Financial Internet Quarterly 2026, vol. 22 / no. 3, p. 99-111
https://dx.doi.org/10.65748/fiqf-2026-0022

FINANCIAL DISTRESS PREDICTION MODELLING USING A COMBINATION OF LOGISTIC REGRESSION AND DECISION TREES

Martina Sponerova[1], Silvie Zlatosova[2], Miroslav Sponer[3]

Abstract

This article examines the prediction of financial distress among micro, small and medium-sized enterprises (MSMEs). The primary research question is whether integrating logistic regression with decision tree techniques enhances the accuracy and practical relevance of bankruptcy prediction compared to logistic regression alone. To address this question, both a standard logistic regression model and a hybrid model combining logistic regression with decision tree–based variable selection were applied to a dataset of 2,770 accounting records from 986 Czech companies, of which 238 were declared insolvent between 2010 and 2016. The results indicate that while logistic regression independently provides valuable insights, its hybrid integration improves model robustness and predictive performance. The combined approach not only enhances the accuracy of bankruptcy prediction but also identifies key financial distress indicators and their critical threshold values. These findings suggest that the hybrid method can serve as a useful analytical tool for financial analysts, policymakers, and business managers seeking reliable early-warning mechanisms for financial distress in MSMEs.

JEL classification: G32, G33
Keywords: Financial Distress, SME, Financial Indicator, Logistic Regression, Decision Trees

Received: 27.04.2026 Accepted: 22.06.2026

Cite this:
Sponerova, M., Zlatosova, S. & Sponer, M. (2026). Financial distress prediction modelling using a combination of logistic regression and decision trees. Financial Internet Quarterly 22(3), pp. 99-111. https://dx.doi.org/10.65748/fiqf-2026-0022.

© 2026 Martina Sponerova et al. published by University of Information Technology and Management in Rzeszów. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.

[1] Faculty of Economics and Administration, Department of Finance, Masaryk University, Czechia, e-mail: martina.sponerova@econ.muni.cz, ORCID: https://orcid.org/0000-0003-0457-4653.
[2] Faculty of Economics and Administration, Department of Finance, Faculty of Science, Department of Mathematics and Statistics , Masaryk University, Czechia, e-mail: silvie.zlatosova@mail.muni.cz, ORCID: https://orcid.org/0000-0002-5176-3505.
[3] Institute of Economic Studies, CTU Masaryk Institute of Advanced Studies, Czechia, e-mail: Miroslav.Sponer@cvut.cz, ORCID: https://orcid.org/0000-0001-6117-8468.

Introduction

The effective prediction of financial distress and credit risk has long been a central concern for financial institutions, risk managers, and academics alike, serving as a cornerstone for institutional survival, capital optimization, and accurate credit pricing. Since the 1990s, this issue has gained increasing prominence, driven by the evolving global regulatory landscape. Historically, framework standards like Basel II - the international regulatory accord introduced in 2004 by the Basel Committee on Banking Supervision (BCBS) to align banking capital with actual risk - shifted risk management toward internal modelling, allowing banks to compute regulatory capital using their own internal risk estimates. However, early calibrations under Basel II often penalized small and medium-sized enterprises (SMEs), prompting corrective regulatory interventions such as the European Union's SME Supporting Factor under the Capital Requirements Regulation (CRR) to alleviate capital burdens. (Basel Committee on Banking Supervision 2017) However, traditional bankruptcy prediction models - primarily developed using data from large corporations - showed relatively weak predictive power when applied to MSMEs, as these firms exhibit distinct financial characteristics. The global financial crisis of 2007–2009 further deepened MSMEs' difficulties in accessing credit. Later, the introduction of Basel III, combined with the economic consequences of the COVID-19 pandemic, created additional challenges for financial institutions in managing risk-weighted and non-performing exposures. With the transition to the finalized Basel III reforms, colloquially referred to as Basel IV and legally enacted in the European Union via the Capital Requirements Regulation III and Capital Requirements Directive VI (CRR III / CRD VI) package (applicable from January 1, 2025), the regulatory environment has fundamentally changed. A cornerstone of this framework is the introduction of the "Output Floor," which heavily restricts the capital-reducing benefits of internal models, forcing banks to balance internal risk assessments against rigid standardized calculations. Because the EU has retained and expanded capital relief multipliers for SME exposures under CRR III, precise and robust default prediction modelling becomes vital. In this tightened regulatory environment, advanced modelling serves as the primary mechanism for banks to minimize capital requirements, optimize risk-weighted assets (RWA), and maintain competitive credit pricing. (Neisen & Röth 2023)

This study investigates the predictive power of financial distress models for MSMEs by applying logistic regression and a hybrid modelling approach that integrates logistic regression with decision tree-based techniques. Importantly, decision trees are not employed as a standalone method; rather, they serve to enhance the logistic regression framework by improving variable selection and identifying potential threshold effects. The analysis draws on an exclusive dataset of 2,770 accounting records representing 986 Czech companies, of which 238 were identified as insolvent between 2010 and 2016. The main contribution of this research lies in its methodological focus: instead of simply comparing separate prediction methods, it proposes and tests a combined approach that proves to be a more effective modelling strategy for MSMEs bankruptcy prediction.

A rich body of literature addresses financial distress prediction using a wide range of methods. While numerous statistical and machine learning techniques have been applied, no single approach has achieved universal dominance. Traditional models such as discriminant analysis and logistic regression are often contrasted with artificial intelligence (AI)-based methods, including decision trees, neural networks, and support vector machines, typically with the aim of demonstrating the superior predictive accuracy of AI models (Alaka et al. 2017; Ling & Wang 2024; Samara & Shinde 2025). In contrast, our objective is not to establish the supremacy of one method over another, but to develop a systematic, hybrid modelling procedure that enhances both interpretability and predictive performance.

Based on the findings, two key recommendations can be drawn. First, models for predicting financial distress should be segmented according to firm-specific characteristics or prevailing economic conditions and regularly updated, as the determinants of financial distress vary across periods of crisis and growth. Second, hybrid modelling that integrates decision tree–based insights into logistic regression is recommended. This combined approach improves both the estimation of bankruptcy probability and the identification of critical variable thresholds, thereby enhancing the practical value of prediction models for credit risk assessment in banks.

In sum, this study contributes to the existing literature by extending the understanding of bankruptcy prediction through the integration of traditional econometric and data-driven techniques, the use of segmented modelling approaches, and, most importantly, the identification of threshold values of key financial distress indicators. First, the theoretical framework of this study is outlined in the Literature Review, focusing on the most commonly researched topics in bankruptcy prediction models. The Methodology section provides a detailed description of the data, variable selection, and the specification of the methods employed. Subsequently, the prediction of financial distress is examined using logistic regression and a hybrid approach combining logistic regression with decision trees. The paper concludes with a discussion of the results.

Literature review

Numerous methods have been developed to assess a firm's financial position, with the most widely cited bankruptcy prediction models including those of Altman, Taffler, Ohlson, and Zmijewski. With advances in information technology, artificially intelligent expert systems (AIES) and other artificial intelligence (AI)-based models have increasingly gained prominence.

Comparisons between traditional statistical models and more recent machine learning and artificial intelligence (AI) methods represent a major theme in the literature on bankruptcy prediction. Traditional statistical models, such as Altman's Z-score (1968), Ohlson's O -score (1980), Zmijewski's Probit (1984), and Taffler's model (1982), are parametric approaches that rely on strict assumptions of multivariate normality and homoscedasticity. While highly interpretable and auditable, they struggle to capture complex non-linear feature interactions.

Conversely, non-parametric Machine Learning techniques - including Support Vector Machines, Artificial Neural Networks, and Decision Trees (DTs) - automatically detect non-linear interactions and threshold effects without requiring rigid distributional assumptions. They employ advanced techniques to address severe class imbalance, which is typical of real-world bankruptcy datasets where insolvent firms represent only a small minority (Dasilas & Rigani, 2024).

Traditional bankruptcy prediction models have predominantly been developed using data from large enterprises. Consequently, their predictive accuracy tends to decline when applied to other segments of firms, such as micro, small and medium-sized enterprises (MSMEs). Previous research (e.g., Altman et al. 2020; Gupta et al. 2018a, b; Wang & Guedes 2025) demonstrates that models tailored to a specific segment of enterprises offer greater validity in predicting financial distress.

Many authors contrast predictive accuracy across models to demonstrate methodological superiority, with traditional approaches—such as discriminant analysis and logistic regression—frequently pitted against AI-based techniques. The prevailing argument, supported by numerous studies (Alaka et al. 2018; Dasilas & Rigani 2024; Svabova et al. 2020; Gregorova et al. 2020, Ling & Wang, 2024; Samara & Shinde, 2025), is that AI models outperform traditional methods. Nonetheless, AI techniques have also faced criticism, particularly with regard to interpretability and robustness (Tran et al. 2022; Yeo et al. 2023; Alangari et al. 2023; Chander et al. 2025).

Another important dimension of the literature concerns the segmentation of bankruptcy prediction models by industry, size, or enterprise type. Industries differ in terms of seasonality, balance sheet structures, and cost characteristics, making industry-specific approaches particularly valuable. Research has confirmed that models tailored to individual sectors achieve higher predictive power (Nugroho & Dewayanto, 2025; Mehmood & De Luca, 2025; Ďuričová et al., 2025; Michalkova & Ponisciakova, 2025; Fedorova et al., 2016; Karas & Reznakova, 2017). Similarly, segment-specific models - especially those developed for SMEs, or even further for MSMEs - tend to provide more accurate results (Altman et al., 2020; Wang & Guedes, 2025; Cheralghali & Molnár, 2025; Croquet, 2024; Gupta et al., 2018a, b).

Recent research has shifted attention from merely comparing traditional statistical and machine learning techniques toward the development of hybrid and ensemble models that combine the strengths of multiple methodologies. Several studies have shown that ensemble learning approaches and hybrid frameworks often achieve higher predictive performance than individual models while maintaining greater robustness across different datasets (Zhao et al. 2024; Dasilas & Rigani 2024; El Madou et al. 2024). At the same time, increasing emphasis has been placed on model interpretability and explainability, leading to the growing adoption of explainable artificial intelligence (XAI) techniques in bankruptcy prediction (El Madou et al. 2024; Rojas-Paredes et al. 2025). Another emerging research direction involves the integration of non-financial and alternative data sources, including textual information, corporate governance indicators, network relationships among firms, and event-based information, which have been shown to improve predictive accuracy beyond that achievable using financial ratios alone (Zhao et al. 2024; Dasilas & Rigani 2024). These developments suggest that future bankruptcy prediction systems will increasingly rely on combining traditional financial indicators with advanced machine learning methods and alternative sources of information.

Overall, the reviewed research suggests that no universally superior bankruptcy prediction method exists. Instead, the choice of methodology should be guided by the nature of the data, the context of the analysis, and the specific goals of the study.

Methodology

Data

This study is based on anonymized data obtained from a commercial bank. The dataset comprised 2,770 financial statements of MSMEs covering the period 2010-2016, representing 986 companies, of which 238 statements were classified as 'in insolvency'. The insolvent accounts correspond to the year preceding the initiation of insolvency proceedings.

To establish greater terminological precision, small and medium-sized enterprises are defined in accordance with the European Commission Recommendation 2003/361/EC. The classification is summarised in Table 1.

Table 1: European Commission SME Classification Tresholds

Enterprise Category	Headcount (FTE)	Annual Turnover	Balance Sheet Total
Medium-sized	< 250	<= EUR 50 million	<= EUR 43 million
Small	< 50	<= EUR 10 million	<= EUR 10 million
Micro	< 10	<= EUR 2 million	<= EUR 2 million

Source: Author's own work.

Firm-size thresholds capture meaningful heterogeneity in the structural and behavioural determinants of corporate cash-flow dynamics. Micro-enterprises are generally characterized by lower financing needs and greater operational adaptability, but they frequently maintain more limited liquidity buffers, which can increase their sensitivity to short-term cash-flow shocks. By comparison, small and medium-sized enterprises tend to exhibit higher fixed-cost intensity, more complex financing structures, and greater resource commitments. As a result, the nature and magnitude of cash-flow risk may differ systematically across size categories. Failure to account for this heterogeneity within a single prediction framework may result in model misspecification and reduced predictive accuracy, thereby providing a rationale for the development of size-specific prediction models.

To structure the dataset for bankruptcy prediction modelling, we considered two key segmentation criteria. The first segmentation criterion is firm size, as firm-size categories capture important differences in the structural and behavioural determinants of cash-flow risk. Given this heterogeneity, prediction models estimated on pooled samples of enterprises may fail to adequately capture size-specific patterns, resulting in reduced predictive performance. The second criterion is the period of economic development, reflecting the findings and recommendations of Altman et al. (2021), who emphasized the importance of updating predictive models in line with changing macroeconomic conditions.

The table below presents the classification of companies by firm size.

Table 2: Classification of enterprises by size

Segment	Healthy	Bankrupt	% Bankruptcy	Total
Micro-enterprises	908	73	7.44%	981
Small enterprises	841	77	8.39%	918
Middle enterprises	783	88	10.10%	871
Total	2,532	238	8.59%	2,770

Source: Author's own work.

Another criterion selected for analysis is the modelling based on the period of economic development, as the dataset spans from the financial crisis to the economic upturn beginning in 2014.

Table 3: Time structure of the data

Year	Healthy	Bankrupt	% Bankruptcy	Total
2010	402	53	11.65%	455
2011	392	45	10.30%	437
2012	400	46	10.31%	446
2013	377	34	8.27%	411
2014	366	30	7.58%	396
2015	362	27	6.94%	389
2016	233	3	1.27%	236
Total	2,532	238	8.59%	2,770

Source: Author's own work.

The crisis period is defined as the years 2010 to 2013, reflecting the aftermath of the 2008 global financial crisis and the ensuing European sovereign debt crisis that affected many economies during this time. The economic growth period covers the years 2014 to 2016, characterized by recovery and expansion in most sectors across the region.

Variables

The selection of explanatory variables was conducted through a theoretical filtration based on a systematic review of 59 academic publications from 2006 to 2020, as well as the comprehensive overview of commonly used variables by Bellovary, Giacomino, and Akers (2007). Ratios that appeared most frequently in the literature focused on Visegrad (V4) MSMEs were prioritized, ensuring region-specific relevance and reducing initial noise. These variables primarily consist of financial ratios derived from financial statements, including measures of liquidity, profitability, leverage, solvency, and efficiency.

Table 4: Key financial indicators

	Indicator	Title
Profitability		
EAT/A	EAT/assets	Return on assets
RETE/A	retained earnings/assets	
EAT/REV	EAT/sales	Return on sales
EAT/C	EAT/equity	Return on equity
EBT/ST. DEBT	EBT/current liabilities	
Liquidity		
CURR.A/ST DEBT	current assets/current liabilities	Current liquidity
ST DEBT/CURR.A	current liabilities/current assets	
NWC/A	working capital/assets	
Activity		
REV/A	sales/assets	Turnover of assets
Indebtedness		
DEBT/A	liabilities/assets	Indebtedness
A/DEBT	assets/liabilities	
C/A	equity/assets	Self-financing coefficient
DEBT/C	liabilities/equity	Debt to equity ratio
C/DEBT	equity/liabilities	
ST DEBT/A	current liabilities/assets	Short-term indebtedness
EBIT/FIN.EXP.	EBIT/interest expense	Interest cover
Other indicators		
CURR.A/DEBT	current assets/liabilities	
C/LT A	equity/fixed assets	
EAT VALUE	1 if net profit is negative, 0 otherwise	
RATIO DEBTtoA	1 if external resources > assets; 0 if external resources < assets	

Source: Author's own work.

Model specifications

As previously noted, one of the employed methods was logistic regression, which allows both the identification of significant predictors of financial distress and the estimation of its probability. Logistic regression is widely used for modelling binary outcomes, particularly in bankruptcy prediction, owing to its interpretability, robustness, and straightforward estimation. Unlike linear regression, which assumes a continuous dependent variable, logistic regression models the probability of an event (e.g., financial distress) based on explanatory variables.

Formally, logistic regression assumes that the log-odds of the dependent variable can be expressed as a linear combination of the predictors:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} \qquad (1)$$

where p_i denotes the probability that firm experiences financial distress, X_{ij} are the explanatory variables (e.g., financial ratios), and are the parameters to be estimated. The model therefore links the explanatory variables to the probability of distress through the logistic (sigmoid) function:

$$p_i = \frac{1}{1+e^{-(\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+\ldots+\beta_k X_{ki})}} \quad (2)$$

One of the key advantages of logistic regression lies in the interpretability of its coefficients. The estimated parameters indicate how a one-unit change in a variable affects the log-odds of financial distress, holding other variables constant. This feature enables both the interpretation and comparison of predictor importance (Hosmer et al., 2013).

Logistic regression is particularly suitable for bankruptcy prediction because it does not require normality or homoscedasticity of predictors and can handle both continuous and categorical variables. However, its assumption of linearity in the logit may overlook nonlinear relationships among financial indicators. To address this limitation, a hybrid approach combining logistic regression with decision tree techniques was applied. Decision trees were not used as a standalone method but rather to enhance the regression model by identifying significant variables and threshold values. They are robust to outliers, missing data, and multicollinearity, and in this study, regression trees were constructed using the CHAID (Chi-square Automatic Interaction Detection) algorithm. The CHAID algorithm constructs a tree by repeatedly splitting the sample into increasingly homogeneous subgroups. At each node, the predictor exhibiting the strongest statistically significant association with the target variable is selected based on the chi-square statistic.

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \quad (3)$$

where O_{ij} and E_{ij} denote the observed and expected frequencies, respectively. The resulting split defines threshold values that maximize differences in the incidence of financial distress between the created groups.

In the hybrid model, significant variables were first identified through decision tree analysis and subsequently incorporated into the logistic regression. To mitigate the influence of outliers, variables were categorised according to the identified threshold values. This combined methodology preserves the interpretability of logistic regression while capturing nonlinear effects and improving predictive accuracy. (Quinlan 1986)

The empirical analysis estimates the probability of financial distress for the overall MSME segment and across firm classifications. Modelling was conducted in IBM SPSS, and model accuracy was evaluated using the Receiver Operating Characteristic (ROC) curve to ensure comparability of results across approaches.

Results

Logistic regression

The results of the financial distress prediction modelling using logistic regression are presented in the Table 5. To achieve accurate classification of firms into healthy and bankrupt categories, it was necessary to lower the cut-off threshold for each segment. This adjustment aimed to maximize the identification of bankrupt firms, even if it resulted in a higher rate of misclassification among healthy firms in certain cases.

Furthermore, variables exhibiting high correlations were isolated to avoid multicollinearity. Based on the correlation coefficients, variables with correlation values of 0.7 or higher were removed. The final set of variables included in the logistic regression is as follows.

Table 5: Final set of variables included in the logistic regression model

Code	Financial indicators
EAT/A	Net profit/asset
RETE/A	Retained earnings/assets
NWC/A	Net working capital/asset
REV/A	Sales/assets
DEBT/A	Foreign resources/asset
DEBT/C	Foreign resources/equity
EBIT/FIN.EXP.	EBIT/interest expense
C/LTA	Equity/non-current assets
EAT VALUE	1 if net profit is negative, 0 otherwise
RATIO DEBTtoA	1 if external resources > assets; 0 if external resources < assets

Source: Author's own work.

Variables with a significance level of 5% or lower are considered statistically significant. In the area of financial ratios, most variables commonly applied in financial analysis have proven to be significant. Across all segments, the most prominent is the total indebtedness ratio (DEBT/A), whose coefficient values increase substantially when indebtedness changes by one unit. This finding indicates that as a firm's indebtedness rises, its probability of bankruptcy increases significantly. Another important variable across all segments is the

ratio of equity to fixed assets (C/LTA), which aligns with the pari-mutuel financing principle stating that fixed assets should be financed by equity. A decline in fixed assets is therefore associated with an increased probability of bankruptcy. Net working capital to assets (NWC/A) also emerges as a key variable. This indicator, previously used in the models of Altman and Ohlson and frequently cited in the literature, shows that a decrease in the net working capital ratio corresponds to a higher probability of bankruptcy.

Table 6: Results of the logistic regression model

Variable	MSP	Micro	Small	Middle	Crisis	Upturn
Cut off	0.090	0.074	0.099	0.103	0.110	0.065
AUC	0.683	0.706	0.797	0.715	0.683	0.775
Constant	0.040***	0.039***	0.013***	0.026***	0.075***	0.006***
EAT/A	0.329	0.337	0.191	2.225	0.403	0.197
RETE/A	0.415*	0.269	0.384	0.775	0.487	0.431
NWC/A	0.441**	0.435	0.086***	2.906	0.215***	3.383
REV/A	0.980	1.020	1.051	0.667***	0.969	1.005
DEBT/A	6.850***	4.582	44.012***	25.103***	3.806***	36.765***
DEBT/C	0.983	0.996	0.887*	1.000	0.978	0.996
EBIT/FIN.EXP.	1.000	0.997**	1.000	1.000	1.000	0.999
C/LTA	1.009***	1.005	1.004	1.049***	1.009***	1.009
EAT VALUE	1.480*	1.410	1.515	1.739	1.126	3.455***
RATIO DEBTtoA	0.295**	0.660	0.000	0.570	0.415	0.075*

Note: *** , ** , * denotes 1%, 5% and 10% significance level

Source: Author's own work.

The findings further suggest that it is appropriate to estimate the probability of financial distress separately for each segment. This is evidenced by the results of the ROC curve and the AUC values, which in most cases demonstrate improved predictive accuracy when segmentation is applied. Each segment possesses unique characteristics that are reflected in the precision of bankruptcy probability estimation.

Integration of Logistic Regression and Decision Tree Methods

Financial distress prediction was further modeled using a combination of logistic regression and decision tree methods. Initially, the decision tree method was applied because logistic regression is highly sensitive to multicollinearity. First, we used a decision tree to identify significant variables. Since these variables may be affected by outliers, the decision tree was also utilized to discretize them into specific thresholds with the highest predictive power. These variables were subsequently incorporated into the logistic regression as categorical inputs. This combined modeling strategy enables the identification of both significant predictors and specific risk thresholds. Consistent with the previous model, predictive accuracy was evaluated using the ROC curve to ensure comparability across methods.

The significant variables were categorized separately for each segment as well as for the entire dataset. The results of the categorization and the thresholds of each variable are presented below.

Categorization of variables for the whole SME segment

Using the decision tree method, the variables C/DEBT, EBT/ST DEBT, REV/A, CURR.A/ST DEBT, RETE/A, EBIT/FIN. EXP., EAT/C, and EAT VALUE were identified as significant. These variables were categorized according to the decision tree method, and the categories with their corresponding values are presented in Table 6.

Table 7: Groups of categorical variables for the whole SME segment

	Category				
	1	2	3	4	5
C/DEBT	≤ 0.16	0.17-0.55	0.56-2.24	> 2.25	
EBT/ST DEBT	≤ 0.02	0.03-0.05	0.06-0.08	0.09-0.2	> 0.3
CURR.A/ST DEBT	≤ 1.19	1.20-2.06	> 2.07		
RETE/A	≤ 0.00	0.01-0.07	0.08-0.24	> 0.25	
EBIT/FIN. EXP.	≤ -0.03	-0.04-0.00	0.01-5.75	> 5.76	
EAT/C	≤ -0.02	> -0.03			

Source: Author's own work.

The REV/A variable was not grouped by the decision tree method, even after lowering the significance level to 0.90. Therefore, it was included in the logistic regression as a continuous variable. The EAT VALUE variable is already categorical and was not further grouped.

Categorization of variables for micro-enterprises

The decision tree method identified DEBT/C, C/LTA, and EBIT/FIN.EXP as significant variables for the micro-enterprise segment. These financial variables were subsequently categorized, and the results are presented in Table 7.

Table 8: Prediction of financial distress with categorical variables for micro-enterprises

	Category			
	1	2	3	4
DEBT/C	≤ 0.73	> 0.73		
C/LTA	≤ 0.92	0.93-4.56	> 4.56	
EBIT/FIN.EXP	≤ -0.32	-0.33-0.00	0.01-8.52	> 8.53

Source: Author's own work.

Categorization of variables for small enterprises

The significant variables for the small business segment selected using the decision tree method are A/DEBT and REV/A. These variables were categorized as shown in the table below. However, asset turnover (REV/A) was not categorized by the decision tree method even after lowering the significance threshold to 0.90; therefore, it was included in the logistic regression as a continuous variable.

None of the selected variables are correlated with each other, and therefore, they all entered the logistic regression.

Table 9: Groups of categorical variables for the small enterprise segment

	Category			
	1	2	3	4
A/DEBT	≤ 1.17	1.18-1.74	> 1.74	

Source: Author's own work.

Categorization of variables for medium -sized enterprises

For the medium-sized enterprise segment, the variables ST DEBT/A, C/LTA, ST DEBT/CURR A, REV/A, DEBT/A, and EBT/ST.DEBT were identified as significant. These financial variables were categorized, except for C/LTA and REV/A, which were not further divided and will therefore be included in the logistic regression as continuous variables.

Table 10: Groups of categorical variables for the medium-sized enterprises segment

	Category			
	1	2	3	4
ST.DEBT/A	≤ 0.24	0.25-0.47	0.48-0.57	> 0.58
ST DEBT/CURR.A	≤ 0.84	0.85-1.06	> 1.07	
DEBT/A	≤ 0.33	0.34-0.48	0.49-0.56	> 0.57
EBT/ST.DEBT	≤ 0.02	> 0.03		

Source: Author's own work.

Categorization of Financial Variables for SMEs During the Financial Crisis

The variables included in the modeling of bankruptcy probability for the financial crisis period are C/A, EAT/REV, REV/A, and RETE/A.

The asset turnover variable (REV/A) has not been further expanded and will therefore enter the logistic regression as a classical variable. None of the selected variables are correlated with each other, and therefore, all of them entered the logistic regression.

Table 11: Groups of categorical variables for SMEs during the financial crisis

	Category			
	1	2	3	4
C/A	≤ 0.18	0.19-0.27	0.28-0.39	> 0.40
EAT/REV	≥ 0.04	> 0.05		
RETE/A	≤ 0.00	0.00-0.07	0.08-0.23	> 0.24

Source: Author's own work.

Categorization of Financial Variables for SMEs During Economic Growth Periods

For the period of economic growth, the significant variables EBIT/FIN.EXP, C/DEBT, NWC/A, EAT/A, C/LTA, and CURR.A/ST DEBT were identified using the decision tree method.

The variables NWC/A, C/LTA were not further branched, so they will enter the logistic regression as classical variables.

Table 12: Groups of categorical variables for SMEs in periods of economic growth

	Category			
	1	2	3	4
EBIT/FIN.EXP	≤ -0.0500	-0.060-8.12	> 8.13	
C/DEBT	≤ 0.3500	0.360-0.87	0.88-2.60	> 2.61
EAT/A	≤ -0.0009	0.001-0.08	> 0.09	
CURR.A/ST DEBT	≤ 1.6400	> 1.65		

Source: Author's own work.

All categorized variables from the individual segments were incorporated into logistic regression. The table below presents the variables that proved significant within each segment. Where a variable is indicated only by an asterisk (*) without an accompanying value, it achieved statistical significance both as a standalone variable and across categorized subsets. Therefore, discrete numerical values for individual segments are omitted in such cases. As noted, the asterisk significance levels are explained in the table footnotes

Consistent with findings from both logistic regression and decision tree analyses, the set of significant variables varies across segments, reflecting the unique characteristics of each segment that must be considered.

Table 13: Significant Variables from the Integration of Logistic Regression and Decision Tree Approaches

Variable	SME	Micro	Small	Middle	Crisis	Upturn
AUC	0.715	0.733	0.781	0.787	0.675	0.787
Constant	0.160***	0.050***	0.018***	0.006***	0.046***	0.003***
RETE/A	***				**	
RETE/A (1)	1.368				1.541	
RETE/A (2)	2.200***				2.335***	
RETE/A (3)	1.602**				1.681**	
EBT/ST DEBT	***					
EBT/ST DEBT (1)	2.414***			1.888**		
EBT/ST DEBT (2)	0.932					
EBT/ST DEBT (3)	1.844**					
EBT/ST DEBT (4)	1.269					
ST DEBT/CURR.A				***		
ST DEBT/CURR.A (1)				8.972***		
ST DEBT/CURR.A (2)				11.679***		
REV/A				0.613***		
DEBT/A				***		
DEBT/A (2)				1.976		
DEBT/A (3)				0.360		
DEBT/A (4)				2.964**		
A/DEBT			***			
A/DEBT (1)			17.895***			

Variable	SME	Micro	Small	Middle	Crisis	Upturn
A/DEBT (2)			8.295***			
C/A					***	
C/A (1)					2.717***	
C/A (2)					1.031	
C/A (3)					2.535***	
DEBT/C						
DEBT/C (2)		2.829***				
C/DEBT	***					***
C/DEBT (1)	4.760***					23.391***
C/DEBT (2)	3.800***					14.822***
C/DEBT (3)	2.320**					6.844*
ST DEBT/A				***		
ST DEBT/A (2)				2.108**		
ST DEBT/A (3)				4.274***		
ST DEBT/A (4)				1.531		
EBIT/FIN.EXP		**				***
EBIT/FIN.EXP (1)		3.476***				8.111***
EBIT/FIN.EXP (2)		1.023				2.241**
EBIT/FIN.EXP (3)		2.186**				
C/LTA		**		1.051***		1.012*
C/LTA (1)		0.641				
C/LTA (2)		0.397***				

Note: *** , ** , * denotes 1%, 5% and 10% significance level

Source: Author's own work.

Discussion

The combined approach of logistic regression and decision trees addresses a key limitation of standalone decision tree models, namely their inability to directly estimate the probability of bankruptcy. By integrating both techniques, the hybrid model not only identifies significant financial distress indicators and their critical threshold values but also provides estimates of bankruptcy probability. As a result, it offers a more comprehensive assessment of financial risk and demonstrates superior predictive performance compared with logistic regression alone.

This improvement is reflected in the Area Under the Curve (AUC) values presented in Table 14. Across all examined firm categories and macroeconomic periods, the hybrid LR+DT model achieved equal or higher predictive accuracy than the logistic regression model.

Table 14: Comparison of AUC Values across Financial Distress Prediction Methods

Variable	SME	Micro	Small	Middle	Crisis	Upturn
AUC LR	0.683	0.673	0.775	0.694	0.670	0.752
AUC LR+DT	0.715	0.733	0.781	0.787	0.675	0.787

Source: Author's own work.

The enhanced predictive ability of the hybrid model also provides valuable insights into the mechanisms of corporate financial distress during the period 2010–2016, which encompassed the aftermath of the global financial crisis, the Eurozone debt crisis, and the subsequent phase of economic recovery supported by accommodative monetary policy. The significant variables and their threshold values identified by the model reflect the structural changes that occurred throughout this period.

One of the most important determining factors was the C/LTA ratio (equity to long-term assets), which highlighted the importance of adhering to the golden rule of financing, particularly for microenterprises with limited access to external financing. Firms with a stronger capital base for financing long-term assets proved to be more resilient to adverse economic shocks. Conversely, a decline in the RETE/A ratio (retained earnings to total assets) below approximately 7% indicated the depletion of internal financial reserves and a sharp increase in the risk of bankruptcy, particularly during times of crisis. Negative interest coverage ratio (EBIT/financial expenses) is a reliable bankruptcy signal for microenterprises, even amidst economic growth. This negative value identifies underperforming entities that conform to the definition of so-called

'zombie firms', as they fail to generate sufficient operating profits to sustain their financing costs during supportive market phases.

Another critical finding concerns financial leverage, measured by the C/DEBT ratio (capital to debt). When this ratio fell below 0.55, corresponding to debt levels exceeding approximately 1.8 times equity, the probability of failure increased nonlinearly. These threshold effects illustrate why the hybrid model outperformed traditional linear methods: while logistic regression identifies statistically significant predictors, decision trees reveal the tipping points at which financial health deteriorates rapidly. Consequently, the hybrid framework is particularly effective in periods of economic transition, when linear relationships may no longer adequately describe corporate behavior.

The practical relevance of these findings extends beyond the study period and is applicable to the current macroeconomic environment characterized by elevated inflation, higher interest rates, and geopolitical uncertainty. Indicators such as EBIT/FIN.EXP and C/DEBT have become even more critical as increasing debt-servicing costs accelerate financial distress among highly leveraged firms. Likewise, insufficient levels of RETE/A limit a company's ability to absorb unexpected cost shocks and maintain operational stability. The results therefore suggest that hybrid modelling approaches capable of identifying nonlinear threshold effects provide a more robust framework for financial distress prediction than traditional linear models. As the boundary between solvency and insolvency becomes increasingly narrow under volatile economic conditions, the combination of logistic regression and decision trees offers a valuable tool for creditors, investors, regulators, and policymakers seeking to monitor and manage corporate risk effectively.

Conclusions

This paper investigates the predictive performance of logistic regression and a hybrid modelling approach that integrates logistic regression with decision tree techniques in forecasting financial distress among small and medium-sized enterprises (MSMEs). In contrast to previous research that primarily compares the performance of individual methods, this paper evaluated whether their integration can enhance predictive accuracy and overall model quality.

Drawing on the existing literature, two modelling frameworks were employed: a standard logistic regression model and a hybrid model incorporating decision tree–based variable selection and categorization. The probability of bankruptcy was first estimated for the overall MSME segment and subsequently analysed across firm-size categories (micro, small, and medium) as well as across two distinct economic periods, the financial crisis of 2010–2013 and the subsequent recovery phase of 2014–2016.

The results confirm that the determinants of corporate financial distress differ across firm-size categories and macroeconomic conditions; however, several robust patterns emerged. The most influential predictors were indicators reflecting capital structure, financial leverage, profitability, and the accumulation of internal financial resources. In particular, the equity-to-long-term-assets ratio (C/LTA), retained earnings to total assets (RETE/A), debt-related indicators (C/DEBT), and interest coverage (EBIT/FIN.EXP) repeatedly demonstrated strong explanatory power. These variables jointly capture a firm's financial stability, its dependence on external financing, and its ability to absorb adverse economic shocks.

The analysis further revealed the existence of important nonlinear threshold effects. Firms whose retained earnings fell below approximately 7% of total assets or whose capital-to-debt ratio declined below 0.55 exhibited a substantially higher probability of failure. Similarly, negative interest coverage ratios were strongly associated with severe financial distress and insolvency risk. These findings indicate that financial deterioration does not occur gradually but often accelerates once specific critical thresholds are crossed.

The comparison of modelling approaches demonstrated that the hybrid framework combining logistic regression with decision trees achieved consistently higher predictive accuracy than logistic regression alone across nearly all firm categories and economic periods. While logistic regression provides interpretable estimates of bankruptcy probability, decision trees complement this approach by identifying critical breakpoints and capturing nonlinear relationships that traditional linear models may overlook. The improvement in AUC values confirms that integrating both methods enhances model robustness and predictive performance.

From a practical perspective, the findings suggest that bankruptcy prediction models should be tailored to specific firm segments and regularly updated to reflect changing macroeconomic conditions. The study also demonstrates that hybrid modelling approaches are particularly valuable in periods of economic uncertainty, as they better capture the complex dynamics underlying corporate financial distress. Given the current environment of elevated inflation, higher interest rates, and increased market volatility, the ability to identify nonlinear warning signals becomes even more important. Consequently, the proposed hybrid model represents a useful tool for creditors, investors, financial institutions, and policymakers seeking to assess and manage corporate default risk more effectively.

References

Alaka, H.A., Oyedele, L.O., Owolabi, H.A., Kumar, V., Ajayi, S.O., Akinade, O.O. & Bilal, M. (2017). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. Expert Systems with Applications, 94, 164-184. https://dx.doi.org/10.1016/j.eswa.2017.10.040.

Alangari, N., El Bachir Menai, M., Mathkour, H. & Almosallam, I. (2023). Exploring evaluation methods for interpretable machine learning: A survey. Information, 14(8), 469-500. https://dx.doi.org/10.3390/info14080469.

Altman, E.I. & Sabato, G. (2007). Modelling credit risk for SMEs: Evidence from the US market. Abacus, 43(3), 332-357. https://dx.doi.org/10.1111/j.1467-6281.2007.00234.x.

Altman, E.I., Esentato, M. & Sabato, G. (2020). Assessing the credit worthiness of Italian SMEs and mini-bond issuers. Global Finance Journal, 43, 100-150. https://dx.doi.org/10.1016/j.gf.2018.09.003.

Basel Committee on Banking Supervision. (2017). Basel III: Finalising post-crisis reforms. Bank for International Settlements, Basel.

Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A review of bankruptcy prediction studies: 1930 to present. Journal of Financial Education, 33, 1-42.

Chander, B., John, C., Warrier, L. & Gopalakrishnan, K. (2025). Toward trustworthy artificial intelligence (TAI) in the context of explainability and robustness. ACM Computing Surveys, 57(6), 1-49. https://dx.doi.org/10.1145/3675392.

Cheraghali, H. & Molnár, P. (2025). Practical insights into predicting defaults in small and medium-sized enterprises. Journal of the International Council for Small Business, 6(4), 685-696. https://dx.doi.org/10.1080/26437015.2024.2430573.

Ciampi, F., Giannozzi, A., Marzi, G. & Altman, E.I. (2021). Rethinking SME default prediction: A systematic literature review and future perspectives. Scientometrics, 126, 1-48. https://dx.doi.org/10.1007/s11192-020-03856-0.

Croquet, M., Cultrera, L., Laroutis, D., Pozniak, L. & Vermeylen, G. (2024). Exploring the Role of Global Value Chain Position in Economic Models for Bankruptcy Forecasting. Econometrics, 12(4), 31-42. https://dx.doi.org/10.3390/econometrics12040031.

Dasilas, A. & Rigani, A. (2024). Machine learning techniques in bankruptcy prediction: A systematic literature review. Expert Systems with Applications, 255, 124-161. https://dx.doi.org/10.1016/j.eswa.2024.124761.

Duričová, L., Kovalová, E., Gazdíková, J. & Hamranová, M. (2025). Refining the best-performing V4 financial distress prediction models: Coefficient re-estimation for crisis periods. Applied Sciences, 15(6), 29-56. https://dx.doi.org/10.3390/app15062956.

El Madou, K., Marso, S., El Kharrim, M. & El Merouani, M. (2024). Evolutions in machine learning technology for financial distress prediction: A comprehensive review and comparative analysis. Expert Systems, 41(2), 134-185. https://dx.doi.org/10.1111/exsy.13485.

Fedorova, E.A., Dovzhenko, S.E. & Fedorov, F.Y. (2016). Bankruptcy-prediction models for Russian enterprises: Specific sector-related characteristics. Studies on Russian Economic Development, 27(3), 254-261. https://dx.doi.org/10.1134/S1075700716030060.

Gregova, E., Valaskova, K., Adamko, P., Tumpach, M. & Jaros, J. (2020). Predicting financial distress of slovak enterprises: Comparison of selected traditional and learning algorithms methods. Sustainability, 12(10), 39-54. https://dx.doi.org/10.3390/su12103954.

Gupta, J., Barzotto, M. & Khorasgani, A. (2018a). Does size matter in predicting SMEs failure? International Journal of Finance & Economics, 23(4), 571-605. https://dx.doi.org/10.1002/ijfe.1638.

Gupta, J., Gregoriou, A. & Ebrahimi, T. (2018b). Empirical comparison of hazard models in predicting SMEs failure. Quantitative Finance, 18(3), 437-466. https://dx.doi.org/10.1080/14697688.2017.13075 14.

Hosmer Jr, D.W., Lemeshow, S. & Sturdivant, R.X. (2013). Applied logistic regression. John Wiley & Sons, New York.

Karas, M. & Reznakova, M. (2017). Predicting the bankruptcy of construction companies: a CART-based model. Engineering Economics, 28(2), 145-154. https://dx.doi.org/10.5755/j01.ee.28.2.16353.

Ling, Y. & Wang, P.P. (2024). Ensemble machine learning models in financial distress prediction: Evidence from China. Journal of Mathematical Finance, 14(2), 226-242. https://dx.doi.org/10.4236/jmf.2024.142013.

Mehmood, A. & De Luca, F. (2025). Financial distress prediction in private firms: Developing a model for troubled debt restructuring. Journal of Applied Accounting Research, 26(6), 205–222. https://dx.doi.org/10.1108/JAAR-12-2022-0325.

Michalkova, L. & Ponisciakova, O. (2025). Bankruptcy prediction, financial distress and corporate life cycle: Case study of Central European enterprises. Administrative Sciences, 15(2), 63-80. https://dx.doi.org/10.3390/admsci15020063.

Neisen, M. & Röth, S. (2023). CRR III: The EU Implementation of Basel IV-the Next Generation of Risk Weighted Assets. John Wiley & Sons, New York.

Nugroho, D.S. & Dewayanto, T. (2025). Application of statistics and artificial intelligence for corporate financial distress prediction models: A systematic literature review. Journal of Modelling in Management, 20(6), 1999–2023. https://dx.doi.org/10.1108/JM2-12-2024-0412.

Ohlson, J.A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. Journal of Accounting Research, 18(1), 109–131. https://dx.doi.org/10.2307/2490395.

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

Samara, K. & Shinde, A. (2025). Bankruptcy Prediction Using Machine Learning and Data Preprocessing Techniques. Analytics, 4(3), 22-35. https://dx.doi.org/10.3390/analytics4030022.

Svabova, L., Michalkova, L., Durica, M. & Nica, E. (2020). Business failure prediction for Slovak small and medium-sized companies. Sustainability, 12(11), 45-72. https://dx.doi.org/10.3390/su12114572.

Taffler, R.J. (1982). Forecasting Company Failure in the UK Using Discriminant Analysis and Financial Ratio Data. Journal of the Royal Statistical Society. Series A (General), 145(3), 342–358. https://dx.doi.org/10.2307/2981867.

Tran, K. L., Le, H.A., Nguyen, T.H. & Nguyen, D.T. (2022). Explainable machine learning for financial distress prediction: Evidence from Vietnam. Data, 7(11), 160-180. https://dx.doi.org/10.3390/data7110160.

Vásquez-Serpa, L.J., Rodríguez, C., Pérez-Núñez, J.R. & Navarro, C. (2025). Challenges of Artificial Intelligence for the Prevention and Identification of Bankruptcy Risk in Financial Institutions: A Systematic Review. Journal of Risk and Financial Management, 18(1), 26-48. https://dx.doi.org/10.3390/jrfm18010026.

Wang, W. & Guedes, M.J. (2025). Firm failure prediction for small and medium-sized enterprises and new ventures. Review of Managerial Science, 19(7), 1949-1982. https://dx.doi.org/10.1007/s11846-024-00742-4.

Yeo, W. J., Van Der Heever, W., Mao, R., Cambria, E., Satapathy, R. & Mengaldo, G. (2025). A comprehensive review on financial explainable AI: WJ Yeo et al. Artificial Intelligence Review, 58(6), 189-198. https://dx.doi.org/10.1007/s10462-024-11077-7.

Zhao, J., Ouenniche, J. & De Smedt, J. (2024). Survey, classification and critical analysis of the literature on corporate bankruptcy and financial distress prediction. Machine Learning with Applications, 15, 100-127. https://dx.doi.org/10.1016/j.mlwa.2024.100527.

Zmijewski, M.E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. Journal of Accounting Research, 22, 59–82. https://dx.doi.org/10.2307/2490859.