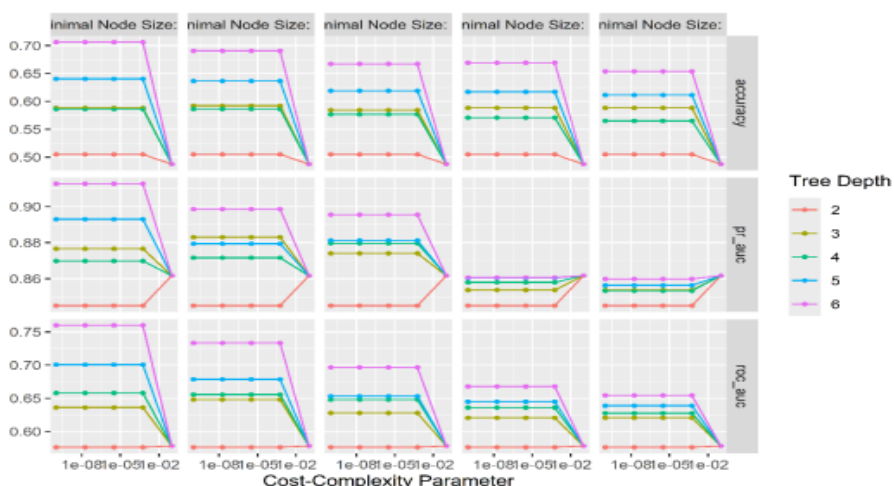
```
> logLik_complet <- logLik(model_complet)
> logLik_reduit <- logLik(model_reduit)
> print(logLik_complet)
'log Lik.' -236.2207 (df=18)
> print(logLik_reduit)
'log Lik.' -251.14 (df=4)
> LR_statistic <- -2 * (logLik_reduit - logLik_complet)
> print(LR_statistic)
'log Lik.' 29.8387 (df=4)
> df <- df.residual(model_reduit) - df.residual(model_complet)
> print(df)
[1] 14
> p_value <- pchisq(LR_statistic, df = df, lower.tail = FALSE)
> print(p_value)
'log Lik.' 0.008031798 (df=4)
```

Appendix4: Results of the McFadden statistic

```
> print(mcfadden_r2)
[1] 0.1298516
```

Appendix 5: Evaluation and Configuration of Decision Tree Models

tree_depth	min_n	.metric	.estimator	mean	n	std_err
<int>	<int>	<chr>	<chr>	<dbl>	<int>	<dbl>
6	2	roc_auc	binary	0.705	5	0.0270
6	40	roc_auc	binary	0.649	5	0.0206
2	2	roc_auc	binary	0.541	5	0.0190
2	40	roc_auc	binary	0.539	5	0.0200
2	2	roc_auc	binary	0.5	5	0
6	2	roc_auc	binary	0.5	5	0
2	40	roc_auc	binary	0.5	5	0
6	40	roc_auc	binary	0.5	5	0



Appendix 6 : MSE results (ANN)

```

##model
{r}
model %>% fit(x = X_train,y = Y_train,batch_size = 36,epochs = 100,verbose = 1,validation_split = 0.17)

> print(auc_score)

Call:
roc.default(response = Y_test, predictor = y_pred_proba)

Data: y_pred_proba in 130 controls (Y_test 0) < 130 cases (Y_test 1).
Area under the curve: 0.8232
>

{r}
model %>% evaluate(x = X_test,y = Y_test)

5/5 [-----] - 0s 3ms/step - loss: 0.7069 - accuracy: 0.8000
      loss accuracy
0.7069061 0.8000000

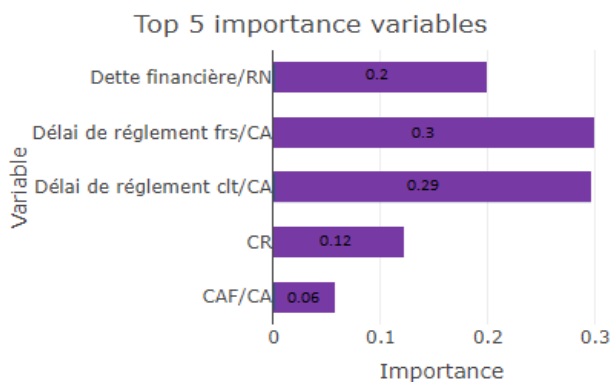
{r}
# Calculer le MSE pour l'ensemble d'entraînement
mse_train <- mean((eval_df_train$y - eval_df_train$y_hat)^2)
print("MSE for Train:")
print(mse_train)

# Calculer le MSE pour l'ensemble de test
mse_test <- mean((eval_df_test$y - eval_df_test$y_hat)^2)
print("MSE for Test:")
print(mse_test)

[1] "MSE for Train:"
[1] 0.2185687
[1] "MSE for Test:"
[1] 0.2

```

Appendix 7: Top 5 importance of the ANN model



Appendix 8: Specific features of the RF model

```

Type of random forest: classification
Number of trees: 500
no. of variables tried at each split: 4

```

Appendix 9: Top 5 importance of the RF model

