JOURNAL OF CULTURAL ANALYSIS AND SOCIAL CHANGE

Journal of Cultural Analysis and Social Change,
2025, 10(4), 1308-1326
ISSN: 2589-1316

A Comparative Analysis of Machine Learning and Fuzzy Logic Models for Credit Risk Prediction

Nadia Belkhir[1*], Sihem Khemakhem[2], Asma Bouzouitina[3]

[1]*Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia*
[2,3]*Faculty of Economics and Management of Sfax, University of Sfax, Tunisia*

***Corresponding Author:** nabelkhir@imamu.edu.sa

Citation: Belkhir, N., Khemakhem, S., & Bouzouitina, A. (2025). A Comparative Analysis of Machine Learning and Fuzzy Logic Models for Credit Risk Prediction. *Journal of Cultural Analysis and Social Change*, *10*(4), 1308–1326. https://doi.org/10.64753/jcasc.v10i4.3014

Published: December 08, 2025

ABSTRACT

Credit risk prediction is a crucial task for financial institutions seeking to minimize defaults and maintain financial stability. This study investigates the most relevant predictive variables and compares the performance of three classification model's fuzzy logic, logistic regression, and random forest, using features selected through variable importance analysis with the random forest method. Based on a dataset of 1,000 Tunisian firms (500 creditworthy and 500 non-creditworthy) described by fourteen financial and non-financial variables, the selected predictors were used as inputs for all three models. Performance was evaluated through accuracy, Type I and II errors, AUC (Area Under the Curve), Cohen's Kappa, and Fuzzy Average Absolute Error (FAE). Results indicate that fuzzy logic achieved the highest accuracy (98.67%), zero Type I error, a low Type II error rate (2.7%), and excellent AUC (0.99) and Kappa (0.97) scores. However, it recorded a higher FAE (1.2645), suggesting a trade-off between classification accuracy and absolute prediction error. Logistic regression and random forest yielded lower FAEs (0.1997 and 0.2266, respectively) but slightly lower accuracies (87% and 90%). Overall, combining variable selection with multiple classifiers enhances both interpretability and performance assessment, and while fuzzy logic delivers superior classification results, metrics like FAE offer valuable insights into model behavior.

Keywords: Credit risk, Default prediction, Logistic regression, Random forests, Variable selection.

INTRODUCTION

One of the main problems faced by financial institutions is risk management, particularly credit risk management. Technological advances now enable the analysis of a large number of borrowers, thereby reducing the risks associated with lending for banks (Nallakaruppan et al., 2024). Credit risk assessment essentially involves estimating the probability that a borrower will fail to meet financial obligations, which is crucial for lenders to manage portfolio risk and ensure financial stability (Roy et al., 2025).

In an economic context characterized by intense competition, increased market volatility, and unforeseen events, risk mastery becomes a strategic imperative for these institutions. Historically, banking crises have dramatically highlighted the consequences of poor credit risk management (Shi et al., 2022). They generate not only direct financial losses but also impacts on market confidence, bank liquidity, and macroeconomic stability (Caprio et al., 1998; Campbell, 2007). The non-performing loan phenomenon thus remains a key indicator of the health of the financial sector, with real implications for banks' ability to continue their activities (Demirguc-Kunt and Detragiache, 1998).

In face of these challenges, academic and professional literature has intensively studied the development of robust models for predicting credit risk. Traditional methods such as logistic regression (Kamstra et al., 2001),

Copyright © 2025 by Author/s. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited..

discriminant analysis (Altman, 1968), or probit regression (Kaplan and Urwitz, 1979) have long been used to improve credit decisions. Credit risk assessment has evolved considerably compared to traditional methods, which relied heavily on human judgment and limited data sources (Abiola, 2025).

The advent of new technologies and artificial intelligence has opened a new era, offering more sophisticated models capable of integrating a wide range of financial and non-financial variables, as well as better capturing nonlinear relationships and the uncertainties inherent in forecasting (Phillips et al., 2023; Suhadoinik et al., 2023).

Among these advanced approaches, fuzzy logic has proven particularly promising as it handles data complexity and imprecision through linguistic rules and membership degrees, thus enabling better modeling of human judgments in the credit evaluation process (De Souza, 2025; Shariatmadari et al., 2024). Complementarily, variable selection methods, such as random forests, optimize the relevance of the data considered, improving model accuracy and interpretability (Roy, 2024). By using sophisticated classification models such as Random Forest and XGBoost, banks can analyze vast datasets with accuracy well beyond traditional methods, leading to more informed lending decisions (Fan, L., 2025).

Despite these advances, a major ongoing challenge is identifying the truly decisive variables that explain default risk. This selection is crucial not only for predictive performance but also for understanding underlying mechanisms and managerial decision-making (Marshall et al., 2010). Indeed, an excess of variables can lead to complex, computationally expensive, and hard-to-interpret models, limiting practical applicability. This study contributes by proposing a comparative analysis between three complementary models logistic regression, random forest, and fuzzy logic applied to a representative sample of Tunisian firms. The aim is to identify the most explanatory variables of credit risk while evaluating the performance and relevance of each approach. The objective is also to demonstrate that integrating traditional and intelligent methods achieves an optimal balance between precision and interpretability.

The central research question is: how to effectively combine classical techniques and innovative approaches to improve credit risk prediction while ensuring reliable and operational diagnostics? This study contributes a robust and innovative methodological framework, integrating recent analyses to guide financial institutions in their risk management decisions.

The article is organized as follows: Section 2 describes the methodology and data used; Section 3 presents detailed empirical results; Section 4 concludes with practical implications and future research directions.

PRIOR LITERATURE

Traditional and Machine Learning Approaches to Credit Risk Prediction

Research on credit risk prediction has evolved along two main strands. Traditional statistical models such as discriminant analysis (Altman, 1968), linear regression (West, 1970), probit regression (Kaplan & Urwitz, 1979), and logistic regression (Ohlson, 1980; Kamstra et al., 2001) are still widely used for their simplicity and interpretability. However, these methods are limited by assumptions of linearity and normality, which often reduce predictive accuracy in practice (Huang, 2004). To overcome these limits, AI techniques including neural networks (Alfaro et al., 2008), support vector machines (Huang et al., 2004), fuzzy logic systems (Hoffmann et al., 2002, 2007), and ensemble methods such as Random Forests (Genuer et al., 2010) have been increasingly adopted. While these models enhance accuracy and robustness, they are often criticized as "black boxes" that lack transparency (James et al., 2013). **Chang et al., 2024** demonstrated the effectiveness of Random Forest and Gradient Boosting for credit risk detection while emphasizing the importance of model explainability to strengthen decision-makers' trust.

Predictors and Explainability in Credit Risk Models

Beyond methodological advances, the choice of predictors plays a crucial role. Early studies focused on financial ratios (Beaver, 1966; Altman, 1968; Ohlson, 1980), but their volatility and short-term focus limit explanatory power (Balcaen & Ooghe, 2006). Subsequent research incorporated non-financial variables such as firm size, age, industry, compliance, and management quality (Keasey & Watson, 1987; Grunert et al., 2005; Altman & Sabato, 2007), showing that combining both types improves performance. More recently, attention has shifted to explainability, as regulatory frameworks such as the GDPR require meaningful explanations of automated decisions (Regulation, 2016). Explainable AI (XAI) tools, including Shapley values, LIME, and ALE plots (Apley & Zhu, 2020; Visani et al., 2020), aim to bridge the gap between accuracy and transparency. **Chang et al., 2024** contributed to the literature on explainable credit risk detection for supply chain financing and provided practical implications for financial institutions to guide decision-making.

METHODOLOGY

Sample

Credit risk analysis has been an integral part of the financial sector for years, enabling the recovery of loans and the identification of insolvent clients (Naili et Lahrichi, 2022). Credit risk forecasting is generally estimated using credit risk assessment methods that rely heavily on statistical models, financial ratios, and historical data (Adefabi et al., 2023).

Our methodological approach consists in studying the management credits granted to the Tunisian companies because they make up the majority of the credits granted by the Tunisian commercial banks. All the credit files for this study were collected from a Tunisian bank for the years 2023 to 2024. Of the total number of cases, 384 enterprises were selected, which were broken down as follows: 192 companies have successfully completed their credit obligations and were classified as creditworthy borrowers and 192 companies have been delayed in fulfilling their obligations and were therefore classified as a set of non-creditworthy borrowers.

Insolvency of companies has been defined on the basis of the rules the Basel New Accord (Oreski*et al.*, 2012). In this study, solvent companies should have been committed to payment in term for at least ninety days during the life of the loan. Otherwise, it is considered insolvent. We randomly divided the sample into a training set (about 70%) and a testing set (about 30%).

Variables of the study

We investigate whether the selected independent variables in our study may account for the credit risk variable "RISK" as a model dependent variable. We consider that credit risk depends on several factors. Therefore, each potential factor was the subject of a hypothesis to be confirmed or invalidated. In this section, we present the variables included in our model.

Binary Response Variable

Credit risk (RISK): we measure the credit risk by a binary variable that takes the value 1 in the case of a repayment of the loan at maturity, 0 otherwise (Like Lee *et al.*, 2002; Bekhet and Eletter, 2014; Chang et al., 2024).

Input Variables

The overall improvement in model accuracy should be related to the selection of variables to be used in a model. In this study, we focus on a number of financial and non-financial variables that appear to have been most successful in previous studies.

Financial Variables

Financial profitability (V1): This ratio reflects the effectiveness of equity as a means of financing a business (Lozinskaia*et al.*, 2017). Grammenos*et al.* (2008) stated that the insolvency of a company decreases return as equity increases.

Economic profitability (V2): This ratio represents the overall profitability of the company's assets. Pompe and Belderbeek (2005) asserted that the profitability of firm assets is crucial to bank decisions to attribute a loan and that the most profitable companies represent the least credit risk. As pointed out by Altman (1968), the ultimate existence of a firm is based on the earning power of its assets.

Solvency (V3): This ratio is a structure ratio that indicates the relative proportion of equity used to finance the assets of a client. It was used as an indicator to measure the company's ability to meet its debt obligations. Pindado and Rodrigues (2001) stated that the solvency of the firm depends on its own capital, they consider that the share of equity guarantees the recovery of receivables and avoids financial distress.

Fixed asset turnover (V4): Fixed assets turnover is a measurement of business activity, and is calculated as the ratio of sales to fixed assets. This ratio is an indicator of the effective use of capital by a company to create sales.

Stock turnover (V5): This ratio provides information on the number of times the stock is shot. A high inventory turnover results in greater liquidity and good sales. On the other hand, low turnover can be interpreted as low liquidity due to over-stocking or obsolete stock (Lucic, 2014).

Leverage ratio (V6): This ratio is commonly used to measure a company's financial leverage (and thus financial risk) by calculating the proportion of the company's assets that have been financed with both short-term and long-term debt. The higher the leverage effect, the greater the financial risk is taken by the firm and the higher its probability of non-repayment (Hernandez Tinoco and Wilson, 2013).

Interest coverage (V7): Interest coverage measures the ability of a company to pay interest on debt (Altman and Sabato, 2007). Generally, a value below 2-2.5 suggests that the company could not meet its financial

 © 2025 by Author/s

obligations; a value below this threshold should therefore be considered as a serious warning: the company does not create enough money from its operations, measured by its earnings before interest, taxes and depreciation, to meet its interests. A value greater than 2.5 is interpreted as a company that can generate funds from its operations to cover interest payments (Hernandez Tinoco and Wilson, 2013). A positive and growing value indicates a growing ability of the company to meet its debt obligations.

Non-Financial Variables

Firm size (V8): Another characteristic of the company that emerged as a factor in the default risk assessment is the size of the companies. It is measured by the logarithm of the firm's sales (Titman and Wessels, 1988; Rajan and Zingales, 1995). Ciampi (2015) mentioned that the different levels of turnover influence the probability of financial distress. Psillaki*et al.* (2010) showed that large firms present a lower bankruptcy risk and face less information asymmetry problems; and are thus more creditworthy.

Age of the firm (V9): According to Farinha and Félix (2015) and Chang et al. (2024), we measure the age of the firm by a binary variable equal to 1 for firms under five and 0 for the older firms. Thornhill and Amit (2003) argued that seniority allows for the accumulation of experience and knowledge of both processes and the environment and thus ensures performance. The lack of experience and know-how related to a domain can thus explain the failure of young companies. Koenig (1985) finds young firms more susceptible to disappearance than senior enterprises with a more solid know-how and a longer experience. Volk (2012) found that one of the most important indicators is the age of a company. Young firms tend to default more often because they are more sensitive to different shocks. According to Fougère *et al.* (2012), the risk is very important for the newly created companies. The latter are characterized by lower performance and a strong dependence on the internal market. Their adaptation is very slow and sometimes inefficient in comparison with the adaptations of older firms. As a result, new companies are riskier than the old ones.

Guarantees (V10): Guarantee is one of the instruments used by banks to induce borrowers to undertake safer projects (Stiglitz and Weiss, 1981). Harris and Raviv (1991) claimed that tangible assets minimize credit risk. So, high collateral increases the probability of access to credit. Guarantees can have a beneficial effect on the behavior of borrowers, i.e. increase the probability of repayment by making borrowers less willing to take too much risk. According to Jiménez and Saurina (2004), creditworthy borrowers are committed to more collateral, since their low risk means they are less likely to lose it. Thus, guarantees serve as a signal that allows the bank to avoid the problem of adverse selection caused by information asymmetry between the bank and the borrower at the time of the loan decision. In the context of asymmetric information between the bank and the borrower, banks divide borrowers into two types when creating loan contracts: risky borrowers choose high and unsecured interest rates, while creditworthy ones opt for guarantees and low interest rates.

Loans granted to companies (V11): There is an inverse relationship between the loan amount and the probability of default because larger loans are more closely examined (Jiménez and Saurina, 2004).

Corporate banking relationship duration (V12): Another issue which has attracted considerable interest in the literature is the role of the bank-firm relationship in credit risk. The company which can maintain a long-term relationship with its bank develops a sound reputation in the credit market. It gives a positive signal to the financial market (Diamond, 1991). As a result, the banking relationship is likely to strengthen firms' incentives to repay loans requested at maturity. Menkhoff and Suwanaporn (2007) found that clients with strong relationships with banks are not risky because their loan shortfall is low. Cayseele and Degryse (2000) reported that the duration of the relationship may inform the bank of the borrower's ability to meet its future obligations. Therefore, the banking relationship allows better control of credit risk. We measure the duration of the relationship by a binary variable that takes the value 1 if the duration of the relationship is greater than the average relationship duration in our sample, 0 otherwise. The duration of the relationship corresponds to the difference between the date of granting of credit and the date of entry into relation with the bank.

Number of bank-borrower relationships (V13): Volk (2012) found that defaulting firms usually have a higher number of banks working with them. According to him, the reason may be in the fact that current creditors refuse to lend risky companies with poor financial statements if they do not pay the loan regularly. So, they seek loans from other banks. In general, the credit history of the borrower is not accessible to new creditors, but in Tunisia, banks can have access to the repayment delays of bank loans in the ratio of interbank debt. Foglia *et al.* (1998) showed that the riskiest firms have a diversified bank debt (more banks, less commitment with a single bank).

Legal form (V14): The legal form of the company can be measured by a binary variable equal to 1 if the firm with limited liability (SARL), 0 otherwise. It limits the liability of shareholders and managers by requiring specific constraints on the company's equity (Omri and Bellouma, 2008). In addition, it has an impact on private information provided by managers, their incentive and their ability to shift risk to the bank (Elsas and Krahnen, 1998).

For each of the companies observed, we provide information on the various variables from the literature and the signs provided in Table 1.

Table 1. Variables definition.

Code	Variables	Description	References
Financial variables			
V1	*Financial profitability*	*Net income / Total Equity*	*Lozinskaiaet al. (2017) ; Khemakhem and Boujelbène (2015)*
V2	*Economic profitability*	*EBIT / Total Assets*	*Altman (1968); Alfaro et al. (2008) ; Barbozaet al. (2017)*
V3	*Solvency*	*Equity / Total assets*	*Grunertet al. (2005); Altman et al. (2015); Ben Jabeur (2017)*
V4	*Fixed asset turnover*	*Sales/ Fixed assets*	*Liang et al. (2016)*
V5	*Stock turnover*	*Sales / Net stocks*	*Lucic (2014); Mselmiet al. (2017)*
V6	*Leverage ratio*	*Total debt/Total assets*	*Beaver (1966); Dewaelheyns and Van Hulle (2004)*
V7	*Interest coverage*	*EBITDA /Interest expenses*	*Hernandez Tinoco and Wilson (2013); Altman and Sabato (2007)*
Non-financial variables			
V8	*Firm size*	*Natural logarithm of firm's sales*	*Titman and Wessels (1988); Rajan and Zingales (1995); Brick and Palia (2007)*
V9	*Age of the firm*	*Binary variable equal to one for firms less than or five years old, and zero for older firms.*	*Farinha and Félix (2015); Chang et al. (2024),*
V10	*Guarantees*	*Natural logarithm of guarantees*	*Karaa and Krichene (2012)*
V11	*Loans granted to companies*	*Amount of total credit*	*Jiménez and Saurina (2004)*
V12	*Corporate banking relationship duration*	*Binary variable equal to 1 if the duration of the relationship is greater than the average duration of the relationship in our sample, 0 otherwise*	*Bousaada (2012)*
V13	*Number of bank-borrower relationships*	*Number of banks with which a firm has relationships*	*Volk (2012); Jiménez and Saurina (2004)*
V14	*Legal form*	*1 = SARL; 0 otherwise*	-

Models for Credit Default Prediction

The proposed framework is structured into three main steps. In the first step, the Random Forest (RF) algorithm is applied to select the most relevant predictor variables based on their importance scores. In the second step, these selected variables are used as inputs to a logistic regression (LR) model to evaluate their impact on credit risk and ensure interpretability. In the third step, the same subset of variables is used in two distinct predictive models: a fuzzy logic system and a Random Forest classifier. This approach allows for a consistent and fair comparison of three classification techniques—Logistic Regression, Random Forest, and Fuzzy Logic—based on the same input variables, thus assessing their respective performances in predicting credit risk.

Random Forests

RF is a popular and highly efficient algorithm introduced by Breiman (2001) and Chang et al. (2024), based on model aggregation ideas, for both classification and regression issues. RF uses a combination bootstrap aggregation (Breiman, 1996) and the random subspace method (Ho, 1998) at each division to generate multiple classes or trees. The principle is to first generate several bootstrap samples $L_1, L_2, \dots, L_n$. Then, for each set of bootstrap data, a tree is constructed as follows: to split a node of a tree, we randomly draw a number of predictive variables and look for the best break only according to the selected variables to find the best distribution of the node. In addition, the tree is thus developed until it reaches its maximum size. No pruning is done. The end result is the average of the results of all trees (Breiman, 2001; Su, 2024). When the result is a class (here solvent or not solvent), the final forecast is the class with the maximum votes. The probability of pulling above is uniform, so it is the same for all trees. The procedure is to build a powerful predictor that better predicts the classes of new individuals. Therefore, RF involves a combination of many trees generated by bootstrap samples and each built independently.

To build the RF method, we follow the steps below (Breiman 2001):

Draw bootstrap samples from the original data. This parameter is named tree and its default value is 500;

Develop a tree for each bootstrap sample. At each node of the tree, randomly select the number of predictors for the slice. Grow the tree so that each terminal node has at least the same number as node size cases;

Aggregate information from the tree trees for new data prediction such as majority voting for classification;

Calculate an out-of-bag (OOB) error rate using the data set that was not used in the bootstrap sample. RF provides measures of varying importance that could be used to rank predictors based on their relationship to the dependent variable. The most popular and reliable criterion is the decrease of classification accuracy when the values in a tree node are randomly exchanged in the OOB sample for that tree (Breiman 2001).

The importance of the variable (VI) equals:

 © 2025 by Author/s

$VI^{(t)}(X_j) = \frac{1}{ntree}\sum_{t=1}^{ntree}(e\widetilde{OOB}_{tj} - eOOB_{tj})$(1)

Where$ntree$ is the number of trees in the forest, $e\widetilde{OOB}_{tj}$ and $eOOB_{tj}$mean the average error of a treet on the whole sample OOBbefore and after randomly permuting the values of$X_j$respectively.

An important advantage of these non-parametric models is the ability to quantify the explanatory power of each variable. Methods like RF can add valuable information about variable importance in terms of mean decrease accuracy. The importance of each variable in the prediction is measured by the accuracy of each tree.

Logistic Regression

LR was applied to the forecast for the first time by Ohlson (1980) and Ayed and Bougatef (2024). In this model, the endogenous variable y is a qualitative variable that is explained by a linear regression, β is a vector of the constant and x is a vector of the explanatory variables:

$y_i = \beta x_i + \varepsilon_i$where$\varepsilon_i$expresses the error associated with the company yi (2)

The endogenous variable y is a dichotomous variable which takes the value 0 or 1 depending on whether the company is solvent or not. If y is 0 for non-creditworthy and 1 for creditworthy companies, then the estimated model is:

$$y_i = \begin{cases} 1\ if & \beta x_i + \varepsilon_i > 0 \\ 0\ if & \beta x_i + \varepsilon_i \leq 0 \end{cases}$$

Posteriori probabilities are decision aids, allowing the construction of risk classes. The subsequent probability that the enterprise i is non-creditworthy is equal to:

$P\ (y_i\ =\ 0)\ =\ P\ (\ \ \beta x_i + \varepsilon_i\ \leq 0)\ =\ P\ (\varepsilon_i\ \leq\ -\ \beta x_i)\ =\ F\ (-\ \beta x_i)$ (3)

Where $F$is the error distribution function ε. We remind that the normal law and the logistic law are symmetric, the error ε is a zero-expectation.

The posterior probability that the firm i is solvent is equal to:

$P\ (y_i = 1) =\ P\ (\beta x_i + \varepsilon_i > 0) = P\ (\varepsilon_i > -\ \beta x_i) = 1 - P\ (\varepsilon_i\ \leq -\ \beta x_i) = F(\beta x_i)$ (4)

Hence:

$$\begin{cases} P\ (y_i = 0) = F(-\beta x_i) \\ P\ (y_i = 1) = F(\beta x_i) \end{cases}$$

The logit model assumes that the errors follow a logistic law, their distribution function is written:

$F\ (x) = (1 + e^{-x})^{-1}$ (5)

The posterior probability that enterprise i is non-creditworthy is then written:

$P\ (y_i = 0) = \left(1 + e^{-\beta x_i}\right)^{-1}$ (6)

Fuzzy Logic

Fuzzy logic, introduced by Zadeh (1965), goes beyond classical binary logic by incorporating intermediate degrees. Initially applied in technical sciences, it gradually expanded into economics and finance, where decisions are inherently multidimensional. For instance, Bojadziev and Bojadziev (2007) argue that a 13% dividend payout ratio cannot simply be labeled “low” or “high,” since its interpretation varies between shareholders and analysts.

In finance, fuzzy logic is particularly relevant given the incomplete and constantly changing nature of information. Duc and Thien (2013) highlight that decisions are made under uncertain and “fuzzy” conditions, while Falavigna (2006) and De Souza (2025) describe it as an effective method for representing complex realities, such as applicant evaluation, system construction, and account management. Nevertheless, they note that its rules are case-specific, which reduces its objectivity.

Fuzzy Sets

Traditional risk models are based on probability and classical set theory, where an element either belongs or does not belong to a set (0 or 1), and are widely used in market, credit, and insurance assessments. In contrast, fuzzy logic models, developed from Zadeh’s (1965) fuzzy set theory, are especially useful when data are imprecise or information is insufficient. In a fuzzy set, each element has a membership degree between 0 and 1, allowing objects to belong to multiple sets with varying levels of truth or confidence, making the model more intuitive and closer to human reasoning. According to Nguyen (2005), a fuzzy set is defined as a collection where each element

x is assigned a real value μ(x) ∈ [0, 1], determined by a membership function that evaluates the extent of association between the element and the set.

The most common representation of fuzzy logic is as follows:

$$\mu_A(X) = \{(x, \mu A(x)) |\, x \in X\}$$

Where:

X: An element belonging to set A

$\mu A(x)$: The membership function

$\mu_A(X)$: The degree of membership of x

Membership Functions

Fuzzy logic, as an extension of classical logic, relies on membership functions that make an element's state continuous and allow for more precise evaluation. A fuzzy set groups members with varying degrees of belonging, such as a borrower's risk level expressed between "low" and "high." These functions are often triangular or trapezoidal, sufficient for most applications, but they may also be Gaussian or sigmoid. Their construction depends on the knowledge and experience of the designer (Shang and Hossen, 2013).

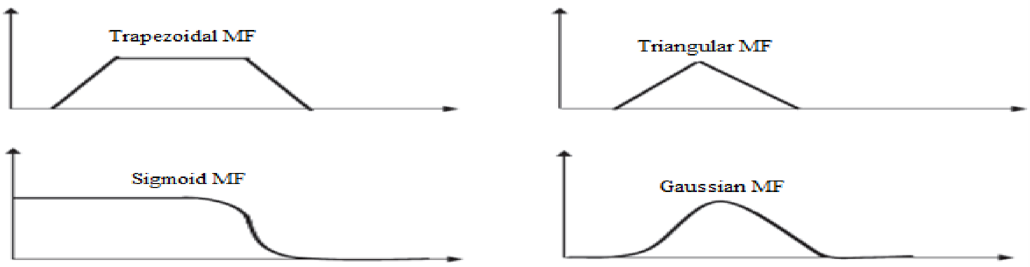

Figure 1. Membership Functions (MFs).

Linguistic Variables

The concept of membership functions, as previously discussed, allows us to define fuzzy systems in natural language. The membership function serves as a bridge between fuzzy logic and linguistic variables. Fuzzy logic relies on fuzzy variables, known as linguistic variables, which take on linguistic values within the universe of discourse U. Each linguistic value represents a fuzzy set within this universe. Chaque valeur linguistique constitue alors un ensemble flou de l'univers du discours.

Inference Rules

Using the logical operators of fuzzy sets, inference rules can be constructed to establish relationships between the input fuzzy variables of a system and its output fuzzy variables. These rules are typically expressed in the following form:

If $x_1$is $A_1$and $x_2$is A_2…and $x_n$is A_n than y is B

That x_1, x_2…and x_n are the linguistic input variables with value of $A_1$to A_n and y is the linguistic output variables with value of B.

Fuzzy Inference System

Bezděk (2014) and Shariatmadari et al. (2024) noted that fuzzy logic is more suitable than classical logic for a range of real-world decision-making tasks, as it simplifies the design of complex control systems. According to these authors, the entire system operates through the following three stages: Fuzzy logic systems accept fuzzy variables as inputs and produce fuzzy variables as outputs. Therefore, it is necessary to fuzzify the input measurements and defuzzify the outputs to obtain precise results.

Fuzzification converts numerical variables into linguistic ones by defining membership functions. **Fuzzy inference** aggregates fuzzy rules by evaluating their truth degrees and assigning corresponding output values (Guenounou, 2009). **Defuzzification** transforms the fuzzy output set into a precise numerical value.

 © 2025 by Author/s

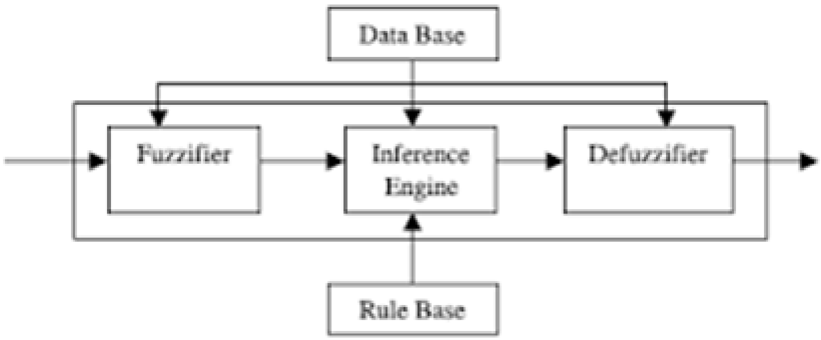

Figure 2. The structure of fuzzy expert systems.

Fuzzy Logic System

With all components, a fuzzy logic system can be built through five steps (Shang and Hossen, 2013): first, selecting independent variables as determinants of the dependent variable; second, creating fuzzy sets for these variables using membership functions to express them linguistically; third, constructing inference rules; fourth, generating fuzzy outputs and applying defuzzification to obtain numerical results; and finally, using these results for decision-making.

Performance Evaluation of Credit Risk Prediction Models

Since classification errors can cause financial losses, evaluating the discriminating power of credit rating systems is crucial for banks. Several performance measures are used to assess the generalization power of forecasting models, including accuracy, type I and type II errors, Cohen's Kappa, Fuzzy Average Absolute Error (FAE), and the Area Under the ROC Curve (AUC). Accuracy measures the proportion of correct classifications:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

Type I error corresponds to solvent firms incorrectly classified as insolvent:

$$Type\ I\ error = \frac{FN}{TP+FN} \quad (8)$$

Type II error refers to insolvent firms incorrectly classified as solvent, which is often more critical due to higher financial costs:

$$Type\ II\ error = \frac{FP}{TN+FP} \quad (9)$$

Where:

- TP (True positives): creditworthy companies classified correctly
- TN (true negatives): non-creditworthy companies classified correctly
- FP (false positives): number of borrowers who default on their loans but are recognized by the model as creditworthy customers
- FN (false negatives): number of borrowers who repay credits but are recognized by the model as non-creditworthy customers.

Cohen's Kappa is a statistic that measures the agreement between two raters or models, while accounting for the agreement that could occur by chance. It is an important metric, especially in classification tasks, to evaluate the performance of a model or the level of agreement between two annotators.

$$Kappa = \frac{Po-Pe}{1-Pe} \quad (10)$$

Where:

Po is the observed agreement, the proportion of instances where the predicted labels match the true labels.

Pe is the expected agreement by chance, the proportion of agreement that would be expected if predictions were made randomly, based on the distribution of the classes.

Fuzzy Average Absolute Error (FAE) is a metric that measures the average absolute error in a fuzzy logic context. Unlike classical error metrics, which are generally defined in a crisp (precise) data framework, FAE takes into account the uncertainty and granularity of values in fuzzy systems.

$$FAE = \frac{1}{N}\sum_{i=1}^{N} \mid y\hat{}_i - y_i \mid \quad (11)$$

Where :

N is the number of companies (or observations).

$y\hat{}_i$ is the fuzzy prediction (i.e., the value inferred by the fuzzy model).

y_i is the actual observed value.

Finally, the ROC curve plots sensitivity against 1-specificity, and its AUC (ranging from 0.5 to 1) indicates predictive power, with values closer to 1 reflecting stronger performance (Barboza et al., 2017).

DESCRIPTIVE STATISTICS AND EMPIRICAL RESULTS

Descriptive Statistics

A Pearson correlation test was used to study the relationship between the independent variables. The obtained results are summarized in Table 2. The results show that the correlation between the variables is relatively low from 0.001 to 0.150. It is within an acceptable interval i.e. <0.50; meaning that there is no problem of multicollinearity.

Table 2. Correlation matrix.

Variables	*V1*	*V2*	*V3*	*V4*	*V5*	*V6*	*V7*	*V8*	*V9*	*V10*	*V11*	*V12*	*V13*	*V14*
V1	1.000													
V2	0.132	1.000												
V3	-0.034	0.066	1.000											
V4	-0.003	0.026	-0.083	1.000										
V5	0.006	0.036	0.076	-0.027	1.000									
V6	0.016	0.043	0.026	-0.018	0.031	1.000								
V7	-0.078	-0.012	0.051	0.024	-0.029	-0.088	1.000							
V8	-0.023	-0.051	0.032	0.012	0.009	-0.002	0.150	1.000						
V9	0.053	-0.019	-0.051	-0.001	-0.043	-0.033	0.080	0.118	1.000					
V10	0.031	0.020	-0.014	-0.026	-0.115	0.055	0.025	-0.111	-0.051	1.000				
V11	0.112	-0.004	-0.041	-0.032	-0.016	-0.082	0.045	0.094	0.046	-0.093	1.000			
V12	-0.023	-0.057	0.028	0.035	-0.063	-0.019	0.076	0.310	0.061	-0.012	-0.018	1.000		
V13	0.102	0.004	0.007	0.144	-0.049	0.077	-0.121	-0.105	-0.118	-0.005	-0.050	-0.115	1.000	
V14	0.032	0.024	0.016	0.052	-0.119	-0.111	-0.113	-0.003	-0.032	-0.110	0.010	0.010	0.013	1.000

To further verify that the multi-collinearity problem is not raised in this study, a variance inflation factor (VIF) is reported in Table 3. As a general rule, a variance inflation factor (VIF) greater than 10 may indicate multi-collinearity and require further investigation (Sayari and Mugan, 2017). VIF ranges from 1.038 to 2.044, which is less than 10, indicating that there is no multicollinearity problem in this study.

Table 3. The result of the variance inflation factor analysis on the chosen variables.

Variables	*VIF*
V1	1.522
V2	2.044
V3	1.569
V4	1.159
V5	1.169
V6	1.074
V7	1.172
V8	1.038
V9	1.131
V10	1.265
V11	1.040
V12	1.094
V13	1.113
V14	1.045

Empirical Results

The computations in our study were done with the use of R. The package randomForest is used for RF, and the package MASS is used to train the LR model.

Random Forest Method

 © 2025 by Author/s

In order to find the relative importance of the explanatory variables, RF are used in this study. We can be interested in constructing a hierarchy of all the variables based on their importance vis-à-vis the response variable. The selection of the most relevant variables to be included in the final model is obtained by ranking the variables according to their decrease in accuracy from most important to least important. For this purpose, we use the index mean decrease accuracy to measure the importance of a variable. This measure is considered very reliable (Genuer *et al.*, 2010).

The two main parameters to construct RF are ntree and mtry, the number of input variables randomly chosen at each split. Thus, RF is built with 500 trees. Variables are listed in Table 4 in descending order of importance.

Table 4. Variable importance classification.

Variables	Mean Decrease Accuracy
V7	77,6351
V2	40,3552
V10	40,3197
V1	37,2252
V11	34,9248
V8	34,0577
V6	30,9508
V12	29,9752
V13	29,5375
V5	28,9382
V9	24,2339
V4	22,1860
V3	21,8353
V14	15,8575

The variable selection procedure is based on the method of Genuer *et al.* (2010). The first step is to rank the variables in a descending order of importance and then remove the variables of low importance. Then, with the selected variables, we compare the OOB error of nested RF models (at each step, we add a variable to the model) and select the model with the lowest error. Finally, the last step aims to select the variables sufficient to predict the dependent variable. This step consists of adding (at each level) a variable in the model only if it makes the OOB error sufficiently small.

An application of this procedure showed us that the importance values of the last three variables (V14, V3, V4) are much less important than the other values. Therefore, we select the rest of the 12 most important variables that are strongly related to the response variable for interpretation (Chan, 1998) and to construct credit risk prediction models (Figure2).

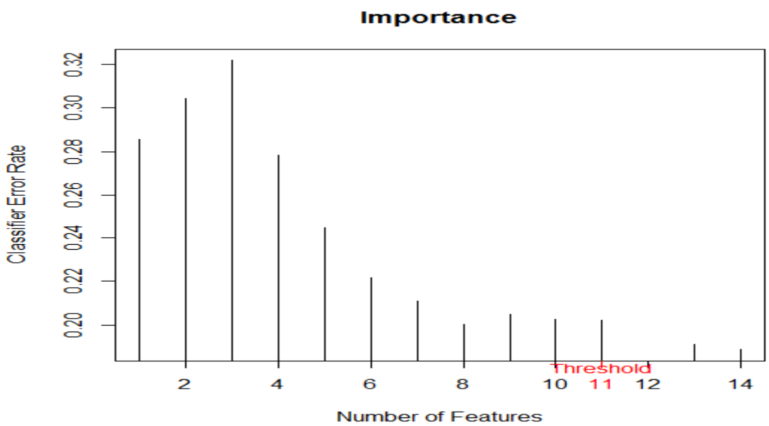

Figure 3. Threshold for variable Selection.

Logistic Regression

After selecting significant independent variables, the LR method is implemented. The eleven most important variables are selected as attributes for classification. Table 5 presents the results of the LR. This study tests the contribution of financial and non-financial variables to develop a model for estimating the likelihood of credit default.

Table 5. RF - LR model.

Variable	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	0,77040	1,49869	0,51405	0,60722
V1	0,81477	0,28235	2,88563	0,00391**

V2	- 5,16398	0,80609	- 6,40625	0,00000***
V5	0,01602	0,00333	4,81077	0,00000***
V6	0,17764	0,05468	3,24843	0,00116 **
V7	- 0,20864	0,02454	- 8,50138	0,00000***
V8	- 0,32003	0,08067	- 3,96734	0,00007***
V9	- 2,05275	0,25801	- 7,95593	0,00000***
V10	0,59488	0,07483	7,95027	0,00000***
V11	- 0,00051	0,00014	- 3,66628	0,00025***
V12	- 2,91593	0,60740	- 4,80068	0,00000***
V13	0,56201	0,12893	4,35917	0,00001***

Note: Signif, codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ; the model is estimated on the training group.

As shown in Table 5, a negative coefficient indicated an inverse relationship to the probability of default, while a positive coefficient indicated a direct relationship to the probability of non-payment.

V1 (0.815): A Positive coefficient indicates that higher financial profitability raises the probability of being solvent. Firms with higher return on equity are more likely to repay loans.

V2 (-5.164): Economic profitability negatively impacts default probability; more profitable companies are more likely to be creditworthy (Alfaro et al., 2008).

V5 (0.016): Higher stock turnover is associated with a higher likelihood of being solvent. Rapid inventory turnover reflects strong sales and liquidity.

V6 (0.178): Leverage ratio positively affects the probability of non-repayment; highly indebted companies face more risk. Low leverage indicates equity-financed assets.

V7 (-0.209): Interest coverage negatively affects default probability; a higher ability to pay interest reduces credit risk. *This variable is significant for lenders and rating agencies.*

V8 (-0.320): Firm size negatively impacts default probability; small Tunisian companies face higher risk (Jiménez and Saurina, 2004; Antão and Lacerda, 2011; Psillaki et al., 2010). Small firms rely on fewer projects and are more financially constrained (Bernanke et al., 1996).

V9 (-2.053): Older firms have lower default probability; younger firms are less solvent and profitable (Cole, 2013).

V10 (0.595): Guarantees are dominant determinants of credit risk; banks require more collateral from risky firms (Jiménez and Saurina, 2004; Manove and Padilla, 1999, 2001).

V11 (-0.0005): Loans granted to companies negatively relate to non-repayment; higher scrutiny for large loans reduces credit risk.

V12 (-2.916): Longer bank-firm relationships reduce default probability; close relationships mitigate information asymmetries (Belaid et al., 2017).

V13 (0.562): More bank-borrower relationships increase the probability of default; risky companies seek credit from multiple banks.

Table 6 presents the overall significance test of the model.

Table 6. Significance test of the model.

Chi-square test	*Degree of freedom*	*P-Value*
724.1997	11	3.493893e-148

The likelihood ratio test compares the full logistic regression model to a null model (with intercept only). The result yields a Chi-square statistic of 724.20 with 11 degrees of freedom, and a 3.493893e-148. This very small p-value indicates that the full model provides a significantly better fit to the data than the null model. In other words, the set of explanatory variables jointly and significantly improves the prediction of the probability of loan repayment.

The Fuzzy Logic Model

In the fuzzy logic model, the explanatory variables selected by RF are represented through three modalities (low, medium, high), allowing each firm to belong simultaneously to several categories with different degrees. This flexible representation captures uncertainty in financial and structural evaluation and serves as the basis for fuzzy inference, which assesses credit risk more realistically than traditional crisp models. Table 7 presents the parameters of the membership functions for each variable.

 © 2025 by Author/s

Table 7. Fuzzy Membership Functions for Each Variable.

Variable	Fuzzy Set (Low)	Fuzzy Set (Medium)	Fuzzy Set (High)
V1: Financial Profitability	(0, 0, 3, 5)	(4, 6, 6, 8)	(7, 9, 11, 11)
V2: Economic Profitability	(0, 0, 4, 6)	(5, 7, 7, 9)	(8, 10, 12, 12)
V5: Stock Turnover	(0, 0, 1, 3)	(2, 4, 4, 6)	(5, 7, 10, 10)
V6: Leverage Ratio	(0, 0, 1, 2)	(1.5, 2, 2.5, 3)	(2.5, 3, 4, 4)
V7: Interest Coverage	(0, 0, 1, 2)	(1.5, 2, 3, 4)	(3, 4, 5, 5)
V8: Firm Size	(0, 0, 10, 20)	(15, 25, 25, 35)	(30, 40, 50, 50)
V9: Age of the Firm	(0, 0, 2, 5)	(3, 6, 6, 9)	(7, 10, 20, 20)
V10: Guarantees	(0, 0, 10, 20)	(15, 25, 25, 35)	(30, 40, 50, 50)
V11: Loans Granted to Companies	(0, 0, 100, 300)	(200, 400, 400, 600)	(500, 700, 1000, 1000)
V12: Banking Relationship Duration	(0, 0, 2, 5)	(4, 6, 6, 8)	(7, 9, 10, 10)
V13: Bank-Borrower Relationships	(0, 0, 1, 2)	(1.5, 3, 3, 5)	(4, 6, 8, 8)

The graphs of fuzzy membership functions allow you to visualize how companies are distributed across different categories for each variable. These visualizations are very useful for understanding areas of certainty and uncertainty in credit assignment decisions. If a company falls within areas of high membership in low-risk categories (such as high values in profitability or size), it may be perceived as less risky. Conversely, companies with low scores in key variables may appear riskier. This provides an overview of companies' behavior in terms of credit risk.

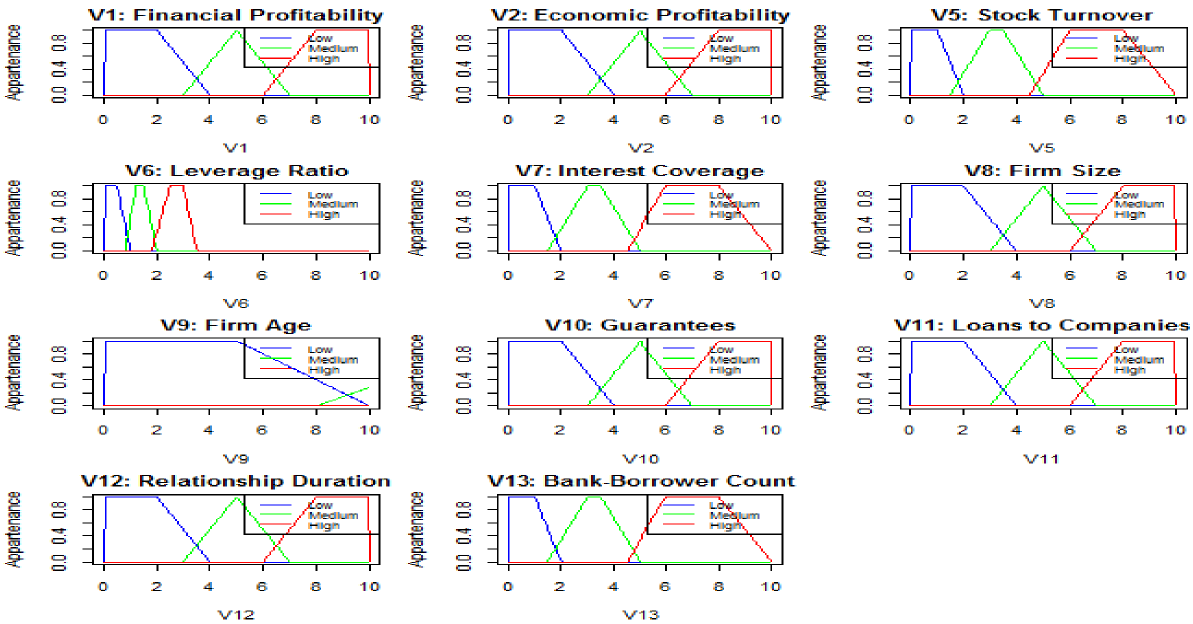

Figure 4. Fuzzy Membership Functions for All Explanatory Variables.

After defining the fuzzy membership functions for each variable, it is essential to formulate fuzzy rules that link these variables to credit risk assessment. These rules are developed by considering the values of various economic and financial variables, associating specific conditions (IF...) with conclusions (THEN...). Fuzzy rules enable a more flexible modeling of credit risk, capturing uncertainties and complex relationships between different variables, which is crucial in a real economic environment.

The following rules are examples of how each variable can be used to evaluate a company's credit risk. They take into account factors such as profitability, company size, debt levels, interest coverage, and other key financial and economic elements

Table 8. Fuzzy Rules for Credit Risk Assessment.

Rule No.	Conditions (IF...)	Conclusion (THEN...)
R1	IF Financial profitability is low AND Debt ratio is high AND Stock turnover is low	THEN Credit risk is high
R2	IF Economic profitability is medium AND Interest coverage is low AND Firm size is small	THEN Credit risk is high
R3	IF Financial profitability is high AND Debt ratio is low AND Duration of banking relationship is long	THEN Credit risk is low
R4	IF Firm age is high AND Collateral is strong AND Loans granted are moderate	THEN Credit risk is low
R5	IF Economic profitability is low AND Stock turnover is low AND Debt ratio is high	THEN Credit risk is high
R6	IF Financial profitability is medium AND Firm size is medium AND Number of bank relationships is high	THEN Credit risk is meduim
R7	IF Economic profitability is high AND Debt ratio is low AND Loans granted are high	THEN Credit risk is medium
R8	IF Interest coverage is high AND Leverage ratio is low AND Collateral is strong	THEN Credit risk is low

© 2025 by Author/s

After fuzzy inference, the results are defuzzied to obtain a single numerical score that classifies firms into credit risk levels (low, medium, high). Using methods such as the centroid approach, this step transforms imprecise fuzzy values into a clear and actionable measure for decision-making. The credit risk score allows financial institutions to adjust loan approvals, interest rates, and risk management. the scores and categories, confirming that the fuzzy model provides a flexible, transparent, and effective tool for credit risk assessment.

Table 9. Excerpt of Credit Risk Scores and Categories Assigned to a Sample of Enterprises.

Company	*Risk Score*	*Risk Category*
Company _1	*9.23*	*High*
Company _2	*9.43*	*High*
Company _3	*3.58*	*Low*
Company _4	*8.47*	*High*
Company _5	*6.78*	*Medium*
Company _6	*5.67*	*Medium*
Company _7	*7.63*	*High*
Company _8	*2.21*	*Low*
Company _9	*6.91*	*Medium*
Company _10	*7.35*	*High*

The table reports credit risk scores and categories for 10 companies. Results show three levels of risk: High, Medium, and Low.

High risk: Companies 1, 2, 4, 7, and 10 have scores between 7.35 and 9.43, indicating greater vulnerability to default. These firms are perceived as less reliable by lenders and may face financial or economic difficulties.

Medium risk: Companies 5, 6, and 9 fall within 5.67 to 6.91, reflecting moderate risk. They possess some financial potential but remain sensitive to economic changes or internal weaknesses.

Low risk: Companies 3 and 8, with scores of 3.58 and 2.21, are considered the safest, showing better solvency, stable management, and stronger repayment capacity.

Overall, this classification highlights the model's ability to distinguish between firms with high default probability and those with stronger financial health, thus supporting lenders in credit decision-making.

Evaluation of Model Performance

The table below displays the results obtained for three classification models: logistic regression, random forest, and fuzzy logic. These models were assessed using various metrics, including accuracy, type I error (false positives), type II error (false negatives), AUC (area under the ROC curve), and Cohen's Kappa.

The comparative analysis shows that fuzzy logic outperforms logistic regression and random forest. It achieves the highest accuracy (98.67%), compared to 90% for random forest and 87% for logistic regression. Error rates confirm this superiority, with no type I errors and only 2.7% type II errors, while random forest records 5% and logistic regression 11.3% type I errors. Performance indicators also validate these results: AUC reaches 0.99 for fuzzy logic (vs. 0.92 and 0.87) and Cohen's Kappa 0.97 (vs. 0.85 and 0.74). Overall, fuzzy logic proves to be the most reliable and accurate model for credit risk assessment.

Table 10. The performance of different prediction methods.

Model	**Accuracy**	**Type I Error (%)**	**Type II Error (%)**	**AUC**	**Kappa**
Logistic Regression	87%	11.3%	14.7%	0.87	0.74
Random Forest	90%	5%	10%	0.92	0.85
Fuzzy Logic	98.67%	0%	2.7%	0.99	0.97

Analysis of the Fuzzy Average Absolute Error (FAE) shows that the fuzzy logic system achieves the best predictive performance among the three models, with an FAE of 0.1325, indicating its predictions are, on average, closest to actual values in a fuzzy context. It is followed by logistic regression, which has an FAE of 0.1997, demonstrating good accuracy but slightly lower performance. Although the random forest model performs well on other traditional metrics, it has the highest average error here, with an FAE of 0.2266, indicating slightly less precise predictions in terms of fuzzy error. These results suggest that, according to this specific metric, fuzzy logic provides a finer generalization capability, likely due to its flexible handling of uncertainty and transitional areas between classes.

Table 11. Performance of Different Prediction Methods.

Model	**Fuzzy Average Absolute Error (FAE)**
Logistic Regression	0.1997
Random Forest	0.2266
Fuzzy Logic System	1.2645

 © 2025 by Author/s

Figure 5 shows the fuzzy membership functions of the fuzzy average absolute error (FAE) for the three models. The fuzzy logic model (FAE = 0.1325) reaches the maximum degree of membership, confirming its precision and stability, while logistic regression (0.1997) and random forest (0.2266) display lower membership degrees, indicating less accuracy. This visualization highlights the superior performance of fuzzy logic in minimizing prediction errors.

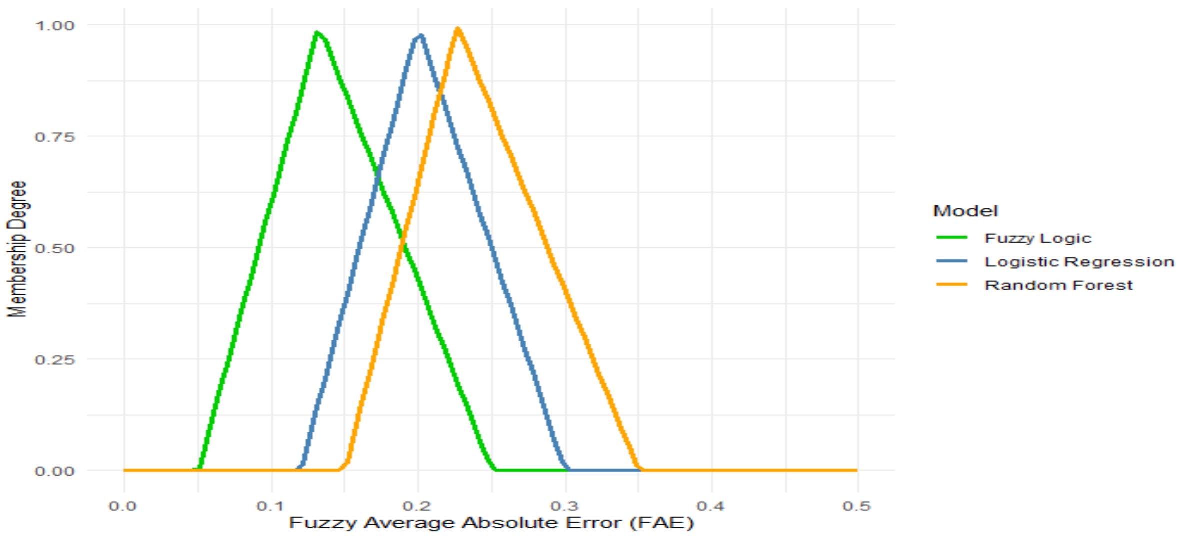

Figure 5. Fuzzy Membership Functions for Fuzzy Absolute Error (FAE) across Models.

CONCLUSION

This study proposed a hybrid approach combining variable selection via Random Forests with three predictive models: logistic regression, random forest, and fuzzy logic to predict credit risk among Tunisian companies. Analyzing a sample of 1000 firms, half of which were non-creditworthy, we examined fourteen financial and non-financial variables to assess their influence on default probability.

The results highlight the crucial role of several factors. Financial profitability (V1) and economic profitability (V2) have a significant positive impact on firms' solvency, confirming that more profitable companies have a lower default risk. Similarly, a high stock turnover rate (V5) reflects better liquidity and strengthens repayment capacity. Interest coverage (V7) and firm size (V8) also play protective roles, while company age (V9) significantly reduces the probability of default.

Conversely, variables such as a high leverage ratio (V6), multiple banking relationships (V13), and the level of guarantees requested (V10) increase the probability of default, underlining the vulnerability of highly indebted or risky firms. The length of banking relationships (V12) emerges as a key factor in reducing risk, illustrating the importance of trust and information sharing between banks and companies.

The results demonstrate that the fuzzy logic model clearly outperforms the other two in predictive performance. Indeed, fuzzy logic achieves an accuracy of 98.67%, with a zero type I error rate (0%) and a very low type II error rate (2.7%). The AUC reaches 0.99, and the Kappa coefficient is 0.97, indicating excellent discrimination ability and strong agreement between predictions and observed outcomes.

In comparison, the random forest model achieves 90% accuracy, with type I and II errors of 5% and 10%, respectively, while logistic regression yields 87% accuracy with higher error rates (11.3% and 14.7%). These performances are corroborated by the AUC values (0.92 for RF and 0.87 for LR) and Kappa scores (0.85 for RF and 0.74 for LR), confirming the superiority of random forest over logistic regression.

Regarding the fuzzy average absolute error (FAE), the fuzzy logic model shows a higher score (1.2645), reflecting the more flexible and approximate nature of this model compared to classical statistical methods (FAE of 0.1997 for LR and 0.2266 for RF).

These findings demonstrate that fuzzy logic, by combining automatic variable selection and modeling uncertainty, represents a very promising approach for credit risk prediction, outperforming traditional models in accuracy and in handling imprecise information.

This study thus makes an important contribution to financial institutions aiming to improve their credit risk assessment tools by adopting more robust and interpretable models. Future research could focus on incorporating additional variables and applying the methodology to datasets from other countries to further validate the robustness and generalizability of the model. Overall, this study contributes to the advancement of credit risk

© 2025 by Author/s

modeling by demonstrating the effectiveness of hybrid approaches that combine statistical rigor with flexible reasoning under uncertainty.

Secondly, understanding the model is of vital importance in the area of credit risk. Indeed, bankers need interpretable methods to justify the causes of credit rejection or acceptance (Crook et al., 2007; Tomczak and Zieba, 2015). While managing credit risk through traditional methods has proven useful, financial institutions still face a significant number of non-performing loans, necessitating the adoption of other techniques to address this growing concern (Suhadoinik et al., 2023). Our model offers a potential solution. It is both precise and understandable, which facilitates its adoption as a management tool by financial institutions.

Funding: This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2504).
Conflict of interest: No potential conflict of interest was reported by the authors.

REFERENCES

Abiola, T. K. (2025). AI-powered credit risk assessment and algorithmic fairness in digital lending: A comprehensive analysis of the United States digital finance landscape.

Adefabi, A., Olisah, S., Obunadike, C., Oyetubo, O., Taiwo, E., and Tella, E. (2023),"Predicting Accident Severity: An analysis of factors affecting Accident Severity using Random Forest Model",*International Journal on Cybernetics & Informatics (IJCI)*,Vol.12, No.6, pp. 107-121.

Alfaro, E., Elizondo, D., Garcia, N. and Gamez, M. (2008), "Bankruptcy forecasting: An empirical comparison of adaBoost and neural networks", *Decision Support Systems*, Vol. 45, No. 1, pp. 110-122.

Altman, E. I. and Sabato, G. (2007), "Modelling credit risk for SMEs: Evidence from the U.S. market" *Abacus*, Vol. 43, No.3, pp.332–357.

Altman, E.I. (1968), "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy", *Journal of finance*, Vol. 23, No. 4, pp. 589-609.

Altman, E.I., Iwanicz-Drozdowska, M., Laitinen, E.K. and Suvas, A. (2015), "Financial and non-financial variables as long-horizon predictors of bankruptcy" Available at SSRN: https://ssrn.com/abstract=2669668 or http://dx.doi.org/10.2139/ssrn.2669668(accessed 18 September 2017).

Altman, E.I., Sabato, G. and Wilson, N. (2010), "The value of non-financial information in Small and Medium-Sized Enterprise risk management", *Journal of Credit Risk*, Vol. 6, No. 2, pp. 95-127.

Antão, P. and Lacerda, A. (2011), "Capital requirements under the credit risk-based framework", *Journal of Banking and Finance*, Vol. 35, No. 6, pp.1380-1390.

Apley, D. W., & Zhu, J. (2020), "Visualizing the effects of predictor variables in black box supervised learning models", *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol.82, No.4, pp.1059-1086.

Apostolik, R.., Donohue, C. and Went, P. (2009), *Foundations of banking risk: An overview of banking, banking risks, and risk-based banking regulation*, New Jersey:JohnWiley&Sons,Inc.

Ayed, N. and Bougatef, K. (2024), "Performance Assessment of Logistic Regression (LR), Artificial Neural Network (ANN), Fuzzy Inference System (FIS) and Adaptive Neuro-Fuzzy System (ANFIS) in Predicting Default Probability: The Case of a Tunisian Islamic Bank", *Computational Economics*, Vol.64, No.3, pp.1803-1835.

Balcaen, S. and Ooghe, H. (2006), "The exit path of failing firms: Which factors determine the time to exit?", Working Paper Series, Vlerick Leuven Gent Management School.

Barboza, F., Kimura, H. and Altman, E. (2017), "Machine learning models and *bankruptcyprediction*", *Expert Systems with Applications*, Vol. 83, pp. 405-417

Barr, R. and Siems,T. (1994), "Predicting bank failure using DEA to quantify management quality", Financial Industry Studies Working Paper, Federal Reserve Bank of Dallas, available at: http://dx.doi.org/10.1007/978-1-4615-4102-8_15 (accessed 20 September 2017).

Beaver, W.H. (1966), "Financial ratios as predictors of failure", *Journal of Accounting Research*, Vol. 4, pp. 71-111.

Becchetti, L. and Sierra, J. (2003), "Bankruptcy risk and productive efficiency in manufacturing firms", *Journal of Banking and Finance*, Vol. 27, No. 11, pp. 2099-2120.

Bekhet, H.A. and Eletter, S.F.K. (2014), "Credit risk assessment model for Jordanian commercial banks: Neural Scoring Approach", *Review of Development Finance*, Vol. 4, No. 1, pp. 20-28.

Belaid, F., Boussaada, R. and Belguith, H. (2017), "Bank-firm relationship and credit risk: An analysis on Tunisian firms", *Research in International Business and Finance*, Vol. 42, pp. 532-543.

Ben Jabeur, S. (2017), "Bankruptcy prediction using Partial Least Squares Logistic Regression", *Journal of Retailing and Consumer Services*, Vol. 36, pp. 197-202.

 © 2025 by Author/s

Bernanke, B., Gertler, M. and Gilchrist, S. (1996), "The financial accelerator and the flight to quality", *The Review of Economics and Statistics*, Vol. 78, No. 1, pp. 1-15

Bezděk, V. (2014), "Using fuzzy logic in business", Procedia Social and Behavioral Sciences, Vol. 124, p. 371- 380.

Bhattacharyya, S., Jha, S., Tharakunnel, K. and Westland, J.C. (2011), "Data Mining for Credit Card Fraud: A Comparative Study", *Decision Support Systems*, Vol. 50 No. 3, pp. 602-613.

Bojadziev, G. and Bojadziev, M. (2007), "Fuzzy logic for business, finance and management", *World Scientific Publishing*, USA.

Bousaada, R. (2012), "L'impact de la gouvernance bancaire et de la relation bancaire sur le risque de crédit : cas des banques tunisiennes", thèse de doctorat en sciences économiques, université Montesquieu -Bordeaux IV, université de TUNIS.

Breiman, L. (2001), "Random forest", *Machine Learning*, Vol. 45, pp. 5-32.

Brick, I. E. and Palia. D. (2007), Evidence of jointness in terms of relationship lending, *Journal of Financial Intermediation*, Vol. 16, No. 3, pp. 452-476.

Campbell, A., (2007), "Bank insolvency and the problem of non-performing loans", *Journal of Banking Regulation*, vol. 9, No. 1, pp. 25- 45.

Caprio, G., Hunter, W., Kaufman, G. and Leipziger, D. (1998), "Preventing bank crisis: lessons from recent global bank failures", *Economic Development Institute Studies*, the World Bank, Washington, D.C. 377p.

Cayseele, P. V. and Degryse, H. (2000), "Relationship lending within a bank based system: Evidence from European small business data", *Journal of Financial Intermediation*, Vol. 9, No.1, pp. 90-109.

Chan, C.C. (1998), "A rough set approach to attribute generalization in data mining", *Information Sciences*, Vol. 107, No. 1–4, pp. 169-176.

Chang, V., Xu, Q. A., Akinloye, S. H., Benson, V., & Hall, K. (2024). "Prediction of bank credit worthiness through credit risk analysis: an explainable machine learning study". *Annals of Operations Research*, 1-25.

Chen, W.H. and Shih, J.Y. (2006), "A study of Taiwan's issuer credit rating systems using support vector machines", *Expert Systems with Applications*, Vol. 30, No. 3, pp. 427–435.

Ciampi, F. (2015), "Corporate governance characteristics and default prediction modeling for small enterprises. An empirical analysis of Italian firms", *Journal of Business Research*, Vol. 68, No. 5, 1012–1025.

Cole, R.A. (2013), "What do we know about the capital structure of privately held US firms? Evidence from the Surveys of Small Business Finance", *financial management*, Vol. 42, No. 4, pp. 777–813.

Crook, J. N., Edelman, D. B. and Thomas, L. C. (2007), "Recent developments in consumer credit risk assessment", *European Journal of Operational Research*, Vol. 183, No. 3, pp. 1447-1465.

Cultrera, L. and Brédart, X. (2016), "Bankruptcy prediction: the case of Belgian SMEs", *Review of Accounting and Finance*, Vol. 15, No. 1 pp. 101 – 119.

De Souza, G. A. (2025). The use of Fuzzy logic for risk assessment in financial investments. Brazilian Journal of Development, 11(4), e78939-e78939.

Demirguç-Kunt, A. and Detragiache, E. (1998), "Determinants of banking crisis: evidence from developed and developing countries", World Bank and IMF papers, E44 G2, available at: http://dx.doi.org/10.5089/9781451947175.001(accessed 25 September 2017).

Dewaelheyns, N. and Van Hulle, C. (2004), "The impact of business group on bankruptcy prediction modeling", *Tijdschriftvoor Economies and Management*, Vol. 49, No. 4, pp. 623-645.

Diamond, D. W. (1991), "Monitoring and reputation: The choice between bank loans and directly placed debt", *Journal of Political Economy*, Vol. 99, pp. 689-721.

Duc, V. H., Thien, N. D. (2013), "A new approach to determining credit rating & its applications to vietnam's listed firms", *Open University, Ho Chi Minh City, Vietnam*.

Diaz-Uriarte, R. and Alvarez de Andres, S. (2006), "Gene selection and classification of microarray data using random forest", *BMC Bioinformatics*, Vol. 7, No. 3, pp. 1-13.

Elsas, R. and Krahnen, J. P. (1998), "Is relationship lending special? Evidence from credit file in Germany". *Journal of Banking and Finance*, vol. 22, No. 10-11, pp. 1283-1316.

Fabozzi, F. J. and Drake, P.P. (2010), *The basics of finance: An introduction to financial markets*, Business, Finance & Portfolio Management, USA, NJ: John Wiley & Sons.

Farinha, L. and Felix, S. (2015), "Credit rationing for Portuguese SMEs", *Finance Research Letters*, Vol. 14, pp. 167-177.

Falavigna, G. (2006), "Models for Default Risk Analysis: Focus on Artificial Neural Networks, Model Comparisons", *Hybrid Frameworks, Institute for Economic Research on Firms and Growth - CERIS Working Paper.*

Fan, L. (2025). Review of theoretical advancements in AI/ML classification models for credit risk assessment. *Journal of Risk Management in Financial Institutions*, *18*(2).

Foglia, A., Lavolia, S. and MarulloReedtz, P. (1998), "Multiple Banking Relationships and the Fragility of Corporate Borrowers", *Journal of Banking and Finance*, vol. 22, No. 10-11, pp. 1441-1456.

Fougère, D., Golfier, C., Horny, G. and Kremp, E. (2012), "Did the 2008 crisis affect the survival of french firms?", paper presented at the Conference on Firms' Financing and Default Risk during and after the Crisis, February 10, Banque de France, available at: https://www.banque-france.fr/sites/default/files/fougere.pdf. (accessed 25 September 2017).

Genuer, R., Poggi, J.-M. andTuleau, C. (2010), "Variable selection using Random Forests", *Pattern Recognition Letters*, Vol. 31, No. 14, pp. 2225-2236.

Grammenos, C. T., Nomikos, N. K. and Papapostolou, N. C. (2008), "Estimating the Probability of Default for Shipping High Yield Bond Issues", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 44, No. 6, pp. 1123 -1138.

Grunert, J., Noreden, L. and Weber, M. (2005), "The role of non financial indicators in internal credit rating", *Journal of Banking and Finance*, Vol. 29, No. 2, pp. 509-531.

Harris, M. and Raviv, A. (1991), "The theory of capital structure", *The Journal of Finance*, Vol. 46, No. 1, pp. 297-355.

Hernandez Tinoco, M. and Wilson, N. (2013), "*Financial distress and bankruptcy* prediction among listed companies using accounting, market and macroeconomic variables", *International Review of Financial Analysis*, Vol.30, pp. 394-419.

Hoffmann, F. Baesens, B. Martens, J. Put, F. and Vanthienen, J. (2002), "Comparing a genetic fuzzy and a neurofuzzy classifier for credit scoring", *International Journal of Intelligent Systems*, Vol.17, No.11, pp.1067-1083.

Hoffmann, F. Baesens, B. Mues, C. Van Gestel, T. & Vanthienen, J. (2007), "Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms", *European Journal of Operational Research*, Vol.177, pp.540-555.

Huang, Z., Chen, H., Hsu, C. J., Chen, W. H. and Wu, S. (2004), "Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study", *Decision Support Systems*, Vol. 37, No. 4, pp. 543-558.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013), *An Introduction to statistical learning with applications*, Springer, New York.

Jiménez, G. and Saurina, J., 2004, "Collateral, type of lender and relationship banking as determinants of credit risk". *Journal of Banking and Finance*, vol. 28, No. 9, pp. 2191-2212.

Jones, S., Johnstone, D. and Wilson, R. (2015), "An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes," *Journal of Banking and Finance*, Vol. 56, pp. 72-85.

Kamstra, M., Kennedy, P. and Suan, T.K. (2001), "Combining bond rating forecasts using logit", *The Financial Review*, Vol. 36, No. 2, pp. 75-96.

Kaplan, R.S. and Urwitz, G. (1979), "Statistical models of bond ratings: a methodological inquiry", *Journal of Business*, Vol. 52, No. 2, pp. 231-261.

Karaa, A. and Krichène A. (2012), "Credit–risk assessment using support vectors machine and multilayer neural network models: A comparative study case of a Tunisian bank", *Journal of Accounting and Management Information Systems*, Vol. 11, No. 4, pp. 587-620.

Keasey, K. and Watson, R. (1987), "Non-Financial Symptoms and the Prediction of Small Company Failure: A Test of Argenti's Hypothesis", *Journal of Business Finance and Accounting*, Vol. 14, No. 3, pp. 335-354.

Keramati, A., Yousefi, N. (2011), "A Proposed Classification of Data Mining Techniques in Credit Scoring", *Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management Kuala Lumpur, Malaysia, January*, pp. 22-24.

Khemakhem, S. and Boujelbène, Y. (2015), "Credit risk prediction: A comparative study between discriminant analysis and the neural network approach", *Accounting and Management Information Systems*, Vol. 14, No. 1, pp. 60-78.

Koenig G. (1985), "Entreprises en difficultés : des symptômes aux remèdes", *Revue Française de Gestion*, pp. 84-92.

Korol, T. and Korodian, A. (2011), "Evaluation of effectiveness of fuzzy logic model in predicting the business bankruptcy", *Romanian Journal of Economic Forecasting*, pp 92-107.

Kurniawan, R. (2024), "Application of Random Forest Algorithm on Credit Risk Analysis", *Procedia Computer Science*, No.245, pp.740-749.

Larivière, B. and Van den Poel, D. (2005), "Predicting customer retention and profitability by using random forests and regression forests techniques", *Expert Systems with Applications*, Vol. 29, No. 2, pp. 472-484.

Lee, J.W., Lee, J.B., Park, M. and Song, S.H. (2005), "An extensive comparison of recent classification tools applied to microarray data", *Computational Statistics and Data Analysis* Vol. 48, No. 4, pp. 869-885.

Lee, M.C. (2012), "Enterprise Credit Risk Evaluation models: A Review of Current Research Trends", *International Journal of Computer Applications (0975 – 8887)*, Vol. 44, No. 11, pp. 37-44.

Lee, T., Chiu, C., Lu, C. and Chen, I. (2002), "Credit scoring using the hybrid neural discriminant technique", *Expert Systems with Applications*, Vol. 23, No. 3, pp. 245-254.

 © 2025 by Author/s

Liang, D., Lu, C.C., Tsai, C.F. and Shih, G.A. (2016), "Financial Ratios and Corporate Governance Indicators in Bankruptcy Prediction: A Comprehensive Study", *European Journal of Operational Research*, Vol. 252, No. 2, pp. 561-572

Liaw, A. and Wiener, M. (2002), "Classification and regression by randomforest", *R News*, Vol. 2-3, pp. 18-22.

Lucic, L. (2014), "Financial ratios in the function of business risk assessment", *Online Journal of Applied Knowledge Management*, Vol. 2, No. 3, pp.21-34.

Malekipirbazari, M. and Aksakalli, V. (2015), "*Risk*assessment in social lending via*randomforests,Expert Systems with Applications*, Vol. 42, No. 10, pp. 4621-4631

Manove, M. and Padilla, A.J. (1999), "Banking (conservatively) with optimists". *RAND Journal of Economics,* Vol. 30, No. 2, pp. 324-350.

Manove, M. and Padilla, A.J., 2001, "Collateral versus project screening: A model of lazy banks". *RAND Journal of Economics*, Vol. 32, No. 4, pp. 726-744.

Marshall, A., Tang, L. and Milne, A. (2010), "Variable reduction, sample selection bias and bank retail credit scoring", *Journal of Empirical Finance*, Vol. 17, No. 3, pp. 501-512.

Menkhoff, L. and Suwanaporn, C. (2007), "On the rationale of bank lending in pre-crisis Thailand", *Applied Economics*, Vol. 39, No. 9, pp. 1077-1089.

Mselmi, N., Lahiani, A. and Hamza, T. (2017), "Financial distress prediction: The case of French small and medium-sized firms", International Review of Financial Analysis, Vol. 50, pp. 67-80.

Naili, M. and Lahrichi, Y. (2022), "Bankscredit risk, systematic determinants and specific factors: recent evidence from emerging markets",Heliyon, vol.8, no.2.

Nallakaruppan, M. K., Chaturvedi, H., Grover, V., Balusamy, B., Jaraut, P., Bahadur, J., ... & Hameed, I. A. (2024). "Credit risk assessment and financial decision support using explainable artificial intelligence". Risks, 12(10), 164.

Ohlson, J.A. (1980), "Financial ratios and the probabilistic prediction of bankruptcy", *Journal of Accounting Research*, Vol. 18, No. 1, pp. 109-131.

Nguyen, P. (2005), "Fuzzy Logic and Its Applications", *the science and technology publisher*.

Omri, A. and Bellouma, M. (2008), "L'impact de la qualité de la relation banquet-entreprise sur la prime de risqué exigée des entreprises tunisiennes". *La Revue des Sciences de Gestion*, No. 229, pp. 95-102.

Oreski, S., Oreski, D. and Oreski, G. (2012), "Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment", *Expert Systems with Applications*, Vol. 39, No. 16, pp. 12605-12617.

Phillips, A., Ojalade, I., Taiwo, E., Obunadike, C. and Oloyede, K. (2023), "Cyber-Security Tactics in Mitigating CyberCrimes: A Review and Proposal", *International Journal on Cryptography and Information Security (IJCIS).*

Pindado J. and Rodrigues L.F. (2001), "Parsimonious models of financial insolvency in small companies", Working Paper, SSRN Working Paper Series.

Pompe, P.M. and Bilderbeek, J. (2005), "The prediction of bankruptcy of small and medium sized industrial firms", *Journal of Business Venturing*, Vol. 20, No. 6, pp. 847-868.

Psillaki, M., Tsolas, L.E and Margaritis, D. (2010), "Evaluation of credit risk based on firm performance", *European Journal of Operational Research*, Vol. 201, No. 3, pp. 873-881.

Quinlan, J.R. (1993), *C4.5: Programs for machine learning*, Morgan Kaufmann Publishers.

Rajan R. G. and Zingales, L. (1995), "What do we know about capital structure? some evidence from international data", *Journal of Finance*, Vol. 50, No. 1, pp. 1421-1460.

Roy K. (2024) "Application of Random Forest Algorithm on Credit Risk Analysis", *Procedia Computer Science*,vol.245,pp.740-749.

Sayari, N. and Mugan, C.S. (2017), "Industry specific financial distress modeling", *Business Research Quarterly*, Vol. 20, No. 1, pp. 45-62.

Shang, K., Hossen, Z. (2013), "Applying Fuzzy Logic to Risk Assessment and Decision-Making", *Casualty Actuarial Society, Canadian Institute of Actuaries, Society of Actuaries*.

Shariatmadari, S., Homayounfar, M., Azadi, K. and Daneshvar, A. (2024), "Fuzzy Inference System for Credit Scoring: Legal Clients in Banking Industry", *Fuzzy Optimization and Modeling Journal (FOMJ)*, Vol.5, No.1, pp.46-58.

Shi, S., Tse, R., Luo, W., D'Addona, S. and Pau, G, (2022), "Machine learning-driven credit risk: a systematic review, *Neural Computing and Applications*, vol.34, pp.14327-14339.

Su, Y. (2024), "Financial technology credit risk modeling and prediction based on random forest algorithm", *In 2024 Second International Conference on Data Science and Information System (ICDSIS)*, pp. 1-4 IEEE.

Suhadolnik, N., Ueyama, J., & Da Silva, S. (2023), "Machine learning for enhanced credit risk assessment: An empirical approach", *Journal of Risk and Financial Management*, Vol.16. No.12, pp.496.

Stiglitz, J. and Weiss, A. (1981), "Credit rationing in markets with imperfect information", *American Economic Review*, Vol. 71, No. 3, pp. 393-410.

Tetteh, M. L., Tetteh, F.A. and Ansah, A.Y. (2014), "Credit default modeling: a logit approach", *Research Journal of Finance and Accounting*, Vol. 5, No. 2, pp. 15-23.

Thornhill, S. and Amit, R. (2003), "Learning about failure: Bankruptcy, firm age and the resource-based view", *Organization Science*, Vol.14, No. 5, pp. 497-509.

Titman S. and Wessels, R. (1988), "The determinants of capital structure choice", *Journal of Finance*, Vol. 43, No. 1, pp. 1-19.

Tomczak, J. M. and Zieba, M. (2015), "Classification restricted Boltzmann machine for comprehensible credit scoring model", *Expert Systems with Applications*, Vol. 42, No. 4, pp. 1789-1796.

Vlachos, D. Tolias, Y. A. (2003), "Neuro-fuzzy modeling in bankruptcy prediction", *Yugoslav Journal of Operations Research,* Vol.13, No.2, pp 165-174.

Volk, M. (2012), Estimating probability of default and comparing it to credit rating classification by banks, *Economic and Business Review*, Vol. 14, No. 4, pp. 299-320.

West, R.R. (1970), "An alternative approach to predicting corporate bond ratings", *Journal of Accounting Research*, Vol. 8, No. 1, pp. 118-125.

Whitrow, C., Hand, D.J., Juszczak, P., Weston, D. and Adams, N.M. (2009), "Transaction aggregation as a strategy for credit card fraud detection", *Data Mining and Knowledge Discovery*, Vol. 18, No. 1, pp. 30-55.

Yu, L., Shouyang, W. and Kin, K. L. (2009), "An Intelligent-Agent-Based Fuzzy Group Decision-Making Model for Financial Multicriteria Decision Support: The Case of Credit Scoring." *European Journal of Operational Research*, Vol.195, No.3, pp.942-59.

Zadeh, L. (1965). Fuzzy sets. Information and Control, Vol.8, No.3, pp.338 -353.

 © 2025 by Author/s