

Article

Corporate Loan Default Prediction in the Slovak Banking Context: An Interpretable and Ensemble CRISP-DM Pipeline for Credit Risk Assessment

Lucia Duricova *  and Veronika Labosova 

Department of Economics, Faculty of Operation and Economics of Transport and Communications, University of Zilina, Univerzitna 1, 01026 Zilina, Slovakia; veronika.labosova@stud.uniza.sk

* Correspondence: lucia.duricova@uniza.sk

Abstract

In bank-dominated financial systems, the accumulation of non-performing loans is a recognised source of systemic vulnerability, as correlated corporate defaults can erode bank capital, impair liquidity, and propagate stress across interconnected portfolios. Firm-level default detection thus constitutes a microprudential foundation of macroprudential stability: the reliable early identification of risky borrowers reduces both individual credit losses and the aggregate exposures that drive system-level fragility. Yet the use of structured data-mining pipelines for this task remains underexplored in Central and Eastern Europe. This study applies the CRISP-DM methodology to predict corporate loan default using data on 302 Slovak corporate borrowers, combining financial ratios from publicly available financial statements with selected company and loan-related information from internal bank records. Seven individual classifiers were developed and compared: decision trees (CART, CHAID, C5.0), logistic regression, discriminant analysis, and neural networks (MLP, RBF), together with a stacked ensemble based on their outputs. Model performance was evaluated using sensitivity, overall classification accuracy, and area under the ROC curve (AUC), with sensitivity treated as the primary criterion because of the asymmetric costs of misclassification in credit risk assessment. The results confirm that historical firm-level information provides a reliable basis for default prediction, with tree-based models consistently outperforming statistical and neural network approaches. The stacked ensemble achieved the strongest overall performance, whereas C5.0 and CHAID showed that interpretable classifiers can also deliver competitive predictive accuracy. A champion–challenger deployment architecture is proposed, in which the ensemble serves as the performance-oriented champion and interpretable models act as challengers; this arrangement contributes to the operational resilience of the credit-risk assessment process and aligns with macroprudential expectations of model governance, auditability, and explainability. The study offers a replicable methodological framework for integrating data-driven decision support into credit evaluation in comparable banking settings.



Academic Editor: Steve Y. Yang

Received: 14 May 2026

Revised: 16 June 2026

Accepted: 23 June 2026

Published: 25 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: financial stability; systemic risk; corporate default prediction; credit risk; CRISP-DM; ensemble learning; interpretable machine learning; macroprudential governance

1. Introduction

Credit risk remains one of the most critical risk categories faced by financial institutions, with direct implications for loan portfolio stability, capital allocation, and banking sector

performance [1,2]. A high proportion of non-performing loans raises provisioning, reduces profitability, and, in extreme cases, threatens institutional solvency [3,4]. Beyond the individual institution, correlated corporate defaults can transmit shocks to the wider financial system by eroding aggregate banking-sector capital and amplifying losses through diversification breakdown [5–7]. The reliable identification of risky borrowers at the firm level, therefore, represents a microprudential foundation of macroprudential stability, particularly in the bank-dominated financial systems of Central and Eastern Europe, making effective credit risk management both an institution-level and a system-level concern [8]. Traditional approaches based on expert judgement or simple scoring models often fail to capture the nonlinear relationships between financial indicators and borrower behaviour, whereas data mining and machine learning classifiers can transform historical financial data into actionable credit risk signals and frequently outperform conventional statistical methods, especially within structured analytical frameworks that ensure reproducibility and transparency [9–12].

Despite this progress, two important gaps remain in the existing literature. First, the majority of published studies rely either on publicly available financial data or on proprietary internal bank scoring systems, rarely combining both within a single analytical framework. The integration of financial ratios derived from publicly available financial statements with loan-specific information, such as loan amount and repayment history, available only from internal bank records, represents an underexplored but practically relevant approach, as it reflects the full informational basis that a lending institution actually possesses at the point of credit evaluation. Second, while data-mining methods have been widely applied to credit risk prediction in Western and Asian contexts, evidence from Central and Eastern European (CEE) banking markets remains limited, and no prior study has applied the full CRISP-DM guideline to bank-specific corporate loan default prediction in the Slovak context.

This study addresses both gaps. Using a dataset of 302 Slovak corporate borrowers obtained from a commercial bank, we apply the CRISP-DM methodology to develop and benchmark seven individual classification models: decision trees (CART, CHAID, C5.0), logistic regression (LR), discriminant analysis (DA), and two neural network (NN) architectures (MLP, RBF), alongside a stacked ensemble meta-learner combining all individual model outputs. The main goal of this study is to develop and evaluate a structured and replicable framework for predicting corporate loan default in the Slovak banking context using historical firm-level financial data. The study pursues three main objectives: (i) to examine whether historical firm-level financial information can reliably predict future corporate loan default; (ii) to identify the modelling approach that provides the strongest predictive performance for practical use in credit risk assessment; and (iii) to position firm-level default classification within a broader framework of banking-sector resilience and macroprudential model governance, in which predictive accuracy, interpretability, and operational resilience jointly support the stability of bank-dominated financial systems. The emphasis is on practical applicability—the proposed guideline is designed as a replicable decision-support framework that banks and creditors can integrate into credit evaluation and risk management workflows.

Based on the above, the following research question and hypotheses were formulated:

RQ: Can historical financial ratios serve as reliable predictors of corporate loan default, and which classification approach provides the strongest predictive performance for practical credit risk assessment?

H1. *Financial ratios derived from firms' historical financial statements are capable of reflecting future loan default risk with a sensitivity of at least 70% on the held-out test sample.*

H2. *Individual classification models employing firm-level financial indicators as predictors achieve a sensitivity of at least 70% and an AUC of at least 0.8 on the held-out test sample, indicating sufficient predictive performance for practical use in credit risk assessment.*

H3. *A stacked ensemble model achieves higher predictive performance (sensitivity, overall accuracy and AUC) compared to individual classifiers and is therefore the most suitable approach for corporate loan default prediction.*

The remainder of the article is organised as follows. Section Literature Review reviews existing research on credit default prediction and the application of data-mining methods. Section 2 describes the CRISP-DM framework, the data, and the modelling procedure. Section 3 presents the results of the comparative model evaluation. Section 4 discusses the findings in the context of the existing literature, addresses the study's practical implications and model governance considerations, and outlines the limitations and directions for future research. Section 5 summarises the key findings and methodological contributions of the study.

Literature Review

Default risk is a key concern for financial institutions, as it poses a threat to the stability of their loan portfolios and the efficiency of capital allocation [13,14] (Naili & Lahrichi, 2022; Bank for International Settlements, 2015). In addition to affecting individual credit decisions, default risk also has systemic implications for banks' overall exposure and financial stability: when multiple corporate clients default simultaneously due to sectoral or geographic linkages, banks may face elevated funding costs and increased vulnerability to liquidity and solvency shocks [5,6]. As Javadi and Osah [7] show, the correlation of credit risk across borrowers in a loan portfolio can amplify losses during economic downturns and erode the effectiveness of diversification strategies. With the increasing complexity of the lending environment and the growing availability of firm-level data, traditional models based on expert scoring systems are progressively being replaced by machine learning methods that promise higher predictive accuracy and better handling of non-linear relationships [15,16].

A complementary strand of research explicitly links firm-level credit risk to system-level financial stability through the structural interconnectedness of borrowers, lenders, and exposures. Default events are rarely independent: shared sectoral exposure, common counterparties, and direct contractual linkages cause defaults to cluster in time and to propagate through banking-system balance sheets, transforming localised credit losses into broader financial stability concerns [17–19]. This perspective implies that firm-level default-prediction models contribute not only to institution-level decision support but also, in aggregate, to the macroprudential surveillance of credit-driven systemic vulnerabilities—a connection that the relational and network-based methods reviewed below, notably graph neural networks [20–22], make computationally tractable. This link between firm-level and system-level risk is the theoretical basis for treating default prediction as a microprudential tool that supports broader financial stability.

Classical statistical approaches—LR and DA—remain widely used in credit scoring due to their simplicity and interpretability [23]. However, their ability to accurately identify borrowers at risk of default declines when applied to complex and multidimensional datasets, particularly under volatile or uncertain financial conditions. To address this shortfall, more recent studies have turned to advanced techniques. Kheneifar and Amiri [24] combine topological data analysis with AdaBoost to capture subtle patterns associated with deteriorating creditworthiness, underscoring the importance of integrating both behavioural and structural financial indicators. Similarly, Guo et al. [25] demonstrate that

analysing temporal behavioural data enables models to detect early changes in borrower behaviour that precede missed payments, improving the timely identification of emerging default risk. A significant recent advancement is the application of graph neural networks (GNNs): Zandi et al. [20], Chen et al. [21], and Ye et al. [22] independently show that GNNs are particularly effective at capturing relational dependencies between clients, improving predictive accuracy for mortgage, corporate, and online loan defaults alike.

Alongside technical performance, a parallel strand of research emphasises the importance of preserving the economic and regulatory meaning of default in model design. Millo et al. [26] criticise the rise in algorithmic self-reference, in which developers prioritise optimising classification metrics over interpreting what constitutes a true default from a financial standpoint. In response, Li and Wu [27] and Monje et al. [28] argue that interpretability is essential for trustworthy default risk assessment, achieved through methods such as SHAP, fuzzy logic, or surrogate trees. Sotiropoulos et al. [29] propose symbolic regression as a mathematically transparent alternative to black-box systems, while Kheneifar and Amiri [24] use LIME to make individual default classifications actionable for credit officers. These efforts reflect a growing consensus that default risk models must balance predictive performance with interpretability to satisfy governance and regulatory requirements [30–32].

The importance of sound data preparation and feature selection is equally well established. Antoniadis et al. [33] identify the strengths of modern classifiers, including random forests and neural networks, when combined with oversampling techniques, and Akinjole et al. [34] confirm the superior robustness of ensemble approaches over stand-alone classifiers. Both studies emphasise, however, that preprocessing and relevant feature selection are at least as important as the choice of algorithm in reducing default misclassification. Research interest has also grown in segment-specific default analysis: Chai et al. [35] find that engineered features outperform standard indicators when assessing default risk among farmers, while Lu et al. [36] and Berrada et al. [37] demonstrate that high classification performance can be achieved using structured financial data alone, without behavioural variables. This supports the practical relevance of models built solely on historical accounting indicators, particularly in contexts where behavioural data is unavailable.

Within the CRISP-DM framework specifically, Berrada et al. [37] applied it to corporate lending and found that random forests achieved the best classification results for defaulting developers after applying SMOTE to address class imbalance. Chang et al. [38] extended this approach by integrating survival analysis with LR to account for the timing of default events—a consideration directly relevant to Basel III capital requirements.

Prior research in Slovakia has mainly focused on predicting corporate financial distress, rather than bank loan defaults. Nevertheless, it has produced a number of valuable contributions, such as Svabova et al. [39], Papik and Papikova [40], Gajdosikova et al. [41], Gavurova et al. [42], Gregova et al. [43], Stefko et al. [44], and Labosova et al. [45]. However, these studies share a common focus on corporate bankruptcy or financial distress as the outcome of interest, which is conceptually distinct from bank loan default, the focus of the current study. Moreover, the earliest work directly targeting Slovak corporate loan defaults, a probit-based study by Fidrmuc and Hainz [46] of nearly 700 SME loans covering the period 2000–2005, relied on classical statistical methods and data that predate the current regulatory and market environment by two decades. To our knowledge, no prior Slovak study has adopted CRISP-DM as a structured analytical framework for predicting loan repayment outcomes using bank-sourced credit data.

Despite the growing body of international research on machine-learning-based corporate default prediction, studies focusing on CEE markets remain scarce, and Slovakia

in particular has received minimal attention in this literature. This gap is consequential: credit market structures, accounting standards, and the regulatory environment in CEE economies differ meaningfully from those in Western Europe or the United States, where most benchmark studies are conducted, and models developed in those contexts may not transfer directly without recalibration. The present study addresses this gap by developing and evaluating a CRISP-DM-based classification framework using real financial data from Slovak enterprises, with the dual objective of identifying the most informative financial predictors of loan default and producing a replicable workflow that can be adapted to similar markets in the region. While the empirical data are Slovak, the contribution extends beyond a single national context: the CRISP-DM pipeline applied here is domain- and country-independent, and Slovakia shares key structural characteristics: EU membership, IFRS-aligned reporting standards, a bank-dominated financial system, and comparable SME credit market dynamics, with several CEE economies including the Czech Republic, Hungary, Poland, and the Baltic states. Findings established here therefore provide a relevant reference point for researchers and practitioners across the region, and replication in other CEE markets is a natural and tractable extension of this work.

2. Methodology and Data

This study employs the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology as its overarching analytical framework. CRISP-DM is a widely recognised, domain-independent framework for conducting data-mining projects, comprising six interconnected phases: business understanding, data understanding, data preparation, modelling, evaluation, and deployment [47–49]. Its domain independence makes it particularly well-suited to financial prediction tasks, where it provides a systematic foundation for translating business objectives into analytical tasks and ensuring the reproducibility and transparency of results [50]. In this study, the framework serves as a structured guide for developing a binary classification model that predicts whether a company will default on its loan obligations based on historical financial ratios. The structure of the CRISP-DM phases is shown in Figure 1.

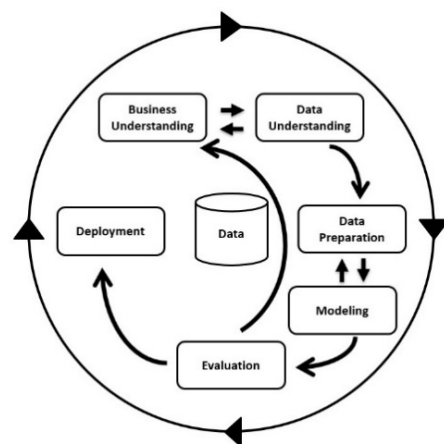


Figure 1. The CRISP-DM process model, illustrating the iterative phases of a data-mining project. The arrows between phases indicate the iterative movement of the process, and the outer circle represents the cyclical nature of data mining. Source: Chapman et al. [48].

The following subsections are organised according to the phases of this framework, with each subsection indicating its corresponding CRISP-DM phase.

2.1. Analytical Framework (CRISP-DM: Business Understanding)

Corresponding to the business understanding phase of CRISP-DM, this section establishes the business motivation and analytical objectives of the study. The central problem addressed is the growing share of non-performing loans, which increases pressure on banks' financial stability and capital adequacy. From a practical standpoint, the ability to identify high-risk companies early is essential for mitigating credit risk and improving the quality of loan approval decisions. The business success criterion is accordingly defined as a reduction in the share of non-performing loans through more effective assessment of applicants' creditworthiness.

The analytical task is formulated as a binary classification problem with the aim of estimating the probability that a company will default on its loan obligations in the following year. This probability corresponds conceptually to the Probability of Default parameter used in the Basel internal ratings-based framework. Loss Given Default, which quantifies the loss severity conditional on default, lies outside the scope of this study, as it requires recovery and collateral data not available in the dataset. For model training, historical data are used, and its predictive ability is evaluated on an independent test sample. Key analytical questions were defined in this planning stage: which financial indicators most determine the risk of default, and to what extent the classification can identify creditworthy companies from risky ones. Besides that, the following techniques were pre-selected for model creation: decision trees, artificial NNs, LR, and DA. A minimum sensitivity of 70% on the test sample was set as the operational success threshold, reflecting the asymmetric cost of misclassifying a defaulting company as creditworthy.

2.2. Data Used in the Study (CRISP-DM: Data Understanding)

Corresponding to the data understanding phase of CRISP-DM, this section describes the data sources, sample composition, and target variable for this study. The dataset comprises financial and non-financial data for 302 enterprises covering the fiscal years 2021 and 2022. Financial ratios were calculated from publicly available financial statements sourced from the Slovak Business Register and the Financial Statements Register, which are accessible to any interested party, including lending institutions. Loan-specific information, including the loan amount granted and the binary target variable Y indicating loan repayment status observed in 2023, was obtained from the cooperating bank's internal records. A value of $Y = 0$ indicates timely repayment; $Y = 1$ indicates default or late repayment. The value $Y = 1$ represents the "hit" category, as identifying defaulting clients is the primary objective: approving a loan to a company that subsequently defaults (false negative) is considerably more costly than declining a loan to a company that would have repaid it (false positive).

The resulting sample contains 267 companies in the repaying class and 35 in the default class, yielding a class imbalance ratio of approximately 8:1. As part of the initial data audit, candidate financial indicators were reviewed across four dimensions of financial performance: activity, liquidity, leverage, and profitability, to assess their availability, completeness, and economic plausibility. This assessment established a reliable basis for the subsequent variable selection and preparation decisions described in the following section.

2.3. Variable Selection and Preparation (CRISP-DM: Data Preparation)

Corresponding to the data preparation phase of CRISP-DM, this section describes the stepwise construction of the analytical dataset. Financial ratios were selected from the candidate indicators identified in the data understanding phase based on two criteria: their theoretical relevance to a company's ability to meet its debt obligations, and their

practical availability in historical financial statements. The selected indicators, calculated separately for 2021 and 2022, cover the four dimensions identified in the data survey: activity (asset turnover period—total, current, and non-current assets; inventory turnover period; accounts receivable turnover period), liquidity (liquidity ratios 1, 2, and 3), leverage (total debt ratio, long-term debt ratio, interest coverage ratio), and profitability (gross return on assets, gross return on equity, net return on sales).

Following variable construction, a data quality check was performed to ensure internal consistency. Values that were economically implausible or inconsistent with accounting logic were removed from the dataset. Outliers were carefully assessed and retained when deemed informationally relevant, particularly for companies exhibiting signs of financial distress. The original data were structured in long format with two rows per company (2021 and 2022). However, for modelling purposes, the data were restructured into wide format, yielding a unique row per company with year-specific indicators, resulting in a final analytical dataset of 302 companies and 29 predictor variables: 13 ratios from 2021, 13 from 2022, and the variable “amount of loan granted”. Descriptive statistics for all indicators (means and standard deviations in parentheses), calculated on the cleaned and restructured dataset, are reported in Table 1.

Table 1. Descriptive statistics of financial indicators used in the study.

Indicator	Year	
	2021	2022
Asset turnover period	17,595.78 (293,287.98)	9661.13 (151,011.85)
Turnover period of noncurrent assets	17,182.66 (290,574.43)	8998.51 (149,706.56)
Turnover period of current assets	410.15 (2748.90)	660.09 (5914.24)
Inventory turnover period	63.39 (134.49)	130.97 (1245.57)
Accounts receivable turnover period	115.80 (245.85)	381.13 (4430.96)
Liquidity 1	0.88 (3.64)	1.56 (8.67)
Liquidity 2	2.33 (7.28)	2.89 (9.79)
Liquidity 3	3.16 (7.89)	3.75 (10.16)
Total debt ratio	62.09 (27.82)	61.84 (29.89)
Longterm debt ratio	14.78 (22.18)	14.71 (23.48)
Interest coverage ratio	440.88 (3276.23)	480.59 (6680.96)
Gross return on assets	6.77 (11.32)	6.87 (11.94)
Gross return on equity	15.87 (119.68)	−2.81 (423.28)
Net return on sales	3.17 (36.97)	−9.43 (181.73)

To improve model stability and address the data's distributional properties, an inverse hyperbolic sine (IHS) transformation was applied to all scale variables, which exhibited high skewness and extreme values. The IHS transformation is defined as:

$$y = \operatorname{asinh} x = \ln \left(x + \sqrt{x^2 + 1} \right) \quad (1)$$

First introduced by Johnson [51] and formally developed for econometric use by Burbidge et al. [52], this transformation was preferred over alternatives such as logarithmic or Box–Cox transformations because several financial ratios in the dataset take zero or negative values—conditions under which log-based transformations are undefined or inapplicable [52,53]. The IHS transformation is applicable to the full range of real values while behaving similarly to a logarithmic transformation for large positive values, effectively compressing the right tail of heavily skewed distributions [54]. The interpretation of coefficients associated with IHS-transformed variables follows a hybrid logic between linear and logarithmic models [53]. For moderate to large values of the original variable x , the estimated coefficient b approximates semi-elasticity: a one per cent increase in x is associated with an approximate change of $b/100$ units in the outcome. For very small values of x , where the transformation behaves approximately linearly, the coefficient is interpreted in the conventional linear sense as a change of b units in the outcome as a result of a one-unit increase in x . The precise marginal effect at any given value of x can be recovered from the derivative of the transformation, which declines toward zero as x grows large. This dual interpretive property makes IHS-transformed coefficients particularly tractable in applied financial modelling, where predictors frequently span a wide range of magnitudes [52,54].

When developed prediction models report split thresholds or other results on the transformed scale, the original-scale values can be recovered by applying the inverse IHS transformation. Since (1) is the forward transformation, the corresponding back-transformation is:

$$x = \sinh y = \frac{e^y - e^{-y}}{2} \quad (2)$$

Throughout the Section 3, split thresholds and other IHS-scale values are reported alongside their back-transformed equivalents to facilitate direct interpretation in terms of the original financial ratios.

In addition to the financial ratio variables, the dataset includes six non-financial variables characterising each borrower: industry sector, ownership type, legal form, company size, and company age. The distribution of these variables is summarised in Table 2. In terms of legal form, the sample is dominated by limited liability companies (nearly 85%), consistent with the structure of the Slovak SME sector. The majority of companies are domestically privately owned (over 80%), and small enterprises represent the largest size category in both observation years (approximately 52%). Company age is a continuous variable measured in years at the end of the respective fiscal year, with a mean (in 2021) of approximately 17.8 years and a right-skewed distribution (skewness = 1.27), indicating a small number of long-established firms. The industry variable comprises 27 categories, several of which contain very few observations; the largest sectors represented are construction (almost 18%), wholesale (over 17%), agriculture and forestry (over 11%), and real estate (nearly 8%).

Table 2. Distribution of additional variables included in the analysis.

Variable	Original Category			Merged Category		
	Category Name	<i>n</i>	%	Category Name	<i>n</i>	%
Industry	Law, Consulting & Accounting	1	0.3	Group 1	2	0.6
	R&D & Testing	1	0.3			
	Automotive industry	3	1.0	Group 2	133	44.0
	Chemistry & Plastics	3	1.0			
	Finance	1	0.3			
	Information Technology	2	0.7			
	Retail	16	5.3			
	Clothing & Footwear	2	0.7			
	Food Industry	2	0.7			
	Agriculture & Forestry	34	11.3			
	Vehicle Sales & Maintenance	8	2.6			
	Engineering & Design	1	0.3			
	Waste Management	2	0.6			
	Wholesale	52	17.2			
	Other Manufacturing	1	0.3			
	Healthcare	6	2.0			
	Tourism & Hospitality	5	1.7			
	Transport & Logistics	11	3.6			
	Wood & Paper	10	3.3			
	Electrical Engineering	6	2.0			
	Energy & Mining	15	5.0	Group 3	167	55.4
	Metalworking & Metallurgy	18	6.0			
	Real Estate	23	7.6			
	Services	6	2.0			
	Intermediation Services	6	2.0			
	Construction	54	17.9			
Ownership type	Mechanical Engineering	13	4.3	retained in original form		
	Domestic private	242	80.1			
	Other	23	7.6			
	International private	21	7.0			
Legal form	Foreign	16	5.3	retained in original form		
	Limited liability company (s.r.o.)	256	84.8			
	Joint-stock company (a.s.)	28	9.27			
	Other	18	5.96			
Company size (2021)	Micro enterprise	54	17.9	SMEs	271	89.8
	Small enterprise	157	52.0			
	Medium enterprise	60	19.9			
	Large enterprise	10	3.3	Other	31	10.2
	Not available (NA)	21	6.9			

Table 2. Cont.

Original Category				Merged Category		
Variable	Category Name	<i>n</i>	%	Category Name	<i>n</i>	%
Company size (2022)	Small enterprise	153	50.7	retained in original form		
	Medium enterprise	61	20.2			
	Micro enterprise	50	16.5			
	Not available (NA)	28	9.3			
	Large enterprise	10	3.3			
Company age (2021)	Min = 1; Max = 68; Mean = 17.76; SD = 9.30; skewness = 1.27					
Company age (2022)	Min = 2; Max = 69; Mean = 18.76; SD = 9.30; skewness = 1.27					
Loan amount	Min = 40,000; Max = 3,323,518; Mean = 405,959.73; SD = 455,629.19; skewness = 2.87					

Prior to modelling, all categorical variables were subjected to supervised category merging. This procedure groups categories that exhibit similar distributions with respect to the target variable (loan repayment status), thereby reducing the number of dummy variables, improving model stability, and enhancing the discriminatory power of each categorical predictor. After this data preparation step, the industry and company size in 2021 variables were merged into fewer categories, whereas ownership type, company size in 2022, and legal form were found suitable for use in their original forms. The resulting merged groupings were used as inputs to all classification models. The two additional continuous variables (company age and loan amount) were retained in their original metric form; the loan amount variable was additionally subjected to the IHS transformation due to its highly skewed distribution, consistent with the transformation procedure applied to the financial ratio variables.

The original category distributions and their corresponding group assignments are presented in Table 2.

To avoid overfitting and ensure reliable evaluation of models' predictive performance, the data were randomly divided into training and test sets in a 60:40 ratio, with class proportions maintained in both partitions. To address the pronounced class imbalance, minority-class oversampling was applied exclusively to the training set to improve the model's ability to identify risky companies. That yields a balanced training sample of 304 observations (153 delayed, 151 repaying). The test set was left unmodified to reflect real-world class distribution and provide an unbiased estimate of out-of-sample performance. The resulting sample distribution is summarised in Table 3.

Table 3. Distribution of repayment classes in the analysed samples.

Sample	Full		Training (Oversampled)		Testing	
	Absolute Frequency	Relative Frequency [%]	Absolute Frequency	Relative Frequency [%]	Absolute Frequency	Relative Frequency [%]
Delayed	35	11.1	153	50.3	13	10.1
Repay	267	88.4	151	49.7	116	89.9
Total	302	-	304	100.0	129	100.0

2.4. Modelling and Evaluation (CRISP-DM: Modelling and Evaluation)

Corresponding to the modelling and evaluation phases of CRISP-DM, this section describes the classification techniques applied and the criteria used to assess their performance. Six individual classification models were developed: a multilayer perceptron (MLP) and a radial basis function network (RBF), both representing neural network approaches capable of capturing non-linear relationships; three decision tree algorithms: CHAID, C5.0 and Classification and Regression Tree (CART); binary LR; and linear DA. All individual models were estimated on the oversampled training set and evaluated on the held-out test set. In addition to the individual models, a stacked ensemble was constructed using a CART as the meta-learner, combining predictions from all other individual models as inputs. The selection of these techniques was guided by the balance between predictive performance and the interpretability and auditability required in regulated lending, which motivated the emphasis on transparent classifiers. The stacked ensemble was included to capture the predictive gains commonly associated with ensemble approaches such as random forests and gradient boosting, while keeping the base learners individually transparent. All models were created using the IBM Statistics SPSS, version 27, and IBM SPSS Modeller software, version 18.0. Model performance was assessed using three metrics:

- sensitivity—the proportion of true defaults correctly identified, i.e., recall for the minority class
- overall classification accuracy—the model's overall ability to correctly classify both repaid and delayed loans
- the area under the ROC curve (AUC)—summarises model's discriminative ability across all possible classification thresholds. The ROC curve plots sensitivity (True Positive Rate) against 1-specificity (False Positive Rate); the closer the curve lies to the upper left corner, the stronger the model.

Given the asymmetric misclassification costs in credit risk management, sensitivity is treated as the primary evaluation criterion throughout this study, as failing to identify a defaulting company poses a greater risk than incorrectly flagging a creditworthy one.

3. Results

This section presents the results of applying the CRISP-DM methodology to predict corporate loan defaults using historical financial ratios. The modelling phase (Section 3.1) comprises seven individual classification models, organised by the type of interpretive information they provide. The first group comprises decision tree models: CHAID, C5.0, and CART, which produce directly readable decision rules and splitting criteria, revealing the threshold values of financial indicators that distinguish creditworthy from high-risk borrowers. The second group includes LR and DA models, which yield explicit predictor coefficients and weights that quantify the direction and relative magnitude of each financial ratio's contribution to the classification. The third group consists of the neural network models: MLP and RBF, which, while not structurally interpretable, provide normalised importance plots that identify the financial indicators most influential in driving the classification outcome. Together, all three groups contribute complementary evidence toward the study's first objective: identifying the most informative financial predictors of loan default. For readability, the predictor importance plots for all seven individual models and the ensemble are collected in Appendixes A and B (Figures A1–A8) and are discussed by model in the corresponding parts below. Consequently, Section 3.2 presents the stacked ensemble model, which combines the outputs of the individual models to maximise predictive performance. Section 3.3 then brings all models together for a comparative evaluation based on sensitivity, overall accuracy, and AUC, addressing the study's second objective of building a high-performing classification model. Finally, Section 3.4 translates the evalua-

tion findings into practical deployment recommendations consistent with the CRISP-DM deployment phase.

3.1. Individual Classification Models (CRISP-DM: Modelling)

Decision Trees Models

Three decision tree algorithms were applied: CART, CHAID, and C5.0. All three are white-box models that produce directly readable decision rules, making their outputs immediately interpretable in terms of the original financial and categorical predictors. Each tree is also accompanied by a predictor importance plot, which quantifies each variable's overall contribution to the classification across all splits, providing a complementary summary of predictor relevance beyond the specific splits shown in the tree structure.

The CART algorithm produces binary splits and selects cut-points that maximise the reduction in node impurity at each step. The resulting tree is shown in Figure 2.

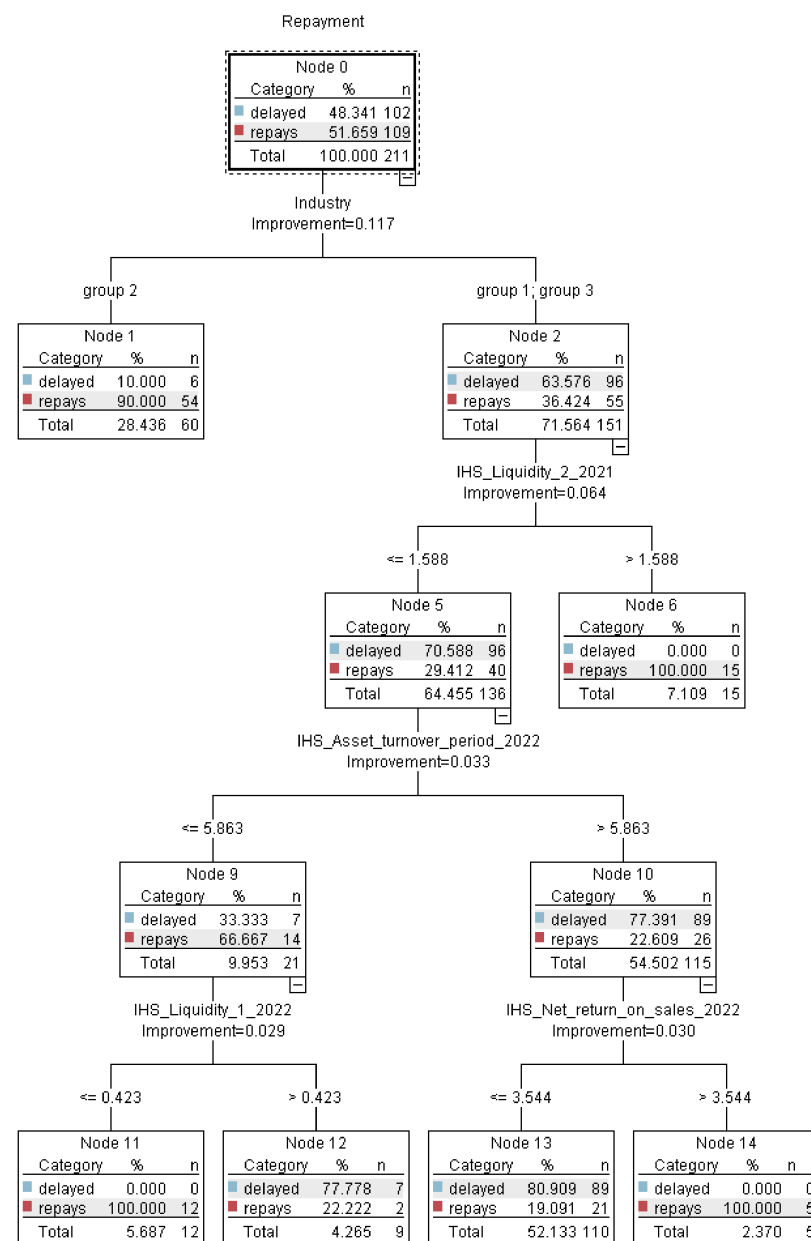


Figure 2. CART prediction model for loan repayment.

The primary split is on Industry: companies belonging to industry group 2 form a predominantly safe node (Node 1: approximately 90% repaying), while companies from groups 1 and 3 are routed to a substantially riskier branch (Node 2: nearly two-thirds delayed). Within this riskier branch, the tree splits further on Liquidity 2 in 2021 at a transformed threshold of 1.688 (original scale: approximately 2.48). Companies with the liquidity 2 ratio above this value are classified as repaying with certainty (Node 6: 100% repaying), suggesting that a quick ratio exceeding approximately 2.48 in 2021 is a reliable indicator of creditworthiness within the higher-risk industry segment. Companies below this threshold (Node 5: roughly 70% delayed) are further divided by Asset turnover period in 2022 at a transformed threshold of 5.863 (original scale: approximately 175.5 days). Companies with a shorter asset turnover period (Node 9) show a more moderate default rate of around one third, and are further split by Liquidity 1 in 2022 at a transformed threshold of 0.423 (original scale: approximately 0.43): those below this value are classified as fully safe (Node 11: 100% repaying), while those above it carry a default rate of nearly 78% (Node 12), making them among the highest-risk profiles in the tree. Companies with a longer asset turnover period above 175.5 days (Node 10: approximately 77% delayed) are further split by Net return on sales in 2022 at a transformed threshold of 3.544 (original scale: approximately 17.2%). Companies with net return on sales below this threshold form the largest high-risk terminal node in the tree (Node 13: roughly 81% delayed), while those exceeding approximately 17.2% net return on sales are classified as fully safe (Node 14: 100% repaying), confirming that adequate profitability in the year immediately preceding the assessment can offset the default risk associated with slow asset conversion.

The predictor importance plot for the CART model (Figure A2 in Appendix B) confirms the dominant predictors, with the first place belonging to the Asset turnover period in 2022 (importance of 0.15), followed by industry (0.05), Liquidity 2 in 2021 and Net return on sales in 2022 (both 0.04). The fifth place belongs to the Liquidity 1 in 2022 (0.03).

The CHAID algorithm allows multi-way partitioning at each node, producing a broader and more complex tree structure than CART. Due to its complexity, the tree is shown in the Appendix A (Figure A1). As with CART, the primary split is on Industry, separating companies into two branches with markedly different risk profiles: Industry Group 2 is predominantly safe (roughly 93% repaying), while Groups 1 and 3 combined carry a default rate of approximately two-thirds. However, both of them are further divided and do not produce the classifications at this step. Within the higher-risk branch, the tree splits on Liquidity 2 in 2021 into five intervals. These intervals range from nearly pure default (Node 4: over 93% delayed, corresponding to a quick ratio from approximately 0.27 to 0.44) to fully safe classification (Node 8: 100% repaying, corresponding to a quick ratio above approximately 2.46). The remaining intermediate nodes split further on the total debt ratio in 2022, the turnover period of non-current assets in 2021, the total debt ratio in 2021, the long-term debt ratio in 2021, the net return on sales in 2021, and liquidity 3 in 2022. Within the safer industry group 2 branch, the tree also splits on Liquidity 2 in 2021 into three intervals. A liquidity 2 transformed value of under 0.949 and of over 1.001 means 100% repayments (Node 9 and 11, original values of approximately 1.10 and 1.19, respectively), and the middle interval contains 75% of delays (Node 10, original values from approximately 1.10 to 1.19). The middle interval is particularly interesting from a credit risk perspective. Companies with a quick ratio in a narrow band between approximately 1.10 and 1.19 carry the highest default risk within the otherwise safer industry group 2. The greater depth and breadth of the CHAID tree compared to CART reflect its sensitivity to interactions among indicators across multiple threshold intervals.

The predictor importance plot for the CHAID model (Figure A3 in Appendix B) identifies Turnover period of non-current assets in 2021 as the most important predictor

(importance of 0.39), followed by Total debt ratio in 2022 (0.24) and Liquidity 2 in 2021 (0.18). Industry ranks fourth (0.13), and the fifth place belongs to the long-term debt ratio in 2021 (0.05). The dominance of the non-current asset turnover period is noteworthy, as it did not appear as a primary splitting variable in the CART model, suggesting that CHAID's multi-way splitting mechanism is more sensitive to the discriminatory power of activity indicators distributed across multiple intervals.

The C5.0 algorithm uses information gain to select splits and incorporates post-pruning to control tree complexity. It is the most complex of the three decision trees applied here, incorporating not only financial ratios but also categorical company-level variables, including industry affiliation, company age, and region. The resulting tree is shown in Figure 3.

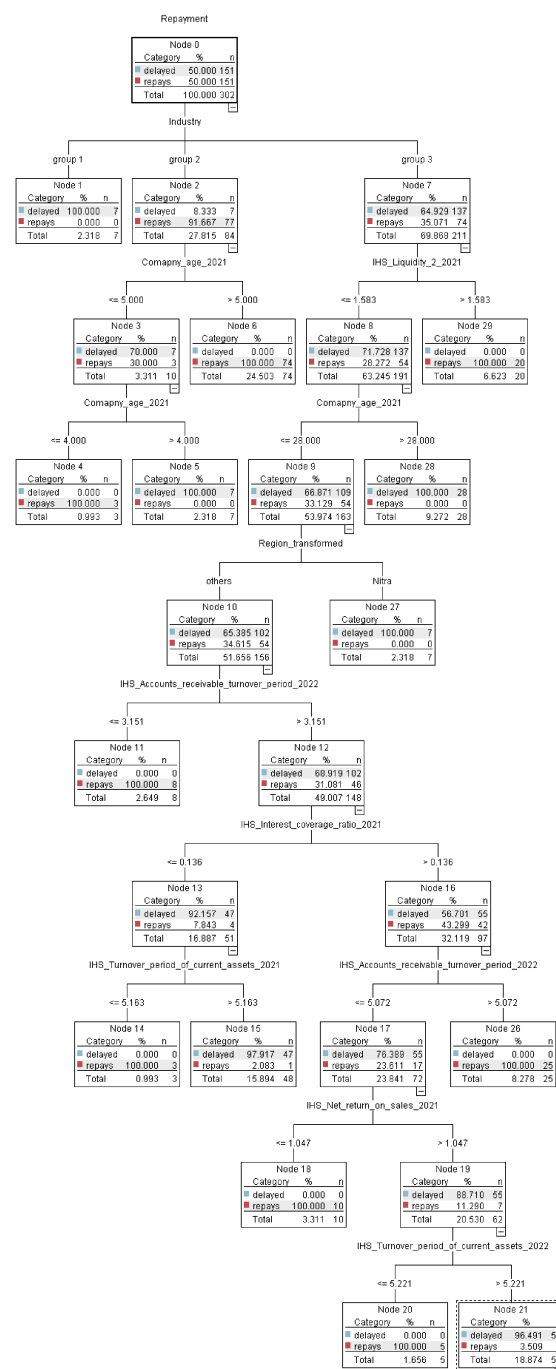


Figure 3. C5.0 prediction model for loan repayment.

As with the previous models, the primary split is on Industry. Companies in group 1 are immediately classified as defaulting with certainty (Node 1: 100% delayed), identifying this industry segment as a homogeneous high-risk group requiring no further splitting. Companies in group 2 are further split by Company age in 2021: younger firms of up to 5 years are further divided by age: those up to 4 years old carry a default rate of approximately 70%, while those older than 4 but no more than 5 years are classified as fully safe, whereas firms older than 5 years are classified as fully repaying (Node 6: 100% repaying). This finding highlights the role of firm maturity as a default risk signal, with very young enterprises exhibiting substantially elevated risk even within the relatively safer industry group 2.

Companies in group 3, the largest and most heterogeneous branch (nearly 65% delayed), are first split by Liquidity 2 in 2021 at a transformed threshold of 1.583 (Node 7, original scale: approximately 2.33). Companies exceeding this liquidity threshold are classified as fully safe (Node 29: 100% repaying), while those below it enter a deeper sequence of splits. Within this high-risk branch, the tree first splits by Company age 2021 at 28 years, with the firms older than 28 years classified as fully repaying (Node 28: 100% repaying), suggesting that long-established enterprises retain sufficient creditworthiness regardless of their short-term liquidity position. Younger firms are further divided by Region and then by Accounts receivable turnover period in 2022, Net return on sales in 2021, Turnover period of current assets in 2021, and Long-term debt ratio in 2022.

The predictor importance plot for the C5.0 model (Figure A4 in Appendix B) identifies the Interest coverage ratio in 2021 as the most important predictor (importance of 0.26), followed by Industry (0.21), Accounts receivable turnover period in 2022 and Liquidity 2 in 2021 (both 0.16), and Net return on sales in 2021 as the fifth contributor (≈ 0.07). The prominence of the interest coverage ratio is noteworthy; it does not appear as an early splitting variable in the tree but carries the highest overall importance, indicating that its discriminatory power is distributed across multiple branches rather than concentrated at a single high-level split. This contrasts with CART and CHAID, where a single variable dominated the importance ranking, suggesting that C5.0 captures a more distributed, interaction-rich representation of default risk.

Across all three decision tree models, industry affiliation consistently emerges as the primary splitting criterion, while liquidity indicators, particularly liquidity ratio 2, appear repeatedly as secondary splitting variables. Activity indicators, notably turnover periods of current and non-current assets and accounts receivable, feature prominently in the importance plots of all three models, though their role varies in depth and prominence across algorithms. The C5.0 model additionally identifies company age, company size, and region as relevant discriminators, suggesting that non-financial firm characteristics carry incremental predictive value beyond financial ratios alone.

Logistic Regression and Discriminant Analysis

Binary LR models the log-odds of the outcome as a linear function of the predictors, with coefficients directly quantifying the direction and magnitude of each predictor's contribution to the probability of loan repayment versus default. The model was estimated using the backward stepwise method. Since all continuous predictors are IHS-transformed (Formula (1)), coefficients cannot be directly interpreted on the original scale. The original-scale marginal effect of a predictor with mean $\bar{x}$ is approximated by the derivative of log-odds, and for key predictors this is reported alongside the coefficient. The resulting model is presented in Table 4.

Table 4. Parameter estimates of the LR model.

Variable	B	S.E.	Wald	Sig.	Exp(B)
Intercept	23.523	8.252	8.125	0.004	-
IHS Asset Turnover Period 2021	3.690	0.912	16.373	<0.001	40.043
IHS Asset Turnover Period 2022	−6.755	1.382	23.873	<0.001	0.001
IHS Turnover period of noncurrent assets 2022	0.608	0.327	3.455	0.063	1.837
IHS Turnover period of current assets 2021	−3.539	1.009	12.304	<0.001	0.029
IHS Turnover period of current assets 2022	5.525	1.273	18.840	<0.001	250.836
IHS Inventory Turnover Period 2022	−0.576	0.159	13.218	<0.001	0.562
IHS Total debt ratio 2021	−2.667	1.092	5.965	0.015	0.069
IHS Total debt ratio 2022	2.240	0.525	18.213	<0.001	9.398
IHS Longterm debt ratio 2021	−2.255	0.487	21.441	<0.001	0.105
IHS Longterm debt ratio 2022	2.026	0.453	20.034	<0.001	7.585
IHS Interest Coverage Ratio 2021	0.428	0.112	14.514	<0.001	1.534
IHS Gross return on assets 2021	−3.530	0.831	18.039	<0.001	0.029
IHS Gross return on assets 2022	−0.704	0.311	5.120	0.024	0.495
IHS Net return on sales 2021	3.185	0.742	18.401	<0.001	24.165
IHS Amount of loan granted	−0.579	0.253	5.253	0.022	0.560
Industry = 0	−18.815	0.000	—	—	—
Industry = 1	7.016	1.290	29.579	<0.001	1114.682
Company size 2021 = 1	−4.036	4.039	0.998	0.318	0.018
Company size 2021 = 2	−3.384	3.961	0.730	0.0393	0.034
Company size 2021 = 3	−6.331	4.044	2.451	0.117	0.002
Company size 2021 = 4	−9.136	4.075	5.027	0.025	0.000
Company size 2022 = 1	0.302	1.584	0.036	0.849	1.353
Company size 2022 = 2	0.463	1.202	0.149	0.700	1.589
Company size 2022 = 3	4.674	1.370	11.647	0.001	107.130

Note: Reference categories (Industry = 2, Company_size_2021 = 5, Company_size_2022 = 4 and 5) are omitted. Industry = 0 exhibits perfect separation; its coefficient should not be interpreted substantively.

Among activity indicators, Asset turnover period 2022 carries a strong negative coefficient ($B = -6.755$, $p < 0.001$), indicating that a longer asset turnover period in 2022 is associated with default; the marginal effect at the sample mean of approximately 9, 7 days is negligible in absolute terms (-0.001 per day), confirming that the variable's discriminatory power is concentrated at lower values consistent with the decision tree threshold. The turnover period of current assets in 2022 is positively associated with repayment ($B = 5.525$, $p < 0.001$), while the turnover period of current assets in 2021 carries the opposite sign ($B = -3.539$, $p < 0.001$), reflecting opposing temporal dynamics of current asset conversion across the two observation years. The inventory turnover period 2022 is negatively associated with repayment ($B = -0.576$, $p < 0.001$).

Among leverage indicators, a consistent pattern of opposing year effects emerges. The total debt ratio in 2021 is negatively associated with repayment ($B = -2.667$, $p = 0.015$), while the total debt ratio in 2022 carries a positive coefficient ($B = 2.240$, $p < 0.001$). Similarly, the long-term debt ratio in 2021 is negatively associated with repayment ($B = -2.255$, $p < 0.001$; marginal effect at mean of 14.78% is -0.152 per percentage point, odds ratio

of approximately 0.86), while the long-term debt ratio in 2022 is positively associated ($B = 2.026$, $p < 0.001$; marginal effect at mean of 14.71% is +0.137 per percentage point, odds ratio of approximately 1.15). The interest coverage ratio in 2021 is positively associated with repayment ($B = 0.428$, $p < 0.001$), with a marginal effect at the mean of 440.88 of approximately +0.001 per unit, indicating that its discriminatory power is concentrated at low values of interest coverage.

Among profitability indicators, the net return on sales in 2021 is the strongest predictor ($B = 3.185$, $p < 0.001$; marginal effect at mean of 3.17% is +0.958 per percentage point, odds ratio of approximately 2.61), confirming that higher net profitability in 2021 is a robust positive signal of repayment capacity. The gross return on assets in 2021 ($B = -3.530$, $p < 0.001$; marginal effect at mean of 6.77% is -0.516 per percentage point, odds ratio of approximately 0.60) and the gross return on assets in 2022 ($B = -0.704$, $p = 0.024$) are both negatively associated with repayment. The amount of loan granted is negatively associated with repayment ($B = -0.579$, $p = 0.022$), indicating that larger loan amounts are associated with higher default risk.

Regarding categorical predictors, Industry 1 is strongly associated with repayment ($B = 7.016$, $p < 0.001$), consistent with all decision tree models. Among company size categories, the size 4 in 2021 ($p = 0.025$) and the size 3 in 2022 ($p = 0.001$, $\text{Exp}(B) = 107.130$) reach statistical significance; the remaining size categories are retained but are not individually significant.

The predictor importance plot for the LR model (Figure A5 in Appendix B) confirms the net return on sales in 2021 as the most important predictor (importance of 0.19), followed by the interest coverage ratio in 2021 (0.14), total debt ratio in 2021 and industry (0.11 each), and asset turnover period in 2022 as the fifth contributor (0.10).

Linear DA constructs a canonical discriminant function as a linear combination of predictors that maximally separates the two groups of repaying and defaulting companies. The model was estimated using the stepwise method based on Wilks' lambda minimisation, retaining only those predictors that contribute unique discriminatory power beyond variables already in the function. Box's M test did not reach significance ($F = 2.010$, $p = 0.156$), confirming that the assumption of equal covariance matrices is satisfied and that the within-groups covariance matrix is the appropriate classification basis. The standardised canonical discriminant function coefficients, together with the structure matrix of pooled within-groups correlations, are presented in Table 5.

Table 5. Standardised canonical discriminant function coefficients and structure matrix correlations.

Variable	Function 1	Correlation with Function 1
IHS Asset Turnover Period 2022	0.272	0.256
IHS Accounts receivable turnover period 2021	0.233	0.261
IHS Total debt ratio 2021	0.390	0.324
IHS Total debt ratio 2022	−0.540	−0.052
IHS Longterm debt ratio 2021	0.990	0.037
IHS Longterm debt ratio 2022	−1.151	−0.199
IHS Interest Coverage Ratio 2021	−0.640	−0.349
IHS Gross return on assets 2021	1.819	−0.109
IHS Net return on sales 2021	−1.728	−0.220
IHS Net return on sales 2022	0.328	−0.111
IHS Amount of loan granted	0.252	0.311

The discriminant function clearly separates the two groups, with group centroids of 0.683 for the defaulting class and -0.643 for the repaying class, indicating that companies with positive discriminant scores tend toward default and those with negative scores toward repayment.

As shown in Table 5, the standardised coefficients indicate the relative contribution of each retained predictor to the discriminant function. The largest contributions come from gross return on assets in 2021 (1.819) and net return on sales in 2021 (-1.728), followed by the long-term debt ratio in 2022 (-1.151) and in 2021 (0.990). The opposing signs of the two long-term debt ratio coefficients mirror the temporal pattern found in the LR model: higher long-term debt in 2021 is associated with default, while higher long-term debt in 2022 is associated with repayment. The interest coverage ratio in 2021 (-0.640) and total debt ratio in 2022 (-0.540) also contribute meaningfully, alongside total debt ratio in 2021 (0.390), accounts receivable turnover period in 2021 (0.233), asset turnover period in 2022 (0.272), net return on sales in 2022 (0.328), and amount of loan granted (0.252). The same table also reports the structure matrix, which complements the standardised coefficients by showing the bivariate correlations of the predictors with the discriminant function, restricted to those retained in the resulting model. The three strongest correlates are the interest coverage ratio in 2021 (-0.349), the total debt ratio in 2021 (0.324), and the amount of loan granted (0.311).

The predictor importance plot for the DA model (Figure A6 in Appendix B) identifies the interest coverage ratio in 2021 as the most important predictor (importance of 0.18), followed closely by the long-term debt ratio in 2022 (0.17), net return on sales in 2021 and long-term debt ratio in 2021 (0.13 each), and gross return on assets in 2021 as the fifth contributor (0.08). The importance structure is notably more evenly distributed than in the LR model, where net return on sales dominated clearly, suggesting that DA captures a more balanced interplay between leverage and profitability indicators in discriminating between the two groups. The convergence of interest coverage ratio and long-term debt ratio from both observation years as the leading contributors is consistent with the structure matrix, which shows these variables among the strongest bivariate correlates of the discriminant function.

Neural Network Models

Two neural network architectures were applied: an MLP and an RBF. Both are black-box models whose internal parameters—network weights and activation functions—cannot be directly interpreted as coefficients or decision rules. Their contribution to the study's first objective of identifying informative predictors is therefore assessed exclusively through normalised predictor importance plots, which quantify each variable's overall contribution to the classification outcome across the network's hidden layer structure.

The MLP importance plot (Figure A7 in Appendix B) shows a highly concentrated importance structure, with the interest coverage ratio in 2021 as the dominant predictor (normalised importance of 0.06), followed by gross return on assets in 2021, company age in 2022, liquidity 1 in 2021 (all 0.05), and liquidity 3 in 2022 (0.04). The overall importance values are notably low across all five displayed predictors; none exceeds 0.10, suggesting that the MLP distributes predictive information across many inputs rather than concentrating it in a small subset.

The RBF importance plot (Figure A8 in Appendix B) shows an even more diffuse importance structure, with liquidity 3 in 2022, total debt ratio in 2021, and liquidity 1 in 2021 as leading predictors (all with relative importance of 0.05), followed by liquidity 2 in 2021 and liquidity 1 in 2022 (both 0.04). Again, no predictor exceeds a normalised importance of 0.10, indicating a very flat importance distribution.

Across both NN models, the overall importance values are substantially lower and more evenly distributed than in the decision tree and coefficient-based models, reflecting the greater complexity of the network architectures and their tendency to distribute predictive information across many inputs simultaneously. The convergence of the interest coverage ratio in 2021 and liquidity 1 in 2021 as contributors in both models provides partial corroboration of the patterns identified by the interpretable models, though the NN importance plots should be interpreted with caution, given the black-box nature of the underlying estimation.

Across all seven individual models, several predictors emerge consistently. Industry affiliation is the dominant first-order discriminator in all three decision tree models and a significant predictor in the LR model. Liquidity indicators, particularly from 2021, appear repeatedly across decision trees, the DA, and both NN importance plots. Long-term debt ratio from both observation years features consistently in the coefficient-based models, with a recurring pattern of opposing temporal effects between 2021 and 2022. Interest coverage ratio and net return on sales, both from 2021, rank among the leading predictors in the LR, DA, and C5.0 importance plots. Activity indicators, particularly asset turnover and current asset turnover, contribute meaningfully to both decision trees and the LR model.

3.2. Ensemble Model (CRISP-DM: Modelling)

The ensemble model was constructed using stacking, a meta-learning approach in which the predicted outputs of individual base models serve as inputs to a higher-level classifier. The meta-learner is a CART model that combines the predicted classes and scores from all individual models to produce a final classification. The inputs available to the meta-learner include the predicted classes and scores from all seven base models: MLP, RBF, CART, CHAID, C5.0, LR, and DA. The resulting meta-tree is shown in Figure 4.

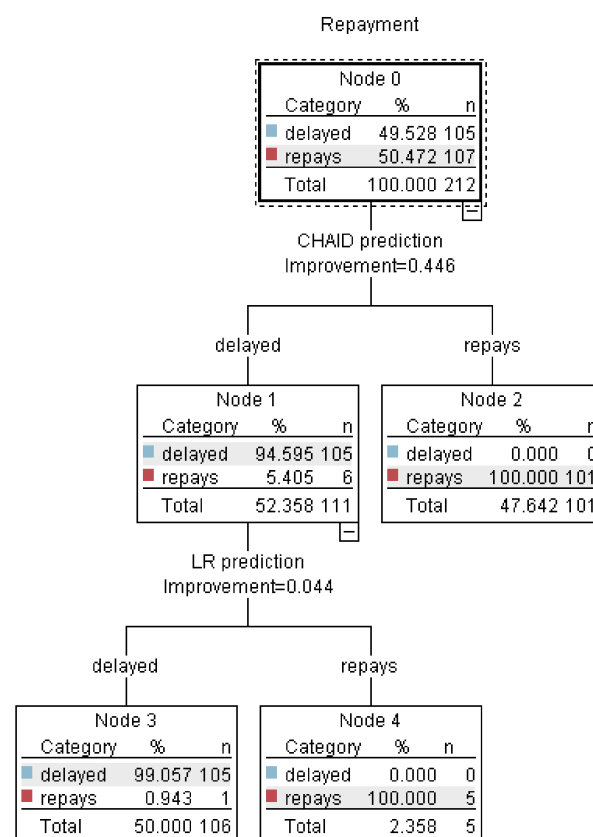


Figure 4. Ensemble prediction model for loan repayment.

The meta-tree is compact, comprising only three terminal nodes reached through two splits. The primary split is on the predicted score from the CHAID model. Companies predicted as repaying by CHAID are immediately classified as repaying with certainty by the ensemble (Node 2: 100% repaying). Companies predicted as defaulting by CHAID proceed to a second split on the predicted class from the LR model. Those predicted as defaulting by LR are classified as defaulting with near certainty (Node 3: approximately 99% delayed), while those predicted as repaying by LR are classified as repaying with certainty (Node 4: 100% repaying). The ensemble thus operates as a sequential filter: CHAID screens out the clearly creditworthy cases first, and LR resolves the remaining ambiguous cases.

The predictor importance plot (Figure A9 in Appendix B) confirms that LR prediction and CHAID prediction (both with a normalised importance of 0.29) are the two dominant contributors to the ensemble decision, together accounting for over 60% of total importance. The RBF, MLP and CHAID predicted scores contribute marginally (all 0.07).

This concentration of importance in the two interpretable white-box models—LR and CHAID—is a notable finding: despite the availability of all seven base model outputs, the meta-learner effectively reduces the ensemble decision to a combination of the two most transparent individual classifiers.

3.3. Evaluation (CRISP-DM: Evaluation)

Model performance was assessed using three indicators: sensitivity, overall classification accuracy, and AUC. While accuracy reflects the model's overall ability to correctly classify both repaid and delayed loans, this study places particular emphasis on sensitivity, as the key objective is to reliably identify companies at increased risk of repayment default. Therefore, to select the champion model, sensitivity was treated as the primary criterion, reflecting the asymmetric cost structure of credit risk assessment, in which the failure to identify a defaulting company carries substantially higher consequences than a false positive classification. The ROC curves for all models are presented in Figure 5, and a comparative summary of all performance metrics is provided in Table 6.

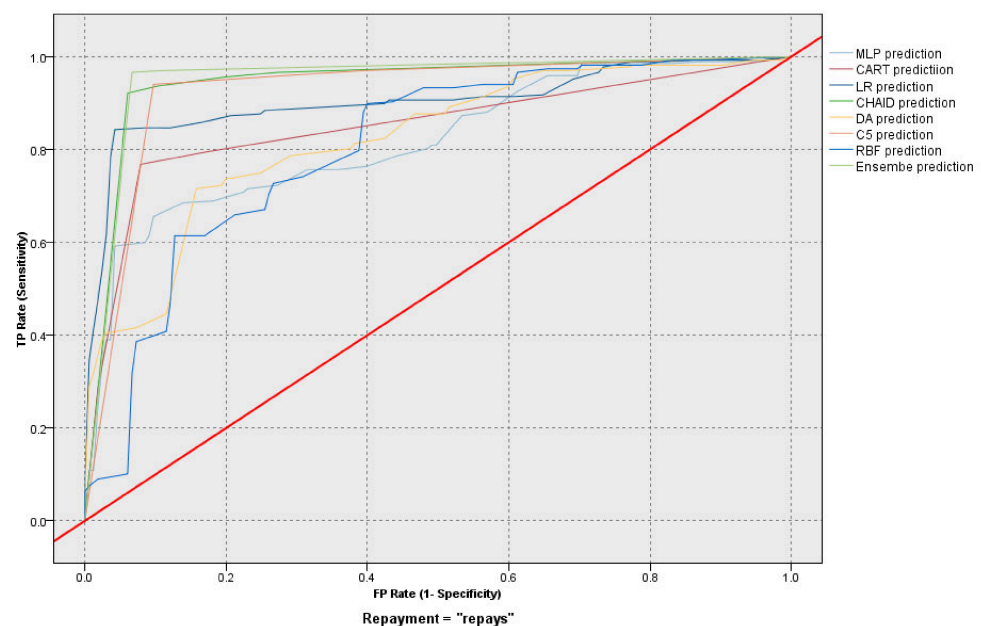


Figure 5. ROC curves for all prediction models. The red diagonal line represents the reference line of a non-discriminating classifier (AUC = 0.50); curves closer to the upper left corner indicate stronger discriminative ability.

Table 6. Comparative summary of the evaluation metrics for all models.

Modelling Technique	Positives		Negatives		Unpredicted	Sensitivity (%)	Accuracy (%)	AUC
	True Positive	False Positive	False Negative	True Negative				
CART	90	1	26	12	0	77.59	79.09	0.850
CHAID	107	1	9	12	0	92.24	92.25	0.944
C5.0	110	1	6	12	0	94.83	94.57	0.926
LR	92	0	24	12	1	79.31	81.25	0.904
DA	86	3	30	10	0	74.14	74.42	0.823
MLP NN	104	7	12	4	2	89.66	85.04	0.820
RBF NN	91	5	25	7	1	78.45	76.56	0.803
Ensemble CART model	113	1	3	12	0	97.41	96.90	0.951

The ensemble CART model achieved the strongest overall performance across all three criteria, reaching a sensitivity of over 97% and an overall accuracy of nearly 97%. Among individual models, the tree-based classifiers consistently outperformed the remaining methods. C5.0 achieved a sensitivity of nearly 95% and an accuracy of slightly below 95%. CHAID recorded both sensitivity and accuracy slightly above 92%. CART produced more moderate tree-based results, with a sensitivity of nearly 78% and an accuracy of slightly above 79%.

Among the statistical models, LR achieved a sensitivity above 79% and an accuracy above 81%, while DA produced lower values across all criteria, with both sensitivity and accuracy slightly above 74%. The NN models showed mixed results. The MLP achieved the highest sensitivity among the two NN architectures (nearly 90%), though its accuracy (85%) was lower than that of the leading tree-based models. The RBF network recorded the lowest overall performance, with a sensitivity of over 78% and an accuracy of almost 77%. All models exceeded the minimum sensitivity threshold of 70% established in the business understanding phase.

The AUC values (Table 6) calculated from the exported ROC curve coordinates using the trapezoidal rule are consistent with the ranking observed in the classification results. The ensemble model achieved the highest AUC (0.951), followed closely by CHAID (0.944) and C5.0 (0.926). LR also attained a high AUC of 0.904, while CART (0.850), DA (0.823), MLP (0.820), and RBF (0.803) recorded moderate values. All models exceeded the conventional threshold of 0.800, indicating acceptable to excellent discriminative ability across the full range of classification thresholds.

3.4. Deployment (CRISP-DM: Deployment)

The stacked ensemble model yields the strongest predictive performance in terms of sensitivity, accuracy, and AUC and is therefore designated as the champion model for deployment. Its practical use in routine credit decisions, however, is constrained by limited interpretability. The final classification is derived from the combined outputs of the base models rather than from a transparent decision-making structure. In credit risk management, where regulatory requirements and internal governance standards demand that individual lending decisions be traceable, explainable, and interpretable, interpretable models retain an important operational role. In that case, CHAID and C5.0 emerge as the most suitable interpretable alternatives, offering explicit decision rules in the form of readable trees while maintaining strong predictive performance. LR provides an additional alternative when a coefficient-based analytical framework is preferred.

These considerations support a champion–challenger deployment architecture, in which the stacked ensemble serves as the performance-oriented champion model, while CHAID, C5.0, and LR are deployed as challengers. In this arrangement, all models are applied to the same incoming data stream, and their outputs are compared continuously; a challenger may replace the champion if it demonstrates superior or more stable performance over a defined monitoring period, or if interpretability requirements take precedence in a given institutional context [55].

Although full operational deployment was beyond the scope of the present study, the findings support several concrete implementation steps within the CRISP-DM deployment phase. The selected model or set of models could be integrated into the bank’s credit evaluation process as a quantitative decision-support tool, generating a probability-of-default score for each applicant based on the applicant’s financial statements for the two most recent fiscal years. Implementation would require developing an automated data processing pipeline capable of ingesting updated financial data, applying the IHS transformation to continuous predictors, and producing risk classifications consistently and in a timely manner. Methodological documentation and training of credit analysts would be necessary to ensure that model outputs are correctly interpreted and appropriately weighted within the broader underwriting process.

Given the dynamic nature of the financial environment, deployment should be understood not as a terminal step but as the initiation of a continuous model governance cycle. Input data should be updated annually as new financial statements become available, and the relevance of retained predictors should be periodically reassessed to detect structural shifts in the relationship between financial indicators and default risk. Classification thresholds should be reviewed in light of prevailing class imbalance and the asymmetric costs of misclassification, with particular attention to sensitivity as the primary performance criterion. Ongoing monitoring of sensitivity, accuracy, and AUC should serve as the basis for determining whether recalibration of the model or methodological revision is required. This monitoring framework ensures that the deployed system remains responsive to evolving economic conditions and firm-level dynamics in the Slovak corporate lending environment. Taken together, the proposed champion–challenger architecture, automated processing pipeline, and continuous monitoring cycle operationalise the study’s third objective: they translate firm-level predictive performance into a governance-aware deployment framework that supports both the operational resilience of the credit-risk assessment process and the broader stability of the bank-dominated financial system in which it operates.

4. Discussion

This study examined whether historical firm-level information can reliably predict future corporate loan default and which modelling approach offers the strongest performance for practical credit risk assessment in the Slovak banking context. The results show that tree-based models were the most effective overall. The ensemble CART model achieved the best performance across all evaluation criteria, while C5.0 and CHAID also produced very strong results among individual classifiers. The remaining approaches showed more moderate or heterogeneous performance. These findings indicate that financial information from the two preceding fiscal years provides a meaningful basis for distinguishing between non-defaulting and defaulting firms, but also that predictive success depends substantially on the choice of modelling technique. From a practical perspective, the results indicate a clear trade-off between predictive power and interpretability, which is highly relevant to credit risk governance and model deployment.

4.1. Support for H1: Predictive Value of Firm-Level Financial Indicators

The findings support H1, as all eight models achieved a sensitivity exceeding the pre-specified threshold of 70% on the held-out test sample, ranging from 74.14% (DA) to 97.41% (Ensemble CART). This confirms that financial ratios derived from historical financial statements contain sufficient signal to discriminate between defaulting and repaying firms at a practically meaningful level. Although model performance varied across techniques, the consistency of this result across methodologically diverse approaches—decision trees, statistical models, and neural networks—strengthens the conclusion that historical firm-level financial information provides a reliable basis for predicting corporate loan defaults.

Industry affiliation emerged as the dominant first-order discriminator in all three decision tree models and a significant predictor in LR, confirming that sector-level characteristics constitute a fundamental structural dimension of corporate default risk. This finding aligns with prior evidence that industry membership captures systematic differences in business-cycle sensitivity, asset tangibility, and competitive dynamics that are not fully reflected in firm-level financial ratios.

Liquidity indicators, particularly from 2021, appeared consistently across decision trees, the discriminant function structure matrix, and both neural network importance plots. The prominence of liquidity ratio 2 as a splitting variable in both CART and CHAID is particularly noteworthy, given that this indicator filters out inventory—an asset class that may be difficult to liquidate rapidly in distress situations. That liquidity indicators from 2021 outperform their 2022 counterparts in several models may reflect the lagged nature of financial deterioration: firms approaching default may have already spent liquid reserves in the earlier period, with the decline becoming visible in financial statements before the default event itself occurs.

Interest coverage ratio and Net return on sales, both from 2021, ranked among the leading predictors in LR, DA, and the C5.0 importance plot. Interest coverage ratio measures a firm's ability to service its debt from operating earnings and is a direct indicator of debt sustainability; its prominence across interpretable models is theoretically well-grounded. Net return on sales captures operational profitability net of all costs, providing a comprehensive signal of earning quality. The fact that both variables from 2021 carry stronger predictive weight than their 2022 counterparts in several models reinforces the interpretation that early deterioration signals are more discriminating than contemporaneous ones, underlining the value of multi-year longitudinal data in credit risk modelling.

A repeated, theoretically interesting pattern in the coefficient-based models is the opposing temporal effects of leverage indicators between 2021 and 2022. Both the long-term debt ratio and the total debt ratio from 2021 are negatively associated with repayment, whereas their 2022 counterparts are positively associated. This finding may indicate that companies with persistently high debt burdens entering the period under the study were already on a deteriorating trajectory, whereas companies increasing their leverage in 2022 may have done so through active refinancing or investment activity—a signal more consistent with ongoing creditworthiness than with distress. This interpretation is consistent with the view that the direction of change in financial indicators, rather than their level alone, carries important information about future default risk, a point also made in the dynamic credit scoring literature.

The negative coefficient of gross return on assets in the LR model, despite its expected positive association with repayment, warrants methodological attention. This pattern is a well-known consequence of multicollinearity among correlated predictors. This finding underlines the importance of interpreting LR coefficients in the context of the full model rather than in isolation and reinforces the complementary value of the structure matrix and importance plots as interpretive tools. This conclusion is broadly consistent with

recent evidence showing that structured borrower-level data can generate useful default prediction performance when appropriate modelling and imbalance-aware evaluation procedures are applied. Li and Wu [27] report that machine-learning models combined with explainability tools can successfully identify default risk patterns in held-out data, while Kohv and Lukason (2021) [56] show that corporate default prediction benefits from combining financial information with broader firm-level signals rather than relying solely on a narrow set of accounting ratios.

4.2. Support for H2: Comparative Performance of Individual Models

All seven individual classification models exceeded both pre-specified performance thresholds on the held-out test sample—a sensitivity of at least 70% and an AUC of at least 0.8—confirming that firm-level financial indicators provide a sufficient basis for practically meaningful default prediction across a range of methodologically diverse classifiers. However, the degree to which individual models surpass these thresholds varies considerably, and the practical suitability of individual approaches is not uniform across techniques. Among the single classifiers, C5.0 and CHAID delivered the strongest and most balanced performance, indicating that interpretable tree-based models can achieve competitive predictive accuracy while retaining transparency in decision-making. LR likewise produced satisfactory results, supporting its continued relevance as an operationally transparent benchmark. DA showed the weakest performance among the statistical models, clearing both thresholds by the narrowest margins (sensitivity: 74.14%, AUC: 0.823). The neural network models yielded mixed outcomes: MLP achieved relatively strong sensitivity (89.66%) but a more moderate AUC (0.820), while RBF narrowly exceeded both thresholds with a sensitivity of 78.45% and an AUC of 0.803.

The underperformance of the neural network models relative to tree-based classifiers is somewhat counter to the expectation that more flexible architectures will outperform simpler ones in complex classification tasks. A plausible explanation lies in the relatively small sample size, particularly the severely imbalanced minority class of 35 defaulting firms, which may have limited the neural networks' ability to learn stable internal representations, even after oversampling. This interpretation is consistent with prior evidence that neural networks tend to require substantially larger training samples to realise their theoretical performance advantages over interpretable classifiers. This finding is further consistent with prior literature suggesting that individual modelling techniques differ considerably in their practical applicability for credit risk prediction and that model choice must reflect not only predictive strength but also interpretability, stability, and implementation feasibility, as noted by Zhu et al. [57].

4.3. Support for H3: Ensemble Superiority

The findings support H3. The stacked ensemble model outperformed all individual classifiers simultaneously across all three criteria specified in the hypothesis—sensitivity (97.41% vs. 94.83% for the next best individual model, C5.0), overall accuracy (96.90% vs. 94.57%), and AUC (0.951 vs. 0.944 for CHAID)—on the held-out test sample. Notably, the ensemble's advantage over the best-performing individual classifiers, while consistent across all criteria, is most pronounced in sensitivity, which is the primary evaluation criterion in this study given the asymmetric costs of misclassification in credit risk assessment. This result is consistent with a broad stream of research showing that ensemble methods often achieve superior discrimination in credit risk prediction, particularly when the data structure is complex, nonlinear, or imbalanced. Dastile et al. [58] show, through a systematic literature survey, that ensemble classifiers generally outperform single classifiers in credit scoring, while Noriega et al. [59] conclude that the strongest predictive results in credit risk

modelling are frequently achieved by boosted approaches. Ayari et al. [60] highlight the growing dominance of machine-learning and ensemble-based methods in financial credit scoring, and Suhadolnik et al. [61] report that boosting-based ensemble models outperform traditional algorithms such as LR in credit classification. Liu et al. [62] and Xia et al. [63] further demonstrate that ensemble tree-based architectures can outperform conventional benchmark models in predictive tasks. Therefore, the strong performance of the ensemble model in the present study fits well with the broader empirical pattern that combining multiple learners improves predictive stability and the detection of minority-class outcomes.

The ensemble's performance advantage should not, however, be interpreted solely as a technical improvement. In credit risk practice, higher sensitivity has clear managerial relevance: failing to identify a defaulting borrower may be substantially more costly than incorrectly classifying a solvent firm, a principle reflected in the asymmetric cost structure underlying the Basel III definition of default used in this study. As emphasised by De Lange et al. [64], the practical adoption of AI-based credit assessment tools in banking also depends on their explainability and transparency. For this reason, even when ensemble models deliver superior predictive performance, their operational use should ideally be complemented by interpretable challenger models or explainability techniques that support governance, auditability, and regulatory acceptability. Recent research increasingly supports this position. Nallakaruppan et al. [65] emphasise that explainable AI can enhance trust and support data-driven policy use in credit assessment, while Onari et al. [66] propose an explainable decision-support framework precisely because credit decisions often need to be understandable to both institutions and affected clients. This concern is particularly relevant in imbalanced credit risk settings, as Chen et al. [67] show that class imbalance can reduce the stability of post hoc explanations generated by methods, suggesting that explainability itself may become less reliable when minority-class detection is central to the modelling task.

4.4. Practical Implications and Model Governance

Addressing the study's third objective, the practical value of the present findings lies not only in identifying the most accurate model, but also in positioning corporate default classification within a broader framework of model governance for banking-sector resilience, in which CHAID, C5.0, and LR remain highly relevant as transparent alternatives. The champion–challenger deployment architecture proposed in Section 3.4 directly addresses the tension between predictive power and interpretability: the ensemble model serves as the performance-oriented champion, while the interpretable classifiers provide transparent, auditable, and regulatorily acceptable decision support. This arrangement is particularly relevant in the Slovak and broader CEE banking context, where regulatory alignment with EU standards and the prevalence of SME lending create simultaneous demands for predictive accuracy and decision transparency.

Beyond the substantive findings, this study makes a methodological contribution by demonstrating the applicability of the full CRISP-DM pipeline to corporate loan default prediction in a bank-specific Slovak dataset. To the best of the authors' knowledge, no prior study has applied CRISP-DM in this specific context, and the structured pipeline approach offers a replicable and transparent framework that can be adapted to other CEE markets sharing similar institutional characteristics. Slovakia's position within the EU regulatory framework, its IFRS-aligned reporting standards, bank-dominated financial system, and SME-heavy corporate sector are broadly representative of comparable economies, including the Czech Republic, Hungary, Poland, and the Baltic states. The pipeline developed in this study, including the IHS transformation of skewed financial ratios, the oversampling strategy applied exclusively to the training set, and the champion–challenger deploy-

ment architecture, therefore carries methodological transferability beyond the specific national context.

The rapid advance of artificial intelligence in credit risk assessment also raises broader questions for managerial decision making. AI-based tools are increasingly capable of supporting and partially automating credit evaluation, which can improve speed, consistency, and the detection of complex risk patterns. At the same time, the full delegation of prognostic and decision-making functions to algorithms raises significant concerns regarding accountability, transparency, and regulatory compliance, particularly when individual lending decisions must remain explainable to clients and supervisors. We therefore argue that such systems should augment rather than replace expert human judgment. The champion–challenger architecture and the interpretable challenger models proposed in this study offer a concrete mechanism for combining the predictive strength of advanced models with the human oversight and auditability that responsible credit risk governance requires.

4.5. Implications for Financial System Resilience and Macroprudential Governance

Beyond firm-level prediction and individual model governance, the findings of this study carry implications for the resilience of bank-dominated financial systems. Three interconnected dimensions emerge: the recognition of sectoral concentration as a structural source of systemic risk, the bridge between microprudential detection and macroprudential stability, and the contribution of governance-aware deployment architectures to the operational resilience of the credit-risk assessment process.

It should be emphasised that the empirical analysis in this study is conducted exclusively at the borrower level. The implications for system-level resilience are therefore developed as a theoretically grounded analytical positioning rather than as a directly tested empirical claim. This positioning rests on two well-documented mechanisms established in the financial-network and contagion literature [17–19]. First, firm-level defaults are not independent: shared sectoral exposure, common counterparties, and contractual linkages cause them to cluster in time and to propagate through bank balance sheets. Second, the resulting aggregate non-performing-loan dynamics constitute a core macroprudential indicator. Improvements in the firm-level identification of high-risk borrowers therefore feed, in aggregate, into system-level stability outcomes, enabling the proposed pipeline to contribute to financial-system resilience.

The consistent emergence of industry as the dominant first-order discriminator across all three decision tree models, and as a significant predictor in logistic regression, is, from a system-level perspective, more than a technical result. It reflects the well-established regularity that default risk is unevenly distributed across economic sectors and that sectoral concentration is a recognised source of portfolio credit risk and systemic vulnerability [5–7]. Loan portfolios with elevated exposure to high-default sectors are more susceptible to correlated losses during cyclical or sector-specific downturns, and macroprudential authorities increasingly monitor sectoral exposure concentrations as part of their systemic-risk surveillance toolkit. The strong predictive role of industry membership in our results therefore aligns the proposed pipeline with broader supervisory concerns about concentration risk in CEE banking sectors.

A second implication concerns the bridge between microprudential default detection and macroprudential stability outcomes. The reliable identification of high-risk borrowers at the loan-origination stage reduces expected losses, supports more accurate provisioning, and contributes to the stability of bank capital and the quality of aggregate loan portfolios. Aggregate non-performing loan ratios are themselves a core macroprudential indicator [9,14], and improvements in firm-level classification accuracy, particularly in sensitivity, where the cost of false negatives is asymmetric, translate directly into measurable

benefits at the system level. The high sensitivity achieved by the stacked ensemble (above 97%) and by the leading interpretable models (C5.0 and CHAID, both above 92%) therefore has implications well beyond the originating institution.

The third implication concerns operational resilience. International supervisory standards increasingly emphasise that financial institutions must maintain the continuity of critical operations, including credit-risk assessment, under conditions of stress, change, and model uncertainty [68]. The proposed champion–challenger deployment architecture directly responds to this requirement by combining a high-performing ensemble with multiple interpretable challengers that can replace the champion if predictive performance deteriorates or if interpretability requirements take precedence. This arrangement reduces operational dependency on any single model and aligns with macroprudential expectations of model governance, auditability, and explainability articulated in recent supervisory and academic discourse [30,31,64].

Finally, these implications take on a distinctly regional dimension. The structural characteristics that motivate the analysis—bank-dominated financial intermediation, EU regulatory alignment, IFRS-aligned reporting, and SME-heavy corporate sectors—are shared across CEE economies. Coordinated deployment of comparable default-prediction pipelines across regional banking systems could therefore contribute not only to institution-level credit-risk management but also to the coherence and resilience of the broader CEE financial system—an underexplored dimension of macroprudential cooperation in the region.

4.6. Limitations and Future Research

As with any empirical study, the present findings should be interpreted in light of several boundary conditions. First, the dataset comprises 302 firms drawn from a single Slovak bank, which reflects the genuine difficulty of obtaining labelled loan-level default data—a constraint common to the majority of published studies in this field. While this limits the direct generalisability of the findings to other institutions or national contexts, it also ensures high data quality and internal consistency, as all observations derive from a uniform credit evaluation process and a single reporting environment. Future research could extend the analytical framework to multi-institution or cross-country samples to test the robustness of the identified predictor signals under more heterogeneous conditions. More explicitly, because all observations originate from a single lending institution, the results are conditioned on that bank’s client base, sectoral composition, and underwriting standards. The predictor signals identified here should therefore be interpreted as institution-specific until corroborated on multi-bank or cross-country samples, and external validity cannot be assumed without such replication.

Then, the class imbalance of approximately 8:1 in favour of repaying firms is an inherent characteristic of real-world credit portfolios rather than a methodological shortcoming, and the oversampling strategy applied exclusively to the training set represents a principled and widely validated response to this challenge. The strong sensitivity results achieved across all models, particularly the ensemble, suggest that the imbalance was managed effectively. Nevertheless, future studies could explore alternative resampling strategies or cost-sensitive learning approaches to further improve minority-class detection under more severe imbalance conditions.

Further, the study relied on structured financial statement data, which constitutes the primary information basis available to banks under standardised credit evaluation procedures and reflects the data actually used in practice. While the inclusion of behavioural, macroeconomic, or relational variables may offer additional predictive gains, the strong performance achieved with financial indicators alone confirms that structured

firm-level data provides a sufficient and practically accessible basis for default prediction. Lee et al. [69] show that relational structures among borrowers can enhance credit risk prediction, while Kohv and Lukason [56] demonstrate that combining financial information with additional behavioural signals improves discrimination. Liu et al. [70] show that incorporating network-topology features can improve predictive performance beyond that of conventional borrower characteristics alone. The integration of such additional data sources represents a natural and promising direction for future research. The integration of Loss Given Default estimation, which would require recovery and collateral data, represents a complementary direction for future research that would extend the present default occurrence model toward a full expected loss framework. Finally, the models were trained and evaluated on a single temporal cross-section covering fiscal years 2021 and 2022. This design reflects the data availability constraints typical of bank-level research and is consistent with the cross-sectional approach adopted in most comparable studies. The two-year observation window nevertheless captures meaningful longitudinal variation in financial indicators, as evidenced by the opposing temporal effects of leverage variables between the two years. Longitudinal validation across different economic cycles, including periods of macroeconomic stress, would further strengthen confidence in the robustness of the identified predictor signals and the stability of the deployed models over time. In addition, model performance in this study was estimated on a single stratified 60/40 hold-out split. Future work should adopt k-fold or repeated stratified cross-validation, with nested cross-validation for any hyperparameter tuning, to obtain more robust and lower-variance estimates of out-of-sample performance. A further extension would be to benchmark the proposed pipeline against widely used ensemble classifiers not evaluated here, in particular random forests and gradient boosting methods such as XGBoost, in order to quantify any incremental predictive gain relative to the interpretability advantages prioritised in this study. The integration of post hoc explainability tools such as SHAP or LIME with the ensemble model would also represent a valuable extension, allowing the performance advantages of the stacked classifier to be combined with the interpretability requirements and generating instance-level explanations for individual credit decisions, thereby aligning with the iterative CRISP-DM framework.

5. Conclusions

This study investigated the predictability of corporate loan default using firm-level financial information within a structured CRISP-DM analytical pipeline. Applying seven individual classification models and a stacked ensemble to a dataset of 302 Slovak corporate borrowers, the results confirm that historical financial indicators, spanning liquidity, leverage, profitability, and activity dimensions, provide a meaningful and practically sufficient basis for distinguishing between defaulting and repaying firms. Industry affiliation, liquidity ratio 2, interest coverage ratio, net return on sales, and long-term debt ratio emerged as the most consistently informative predictors across methodologically diverse approaches, with variables from the earlier observation year frequently carrying a stronger predictive signal than their contemporaneous counterparts—a finding that underlines the value of multi-year financial data in credit risk assessment.

The comparative evaluation revealed substantial differences in predictive performance across modelling techniques. Tree-based classifiers consistently outperformed statistical and neural network models, with the stacked ensemble achieving the strongest results across all three evaluation criteria: sensitivity, overall accuracy, and AUC. Among individual models, C5.0 and CHAID demonstrated that interpretable decision tree classifiers can achieve competitive predictive performance while retaining the transparency required for practical credit risk governance. LR remained a viable and operationally transparent

alternative, while discriminant analysis and the neural network models produced more moderate results, the latter likely constrained by the limited size of the minority class of defaulted loans. These findings confirm that the choice of modelling technique is a critical determinant of predictive success and that ensemble approaches offer a meaningful performance advantage in imbalanced corporate default prediction tasks.

From a practical standpoint, the study contributes a fully documented and replicable CRISP-DM pipeline for corporate loan default prediction that can be adapted to comparable banking environments across CEE. The proposed champion–challenger deployment architecture, positioning the ensemble as the performance-oriented champion alongside interpretable challengers, provides a concrete operational framework that addresses the simultaneous demands of predictive accuracy, model transparency, and regulatory acceptability. The methodological choices, including the IHS transformation of skewed financial ratios and the training-set-only oversampling strategy, offer transferable solutions to recurring data challenges in credit risk research. Taken together, the findings improve both the empirical understanding of corporate default risk and the methodological guidance for researchers and practitioners working on credit risk prediction in data-limited environments.

Beyond these methodological contributions, the study positions firm-level default classification within a broader framework of banking-sector resilience and macroprudential model governance—its third objective. The dominant role of sectoral affiliation as a predictor highlights concentration risk as a systemic concern. The asymmetric value of high sensitivity, translating directly into reduced aggregate non-performing loan exposures, and the operational-resilience properties of the champion–challenger architecture link borrower-level analytics to system-level outcomes. In bank-dominated CEE financial systems, where institutional and regulatory characteristics are broadly shared, the transferable nature of the proposed pipeline supports a coordinated, resilience-oriented approach to credit-risk modelling whose value extends beyond any single institution.

Author Contributions: Conceptualization, L.D. and V.L.; methodology, L.D.; software, L.D. and V.L.; validation L.D. and V.L.; formal analysis, V.L.; investigation V.L.; resources, L.D.; data curation, V.L.; writing—original draft preparation, L.D. and V.L.; writing—review and editing, L.D. and V.L.; visualization, L.D. and V.L.; supervision, L.D.; project administration, L.D.; funding acquisition, L.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Slovak Grant Agency for Science (VEGA), grant number 1/0509/24.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: This research was supported by the Slovak Grant Agency for Science (VEGA) under Grant No. 1/0509/24.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. CHAID Model

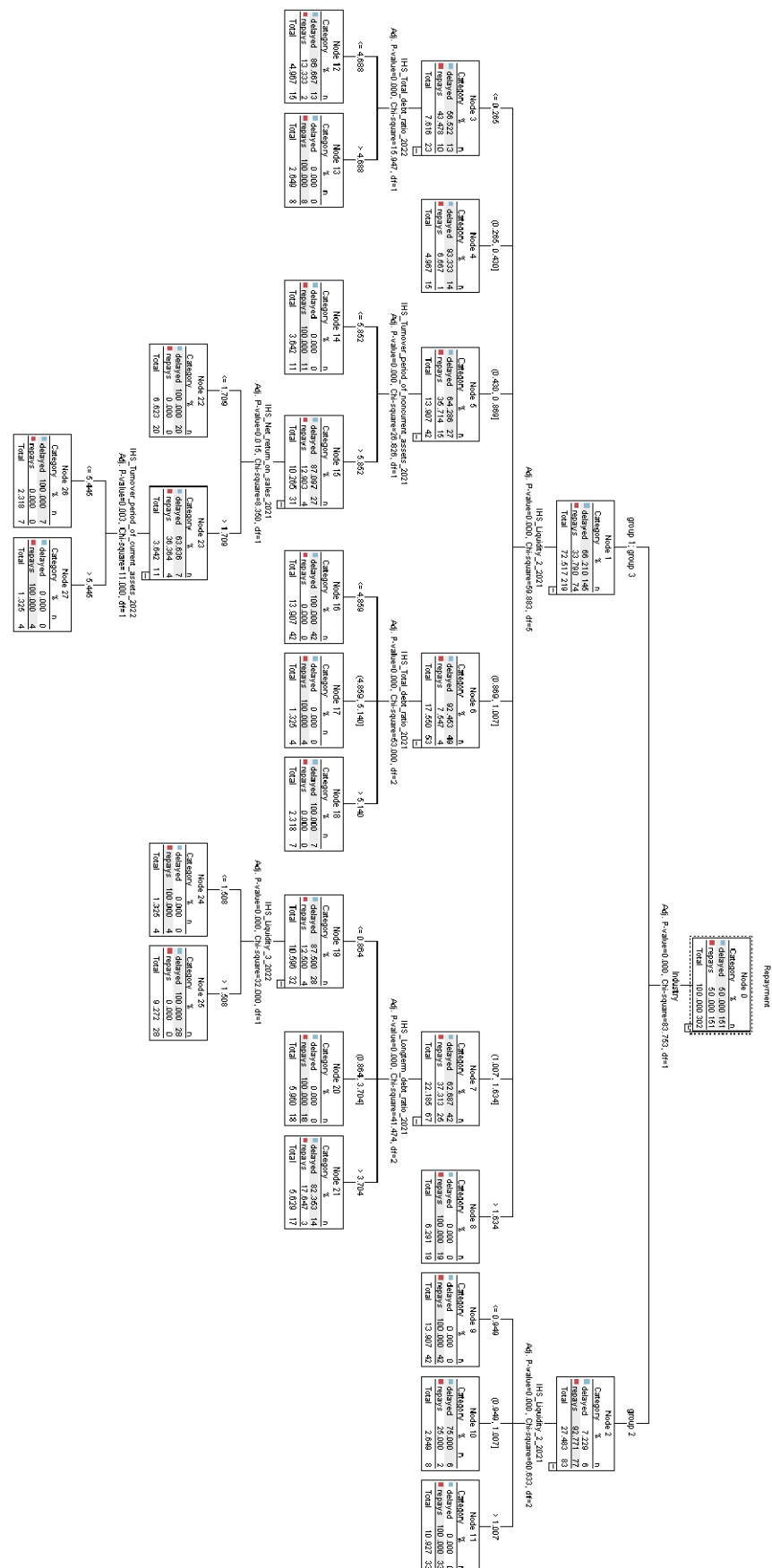


Figure A1. CHAID prediction model for loan repayment.

Appendix B. Predictor Importance Plots

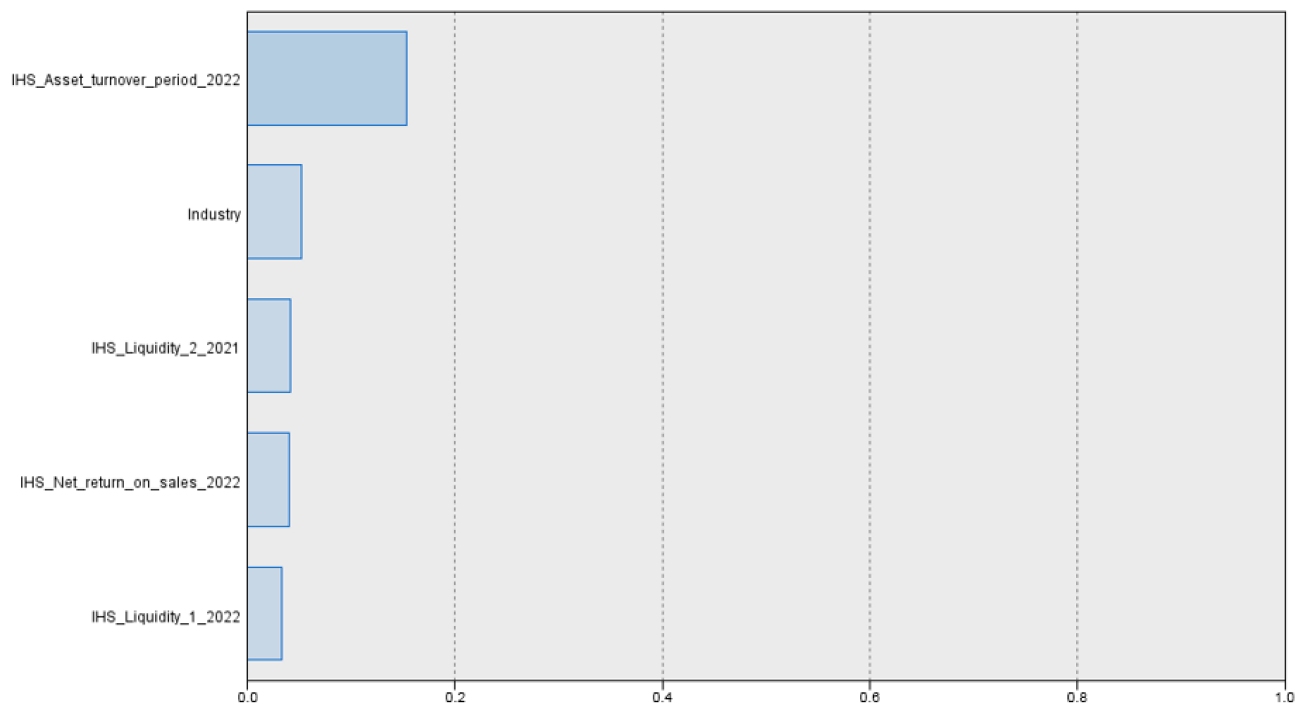


Figure A2. Predictor importance plot for CART model.

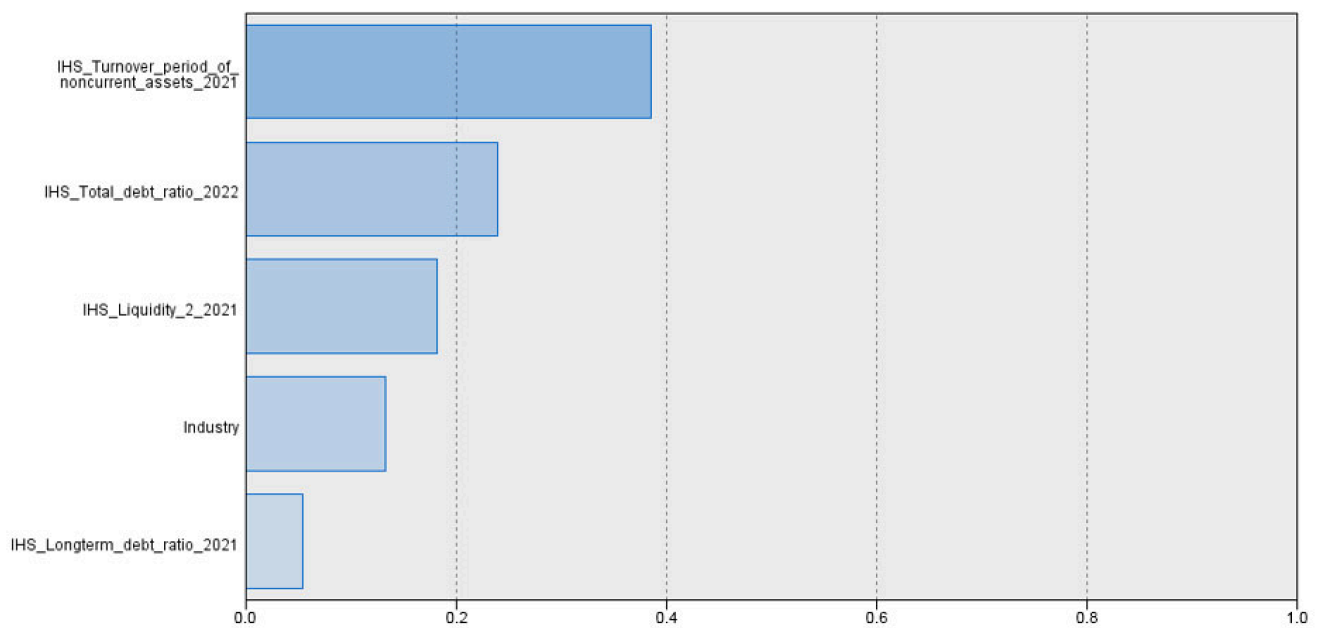


Figure A3. Predictor importance plot for CHAID model.

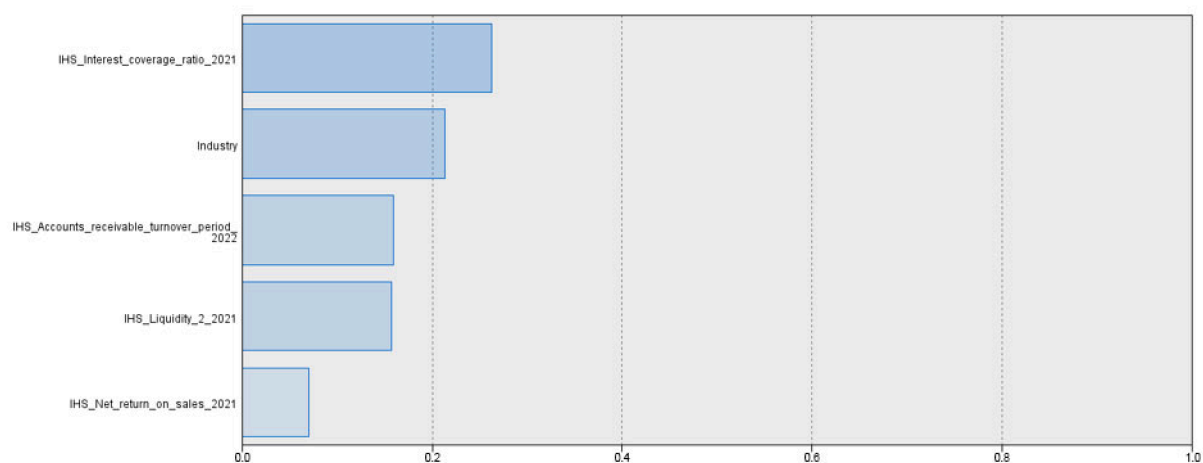


Figure A4. Predictor importance plot for C5.0 model.

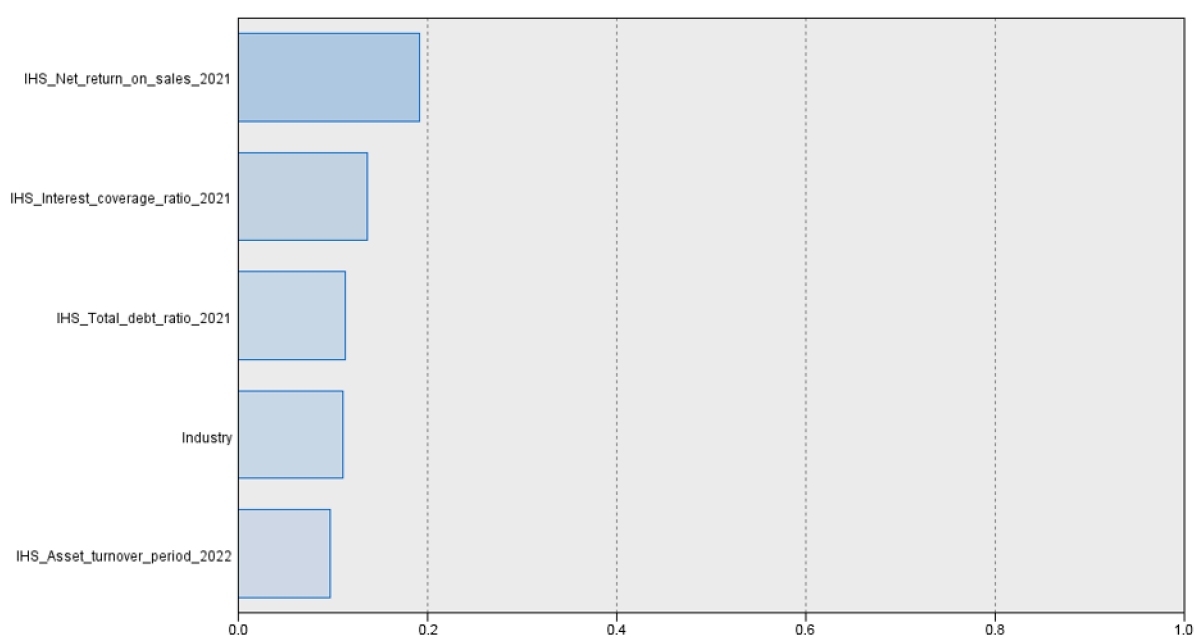


Figure A5. Predictor importance plot for the LR model.

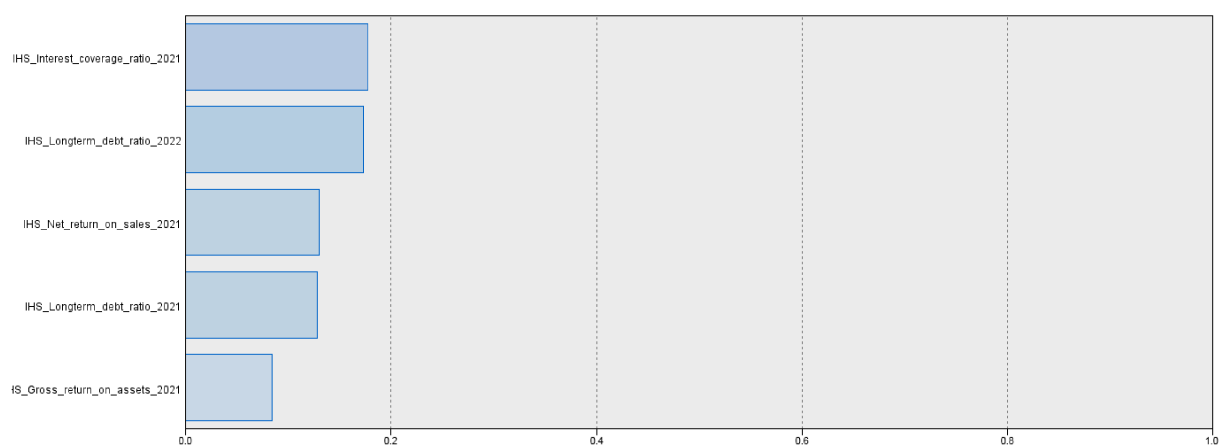


Figure A6. Predictor importance plot for DA model.

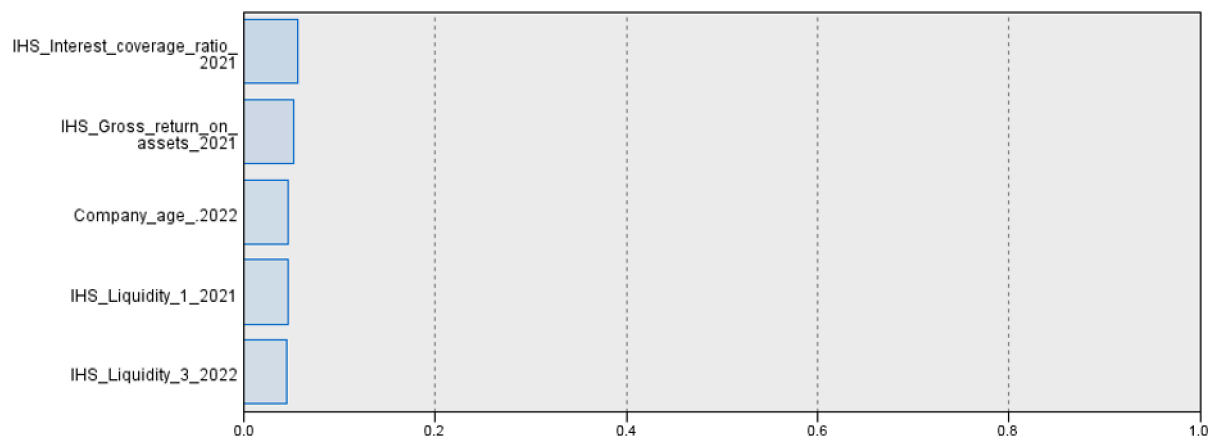


Figure A7. Predictor importance plot for MLP NN model.

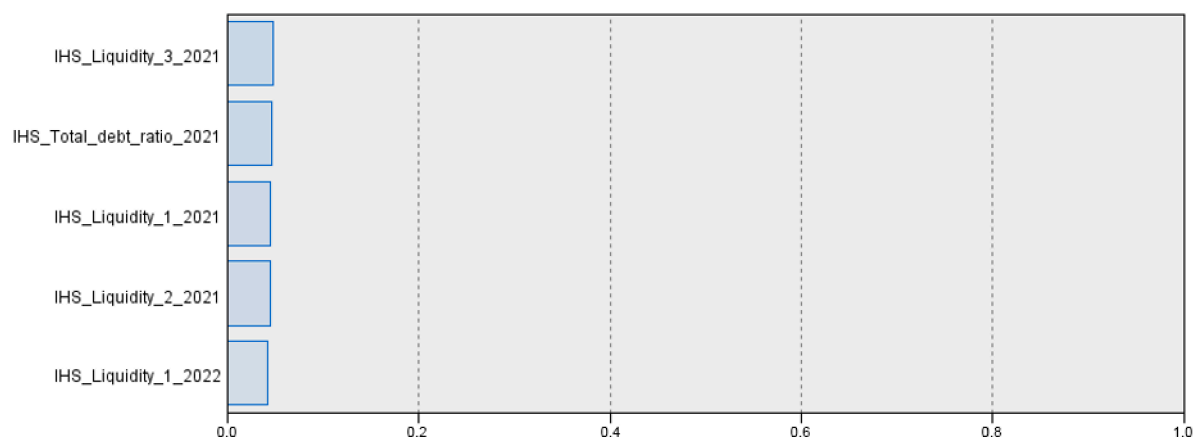


Figure A8. Predictor importance plot for RBF NN model.

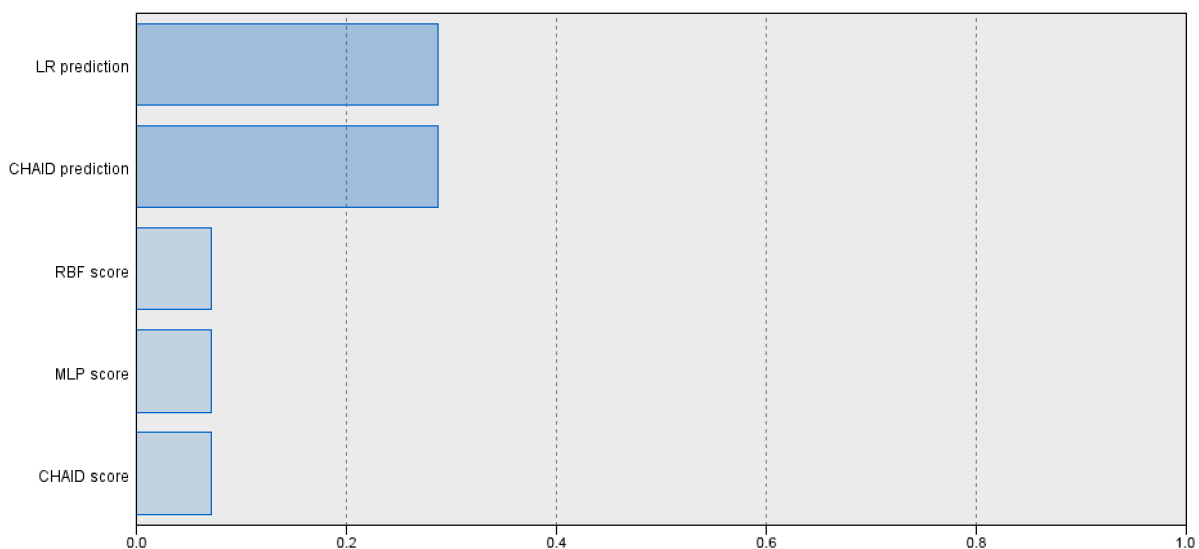


Figure A9. Predictor importance plot for ensemble model.

References

1. Wang, C.; Dong, Y.; Ge, R. Bank branching deregulation and the credit risk of the regional banking sector: Evidence from city commercial banks in China. *Emerg. Mark. Rev.* **2022**, *55*, 100969. [[CrossRef](#)]
2. Nguyen, N.B.; Nguyen, H.D. Impacts of digital transformation and Basel III implementation on the credit risk level of Vietnamese commercial banks. *Int. J. Financ. Stud.* **2024**, *12*, 91. [[CrossRef](#)]

3. Yagli, I.; Topcu, M. Determinants of Credit Risk in the Turkish Banking Sector: Does Ownership Matter? *Sosyoekonomi* **2023**, *31*, 49–67. [\[CrossRef\]](#)
4. Barra, C.; Ruggiero, N. Bank-specific factors and credit risk: Evidence from Italian banks in different local markets. *J. Financ. Regul. Compliance* **2022**, *31*, 316–350. [\[CrossRef\]](#)
5. Di Clemente, A. Modeling portfolio credit risk taking into account the default correlations using a copula approach: Implementation to an Italian loan portfolio. *J. Risk Financ. Manag.* **2020**, *13*, 129. [\[CrossRef\]](#)
6. Dullmann, K.; Scheicher, M.; Schmieder, C. *Asset Correlations and Credit Portfolio Risk: An Empirical Analysis*; Bundesbank Discussion Paper Series 2: Banking and Financial Studies; Deutsche Bundesbank: Frankfurt am Main, Germany, 2007. [\[CrossRef\]](#)
7. Javadi, S.; Osah, T. Credit risk correlation and the cost of bank loans. *Financ. Manag.* **2024**, *53*, 795–832. [\[CrossRef\]](#)
8. Zhang, Y.; Ye, S.; Liu, J.; Du, L. Impact of the development of FinTech by commercial banks on bank credit risk. *Financ. Res. Lett.* **2023**, *55*, 103857. [\[CrossRef\]](#)
9. Bandyopadhyay, A.; Saxena, M. Interaction between credit risk, liquidity risk, and bank solvency performance: A panel study of Indian banks. *Indian Econ. Rev.* **2023**, *58*, 311–328. [\[CrossRef\]](#)
10. Chang, V.; Sivakulasingam, S.; Wang, H.; Wong, S.T.; Ganatra, M.A.; Luo, J. Credit risk prediction using machine learning and Deep Learning: A study on Credit card customers. *Risks* **2024**, *12*, 174. [\[CrossRef\]](#)
11. Chang, V.; Xu, Q.A.; Akinloye, S.H.; Benson, V.; Hall, K. Prediction of bank credit worthiness through credit risk analysis: An explainable machine learning study. *Ann. Oper. Res.* **2024**, *354*, 247–271. [\[CrossRef\]](#)
12. Antony, T.M.; Kumar, B.S. Predicting of credit risk using machine learning algorithms. In *Lecture Notes in Networks and Systems*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 99–114. [\[CrossRef\]](#)
13. Naili, M.; Lahrichi, Y. Banks' credit risk, systematic determinants and specific factors: Recent evidence from emerging markets. *Heliyon* **2022**, *8*, e08960. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Bank for International Settlements. Developments in Credit Risk Management Across Sectors: Current Practices and Recommendations. Basel Committee on Banking Supervision, International Organization of Securities Commissions, International Association of Insurance Supervisors. 2015. Available online: <https://www.bis.org> (accessed on 12 March 2026).
15. Kim, H.; Cho, H.; Ryu, D. Corporate default predictions using machine learning: Literature review. *Sustainability* **2020**, *12*, 6325. [\[CrossRef\]](#)
16. Moscatelli, M.; Parlapiano, F.; Narizzano, S.; Viggiano, G. Corporate default forecasting with machine learning. *Expert Syst. Appl.* **2020**, *161*, 113567. [\[CrossRef\]](#)
17. Acemoglu, D.; Ozdaglar, A.; Tahbaz-Salehi, A. Systemic risk and stability in financial networks. *Am. Econ. Rev.* **2015**, *105*, 564–608. [\[CrossRef\]](#)
18. Battiston, S.; Puliga, M.; Kaushik, R.; Tasca, P.; Caldarelli, G. DebtRank: Too central to fail? Financial networks, the FED and systemic risk. *Sci. Rep.* **2012**, *2*, 541. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Glasserman, P.; Young, H.P. Contagion in financial networks. *J. Econ. Lit.* **2016**, *54*, 779–831. [\[CrossRef\]](#)
20. Zandi, S.; Korangi, K.; Oskarsdottir, M.; Mues, C.; Bravo, C. Attention-based dynamic multilayer graph neural networks for loan default prediction. *Eur. J. Oper. Res.* **2025**, *321*, 586–599. [\[CrossRef\]](#)
21. Chen, H.; Li, S.; Zhang, H.; Zhang, H.; Linag, Y. Research on Loan Default Risk Prediction Methods Based on Graph Neural Networks. In Proceedings of the 8th International Symposium on Computer Science and Intelligent Control, Zhengzhou, China, 6–8 September 2024; pp. 79–83. [\[CrossRef\]](#)
22. Ye, B.; Yang, S.; Zhu, M.; Hu, B.; Wang, D.; Zhang, Z.; He, Y.; Hu, Z.; He, H.; Zhou, J. Granola: Graph Neural network Tackling tabular data for online loan default prediction. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 436–440. [\[CrossRef\]](#)
23. Montevechi, A.A.; De Carvalho Miranda, R.; Medeiros, A.L.; Montevechi, J.A.B. Advancing credit risk modelling with Machine Learning: A comprehensive review of the state-of-the-art. *Eng. Appl. Artif. Intell.* **2024**, *137*, 109082. [\[CrossRef\]](#)
24. Kheneifar, M.A.; Amiri, B. A Novel Hybrid Model for Loan Default Prediction in Maritime Finance Based on Topological Data Analysis and Machine Learning. *IEEE Access* **2025**, *13*, 81474–81493. [\[CrossRef\]](#)
25. Guo, K.; Luo, S.; Liang, M.; Zhang, Z.; Yang, H.; Wang, Y.; Zhou, Y. Credit Default Prediction on Time-Series Behavioral Data Using Ensemble Models. In Proceedings of the 2023 International Joint Conference on Neural Networks (IJCNN), Gold Coast, Australia, 18–23 June 2023. [\[CrossRef\]](#)
26. Millo, Y.; Spence, C.; Xu, R. Algorithmic self-referentiality: How machine learning pushes calculative practices to assess themselves. *Account. Organ. Soc.* **2024**, *113*, 101567. [\[CrossRef\]](#)
27. Li, H.; Wu, W. Loan default predictability with explainable machine learning. *Financ. Res. Lett.* **2023**, *60*, 104867. [\[CrossRef\]](#)
28. Monje, L.; Carrasco, R.A.; Sanchez-Montanes, M. Machine Learning XAI for early loan default prediction. *Comput. Econ.* **2025**, *67*, 4033–4062. [\[CrossRef\]](#)
29. Sotiropoulos, D.N.; Koronakos, G.; Solanakis, S.V. Evolving Transparent Credit Risk Models: A Symbolic Regression approach using Genetic Programming. *Electronics* **2024**, *13*, 4324. [\[CrossRef\]](#)

30. Bucker, M.; Szepannek, G.; Gosiewska, A.; Biecek, P. Transparency, auditability and explainability of machine learning models in credit scoring. *arXiv* **2020**. [CrossRef]
31. Busmann, N.; Giudici, P.; Marinelli, D.; Papenbrock, J. Explainable machine learning in credit risk management. *Comput. Econ.* **2021**, *57*, 203–216. [CrossRef]
32. Markov, A.; Seleznyova, Z.; Lapshin, V. Credit scoring methods: Latest trends and points to consider. *J. Financ. Data Sci.* **2022**, *8*, 180–201. [CrossRef]
33. Antoniadis, A.; Georgiou, K.; Charmanas, K.; Mittas, N.; Angelis, L. Benchmarking Classifiers for Loan Default Prediction using Archetypal Analysis. In Proceedings of the 7th International Conference on Algorithms, Computing and Systems-ICACS, Larissa, Greece, 19–21 October 2023; pp. 13–18. [CrossRef]
34. Akinjole, A.; Shobayo, O.; Popoola, J.; Okoyeigbo, O.; Ogunleye, B. Ensemble-Based Machine learning Algorithm for loan default risk prediction. *Mathematics* **2024**, *12*, 3423. [CrossRef]
35. Chai, N.; Abedin, M.Z.; Wang, X.; Shi, B. Growth potential of machine learning in credit risk predicting of farmers in the industry 4.0 era. *Int. J. Financ. Econ.* **2024**, *30*, 2163–2185. [CrossRef]
36. Lu, Z.; Li, H.; Wu, J. Exploring the impact of financial literacy on predicting credit default among farmers: An analysis using a hybrid machine learning model. *Borsa Istanbul Rev.* **2024**, *24*, 352–362. [CrossRef]
37. Berrada, I.R.; Barramou, F.; Alami, O.B. Towards a machine learning-based model for corporate loan default prediction. *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15*, 565. [CrossRef]
38. Chang, Y.; Chang, K.; Lin, Y. Establishment of Business Loan Default Prediction Model by Integrating Survival Analysis with Logistic Regression. *Sci. Iran.* **2024**, *31*, 2139–2147. [CrossRef]
39. Svabova, L.; Michalkova, L.; Durica, M.; Nica, E. Business Failure Prediction for Slovak Small and Medium-Sized Companies. *Sustainability* **2020**, *12*, 4572. [CrossRef]
40. Papik, M.; Papikova, L. Impacts of crisis on SME bankruptcy prediction models' performance. *Expert Syst. Appl.* **2022**, *214*, 119072. [CrossRef]
41. Gajdosikova, D.; Valaskova, K.; Durana, P. Cross-National Benchmarking of Bankruptcy prediction models across V4 economies. *Int. J. Econ. Sci.* **2026**, *15*, 158–176. [CrossRef]
42. Gavurova, B.; Jencova, S.; Bacik, R.; Miskufova, M.; Letkovsky, S. Artificial intelligence in predicting the bankruptcy of non-financial corporations. *Oeconomia Copernic.* **2022**, *13*, 1215–1251. [CrossRef]
43. Gregova, E.; Valaskova, K.; Adamko, P.; Tumpach, M.; Jaros, J. Predicting Financial distress of Slovak Enterprises: Comparison of selected traditional and learning algorithms methods. *Sustainability* **2020**, *12*, 3954. [CrossRef]
44. Stefko, R.; Horvathova, J.; Mokrisova, M. Bankruptcy Prediction with the Use of Data Envelopment Analysis: An Empirical Study of Slovak Businesses. *J. Risk Financ. Manag.* **2020**, *13*, 212. [CrossRef]
45. Labosova, V.; Duricova, L.; Durana, P. One model fits all? Evaluating bankruptcy prediction across different economic periods. *Economies* **2025**, *13*, 361. [CrossRef]
46. Fidrmuc, J.; Hainz, C. Default rates in the loan market for SMEs: Evidence from Slovakia. *Econ. Syst.* **2010**, *34*, 133–147. [CrossRef]
47. Wirth, R.; Hipp, J. CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining (PADD), Manchester, UK, 11–13 April 2000; pp. 29–39.
48. Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.; Wirth, R. CRISP-DM 1.0-Step-by-Step Data Mining Guide. 2000. Available online: <https://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/es/CRISP-DM.pdf> (accessed on 20 February 2026).
49. Plotnikova, V.; Dumas, M.; Nolte, A.; Milani, F. Designing a data mining process for the financial services domain. *J. Bus. Anal.* **2023**, *6*, 140–166. [CrossRef]
50. Mariscal, G.; Marban, O.; Fernandez, C. A survey of data mining and knowledge discovery process models and methodologies. *Knowl. Eng. Rev.* **2010**, *25*, 137–166. [CrossRef]
51. Johnson, N.L. Systems of frequency curves generated by methods of translation. *Biometrika* **1949**, *36*, 149–176. [CrossRef]
52. Burbidge, J.B.; Magee, L.; Robb, A.L. Alternative transformations to handle extreme values of the dependent variable. *J. Am. Stat. Assoc.* **1988**, *83*, 123–127. [CrossRef]
53. Bellemare, M.F.; Wichman, C.J. Elasticities and the inverse hyperbolic sine transformation. *Oxf. Bull. Econ. Stat.* **2020**, *82*, 50–61. [CrossRef]
54. Aihounton, G.B.D.; Henningsen, A. Units of measurement and the inverse hyperbolic sine transformation. *Econom. J.* **2021**, *24*, 334–351. [CrossRef]
55. van Gestel, T.; Baesens, B. *Credit Risk Management: Basic Concepts, Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital*; Oxford University Press: Oxford, UK, 2009.
56. Kohv, K.; Lukason, O. What best predicts corporate bank loan defaults? An analysis of three different variable domains. *Risks* **2021**, *9*, 29. [CrossRef]

57. Zhu, X.; Chu, Q.; Song, X.; Hu, P.; Peng, L. Explainable prediction of loan default based on machine learning models. *Data Sci. Manag.* **2023**, *6*, 123–133. [\[CrossRef\]](#)
58. Dastile, X.; Celik, T.; Potsane, M. Statistical and machine learning models in credit scoring: A systematic literature survey. *Appl. Soft Comput.* **2020**, *91*, 106263. [\[CrossRef\]](#)
59. Noriega, J.P.; Rivera, L.A.; Herrera, J.A. Machine Learning for Credit Risk Prediction: A Systematic Literature Review. *Data* **2023**, *8*, 169. [\[CrossRef\]](#)
60. Ayari, H.; Guetari, P.R.; Kraiem, P.N. Machine learning powered financial credit scoring: A systematic literature review. *Artif. Intell. Rev.* **2025**, *59*, 13. [\[CrossRef\]](#)
61. Suhadolnik, N.; Ueyama, J.; Da Silva, S. Machine Learning for Enhanced Credit Risk Assessment: An Empirical approach. *J. Risk Financ. Manag.* **2023**, *16*, 496. [\[CrossRef\]](#)
62. Liu, W.; Fan, H.; Xia, M. Step-wise multi-grained augmented gradient boosting decision trees for credit scoring. *Eng. Appl. Artif. Intell.* **2020**, *97*, 104036. [\[CrossRef\]](#)
63. Xia, Y.; He, L.; Li, Y.; Fu, Y.; Xu, Y. A dynamic credit scoring model based on survival gradient boosting decision tree approach. *Technol. Econ. Dev. Econ.* **2020**, *27*, 96–119. [\[CrossRef\]](#)
64. De Lange, P.E.; Melsom, B.; Vennerød, C.B.; Westgaard, S. Explainable AI for credit assessment in banks. *J. Risk Financ. Manag.* **2022**, *15*, 556. [\[CrossRef\]](#)
65. Nallakaruppan, M.; Balusamy, B.; Shri, M.L.; Malathi, V.; Bhattacharyya, S. An Explainable AI framework for credit evaluation and analysis. *Appl. Soft Comput.* **2024**, *153*, 111307. [\[CrossRef\]](#)
66. Onari, M.A.; Rezaee, M.J.; Saberi, M.; Nobile, M.S. An explainable data-driven decision support framework for strategic customer development. *Knowl.-Based Syst.* **2024**, *295*, 111761. [\[CrossRef\]](#)
67. Chen, Y.; Calabrese, R.; Martin-Barragan, B. Interpretable machine learning for imbalanced credit scoring datasets. *Eur. J. Oper. Res.* **2023**, *312*, 357–372. [\[CrossRef\]](#)
68. Basel Committee on Banking Supervision. Principles for Operational Resilience. Bank for International Settlements. 2021. Available online: <https://www.bis.org/bcbs/publ/d516.htm> (accessed on 15 March 2026).
69. Lee, J.W.; Lee, W.K.; Sohn, S.Y. Graph convolutional network-based credit default prediction utilizing three types of virtual distances among borrowers. *Expert Syst. Appl.* **2020**, *168*, 114411. [\[CrossRef\]](#)
70. Liu, Y.; Baals, L.J.; Osterrieder, J.; Hadji-Misheva, B. Leveraging network topology for credit risk assessment in P2P lending: A comparative study under the lens of machine learning. *Expert Syst. Appl.* **2024**, *252*, 124100. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.