International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 282-303

Machine Learning–Based Financial Distress Prediction: Evidence from Borsa Istanbul Manufacturing Firms

Anıl GACAR[1]
Burcu KOCARIK GACAR[2]

Received: 28.12.2025, Accepted: 03.05.2026
DOI Number: 10.5281/zenodo.20563216

Abstract

Accurate financial distress prediction is crucial for investors, regulators, and firm managers to mitigate risk and enhance early warning systems. This study investigates the effectiveness of alternative machine learning approaches for financial distress prediction using firm-level financial ratios. Financial distress is modeled as a binary classification problem, where profitability, liquidity, leverage, and activity ratios are employed as explanatory variables. To assess the impact of alternative modeling approaches, three complementary classification models are implemented: Logistic Regression as a linear model, Support Vector Machine with Principal Component Analysis as a kernel-based nonlinear approach, and Random Forest as a tree-based ensemble method. Model performance is evaluated using accuracy, macro F1-score, recall for the distressed class, and the area under the ROC curve, alongside stratified cross-validation to assess robustness and generalization ability. The empirical results indicate that while LR provides strong and interpretable baseline performance, the RF model consistently outperforms alternative approaches across all key evaluation metrics, achieving the highest macro F1-score and ROC–AUC values. Feature importance analysis reveals that profitability indicators are the most influential predictors of financial distress, followed by liquidity measures. The findings demonstrate that ensemble-based machine learning models offer substantial improvements in financial distress prediction by capturing nonlinear relationships and complex interactions among financial ratios. The results highlight the importance of profitability and liquidity in explaining financial distress and support the use of RF models as a robust and effective framework for early warning systems.

Key Words: Financial Distress Prediction, Support Vector Machine, Ensemble Learning, Random Forest Classifier, Financial Ratios, Early Warning Systems

Jel Codes: G33, C45, G32

[1] Assoc. Prof, Manisa Celal Bayar University, Türkiye, anil.gacar@cbu.edu.tr, http://orcid.org/0000-0002-4571-3886
[2] Assist. Prof. Manisa Celal Bayar University, Türkiye, burcu.kocarik@cbu.edu.tr, http://orcid.org/0000-0001-5944-4456

1. Introduction

Financial distress represents a critical stage in a firm's life cycle, characterized by a deterioration in financial performance that may ultimately lead to bankruptcy if not addressed in a timely manner. The identification of firms exposed to financial distress has therefore been a long-standing concern in corporate finance and accounting research due to its implications for investors, creditors, regulators, and policymakers. In capital markets, accurate assessment of distress risk contributes to efficient capital allocation and financial stability, particularly in economies exposed to macroeconomic uncertainty and financial volatility (Altman et al., 2017). Macroeconomic conditions, especially interest rate and exchange rate movements, are closely linked to financial market performance and may contribute to financial vulnerability at the firm level (Güler & Özçalık, 2018).

The academic literature on financial distress prediction has traditionally emphasized the role of accounting-based financial ratios as early warning indicators. Seminal studies demonstrate that ratios reflecting liquidity, profitability, leverage, and operational efficiency possess significant explanatory power in distinguishing financially distressed firms from healthy ones (Beaver, 1966; Altman, 1968). Despite substantial methodological advancements over time, financial ratios remain widely employed due to their accessibility, interpretability, and strong theoretical foundations in firm performance analysis (Balcaen and Ooghe, 2006).

More recent research has revisited financial distress prediction within increasingly sophisticated empirical frameworks. While machine learning techniques have gained prominence in the literature, empirical evidence suggests that traditional statistical models (such as logistic regression) continue to perform competitively when applied to carefully selected financial ratio datasets (Barboza et al., 2017; Duan et al., 2021). These findings indicate that predictive accuracy depends not only on methodological complexity but also on the relevance and economic meaning of the financial indicators used.

An important strand of the literature emphasizes that the determinants of financial distress vary across industries due to differences in capital intensity, operating structures, and exposure to economic shocks. Manufacturing firms are typically characterized by high fixed asset investments and substantial reliance on external financing, rendering them more sensitive to fluctuations in interest rates, exchange rates, and input costs. Empirical studies provide strong evidence that sector-specific models yield more accurate and informative results than generalized models applied across heterogeneous industries (Ohlson, 1980; Charitou et al., 2004). From an international perspective, emerging markets have attracted increasing attention in financial distress research due to their unique institutional environments and heightened macroeconomic risks. Studies applying established distress prediction models in emerging market contexts generally confirm the relevance of financial ratios, while also emphasizing the need for context-specific empirical analysis (Altman et al., 2017). Stock exchanges in such markets often exhibit higher volatility and structural uncertainty, which may amplify firm-level financial distress risk.

Within this framework, Borsa Istanbul constitutes a particularly relevant setting for examining financial distress, especially in the manufacturing sector, which plays a central role in economic growth, employment, and export performance. Despite the extensive international literature on financial distress prediction, empirical evidence focusing specifically on manufacturing firms in emerging capital markets using recent financial data remains limited.

Accordingly, this study aims to conduct an empirical analysis of financial distress risk among firms operating in the manufacturing sector of Borsa Istanbul by employing 12 widely used financial ratios derived from 2024 financial statements. In this context, the study seeks to address the following research questions:

• Is there a statistically significant difference between machine learning models and traditional statistical models in terms of financial distress prediction performance among manufacturing firms listed on Borsa Istanbul?
• Which financial ratios are the most influential predictors of financial distress within the manufacturing sector?

By addressing these questions, the study aims to provide updated, sector-specific evidence within an emerging market context, thereby contributing to the growing international literature on financial distress prediction and offering practical insights for investors, creditors, and other market participants engaged in financial risk assessment.

2. Literature Review

The prediction of financial distress has been a central topic in corporate finance and accounting research for decades. Early empirical studies primarily relied on accounting-based financial ratios to identify firms experiencing financial difficulties. Beaver (1966) demonstrated that individual financial ratios (particularly liquidity and profitability measures) exhibit strong discriminatory power between failed and non-failed firms. Building on this foundation, Altman (1968) introduced the multivariate Z-score model, which combined several financial ratios into a single discriminant function and significantly improved prediction accuracy. These seminal contributions established financial ratio analysis as a core methodology in financial distress research.

Subsequent studies expanded distress prediction models by adopting alternative statistical techniques. Ohlson (1980) proposed a probabilistic framework using logistic regression, addressing some limitations of discriminant analysis and allowing for greater flexibility in model assumptions. Later comprehensive reviews confirm that, despite methodological diversification, ratio-based models remain widely used due to their interpretability, robustness, and consistency across different economic environments (Balcaen and Ooghe, 2006).

In recent years, financial distress research has increasingly incorporated machine learning and artificial intelligence techniques. Empirical comparisons suggest that models such as neural networks, support vector machines, and random forests may outperform classical statistical methods under certain conditions, particularly when complex and non-linear relationships exist in the data (Barboza et al., 2017). Nevertheless, systematic reviews indicate that traditional approaches (especially logistic regression) continue to perform competitively and remain preferable in many empirical applications due to their transparency and ease of interpretation (Duan et al., 2021). Serdarer Kuzu and Giray Yakut (2020) examine 11 financial ratios and a technology intensity categorical variable for 153 manufacturing firms listed on Borsa Istanbul over the 2012–2015 period, the study compares the predictive performance of SVM kernel functions in financial distress classification across firms with varying levels of technology intensity.

Another important dimension highlighted in the literature is industry specificity. Financial distress determinants differ substantially across sectors due to variations in capital intensity, operating cycles, and exposure to macroeconomic shocks. Manufacturing firms are characterized by high fixed asset investments and significant reliance on external financing, making them especially sensitive to interest rate changes, exchange rate volatility, and fluctuations in input costs. Empirical evidence suggests that sector-specific models yield more accurate and meaningful predictions than generalized models applied across heterogeneous firm samples (Charitou et al., 2004; Ohlson, 1980; Ackuayi, 2025).

Mai et al. (2018) utilize deep learning models by incorporating both financial indicators and textual disclosures, demonstrating that advanced neural network architectures significantly improve bankruptcy prediction performance. Focusing on ensemble learning, Du Jardin (2021) shows that combining multiple classification models based on failure patterns yields superior predictive results compared to single-model approaches. Kou et al. (2021) contribute to the literature by evaluating machine learning-based clustering and classification algorithms for financial risk analysis, emphasizing the importance of model performance and interpretability in decision-making contexts. Complementing these empirical studies, Sun et al. (2019) provide a comprehensive review of state-of-the-art machine learning methodologies for financial distress prediction, highlighting key modeling choices related to sampling strategies, feature selection, and algorithm design.

Emerging markets have attracted growing attention in financial distress research because of their distinct institutional settings and heightened economic volatility. Studies applying established distress prediction models in emerging economies generally confirm the relevance of financial ratios while emphasizing the importance of contextual factors such as inflation, financial market development, and regulatory structures (Altman et al., 2017). These findings underscore the necessity of country- and sector-specific empirical analyses. Gülal

International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 282-303

et al. (2023) examine the prediction of financial distress among firms listed in the BIST Industrials Index by comparing traditional financial distress models with clustering-based techniques. The study evaluates the performance of well-established models, such as ratio-based approaches, alongside unsupervised learning methods to assess their effectiveness in identifying financially distressed firms. Using firm-level financial data, the authors demonstrate that clustering techniques can capture hidden patterns in financial ratios that may not be fully reflected by traditional models. The findings suggest that combining classical distress prediction models with data-driven clustering approaches improves classification accuracy. Overall, the study contributes to the financial distress literature by highlighting the potential of hybrid methodologies in emerging market contexts such as Türkiye. Empirical studies focusing on firms listed on Borsa Istanbul (BIST) commonly employ accounting-based ratios and traditional distress models. Several studies confirm that leverage, profitability, and liquidity ratios are among the most significant predictors of financial distress (Toraman and Karaca, 2016). Aksoy and Boztosun (2021) compare various machine learning classification techniques for predicting financial distress among manufacturing firms listed on BIST and report that non-linear models improve classification performance relative to traditional approaches. Similarly, Arslan and Tutkavul (2024) analyze firms traded in the BIST Watchlist Market and find that accounting-based distress models such as Altman Z-score and Springate remain effective in identifying financially vulnerable firms.

Engin and Durer (2023) employ the XGBoost algorithm to forecast financial distress among BIST100 firms and show that ensemble learning techniques significantly outperform traditional ratio-based models by capturing nonlinear patterns in financial data. While not purely machine learning–based, Özarı (2023) integrates classical financial distress indicators into a VIKOR multi-criteria decision-making framework, which shares similarities with data-driven classification approaches by aggregating multiple predictive signals into a structured ranking mechanism. From a complementary perspective, Terzi (2024) incorporates advanced econometric modeling to examine financial distress, providing a benchmark against which machine learning models can be evaluated. Collectively, these studies underline a growing methodological shift in Türkiye toward data-driven and machine learning–oriented approaches in financial distress prediction. Aslan (2024) applies the Altman Z-score model to electricity generation firms listed on BIST and identifies significant distress risk for several firms under changing financial conditions. Yıldırım and Özekenci (2025) conduct a comparative analysis of financial distress prediction models in the BIST food and beverage sector and show that model performance varies notably across industries. Overall, the literature demonstrates that financial ratios continue to serve as fundamental tools for prediction of financial distress at both international and national levels.

3. Methodology and Empirical Analysis

This section presents the empirical application of machine learning models developed to predict financial distress using firm-level financial ratios. The models are evaluated using multiple performance metrics, including macro F1-score, accuracy, recall for the distressed class, and AUC, in order to ensure a balanced assessment of classification performance. To assess robustness and generalization ability, all models are further validated using stratified cross-validation, allowing for reliable performance comparison across alternative modeling approaches. The analysis aims to identify the most effective modeling approach for accurately distinguishing financially distressed firms from non-distressed ones. Data preprocessing and modeling were conducted in Python using pandas, NumPy, and scikit-learn library.

Data and Variable Definition

This subsection describes the dataset and the variables used in the empirical analysis (Table 1). Financial distress is defined as the binary target variable, while firm-level financial ratios capturing profitability, liquidity, leverage, and activity are employed as explanatory variables. The financial ratios used in this study were obtained from the Finnet database and cover 241 manufacturing firms (www.finnet.com.tr). Drawing on 12 widely utilized financial ratios derived from the 2024 financial statements of manufacturing firms listed on Borsa Istanbul, this study undertakes an empirical examination of financial distress risk at the firm level. In doing so, it aims to extend the international literature on financial distress prediction by providing sector-specific evidence from an emerging market context, while simultaneously delivering practical guidance for investors, creditors, and other market participants engaged in the assessment of financial risk.

Table 1. Financial ratios and financial distress indicator

Variable	Definition	Dtype
Current Ratio	Current Assets / Current Liabilities	Numeric
Quick Ratio	(Current Assets - Inventories) / Current Liabilities	Numeric
Cash Ratio	Cash and Cash Equivalents / Current Liabilities	Numeric
Total Debt to Total Assets	Total Debt / Total Assets	Numeric
Total Debt / Shareholder’s Equity	Total Debt / Shareholder’s Equity	Numeric
Short-Term Debt to Total Debt	Short-Term Liabilities / Total Liabilities	Numeric
Return on Assets (RoA)	Net Profit / Total Assets	Numeric
Return on Equity (RoE)	Net Profit / Shareholders’ Equity	Numeric
Asset Turnover	Net Sales / Total Assets	Numeric
Inventory Turnover	Cost of Goods Sold / Inventories	Numeric
Receivables Turnover	Net Sales / Trade Receivables	Numeric
Target	**Definition**	**Dtype**
Financial Distress	Binary variable equal to 1 if a firm reports negative operating profit or negative net income in a given year, and 0 otherwise.	Binary

Table 2 presents the descriptive statistics of the financial ratios employed in the analysis. The results reveal substantial heterogeneity across firms, particularly in leverage and profitability measures. Liquidity ratios, such as the Current Ratio and Quick Ratio, exhibit moderate dispersion, suggesting notable differences in short-term solvency positions among firms.

Profitability indicators display considerable variability, with both RoA and RoE showing wide ranges and negative minimum values. This indicates that a significant portion of firms experienced operating and equity-based losses during the observation period. Leverage-related ratios, especially Total Debt to Shareholders' Equity, demonstrate extreme dispersion and high maximum values, reflecting pronounced differences in capital structure and potential financial vulnerability.

Table 2. Descriptive statistics of financial ratios

Variable	Count	Mean	Std	Min	25%	50%	75%	Max
Current Ratio	241	2.16	2.20	0.20	1.12	1.49	2.28	16.76
Quick Ratio	241	1.41	1.83	0.13	0.60	0.94	1.49	16.04
Cash Ratio	241	56.65	116.56	0.00	5.50	21.33	56.77	1008.66
Total Debt to Total Assets	241	39.72	18.67	2.61	25.49	36.59	52.59	98.57
Total Debt / Shareholder's Equity	241	123.64	454.20	6.75	34.85	58.74	111.78	6871.97
Short-Term Debt to Total Debt	241	72.63	18.64	4.41	62.09	75.73	86.34	99.89
Return on Assets (RoA)	241	0.69	8.36	-23	-4.00	1.08	4.81	46.84
Return on Equity (RoE)	241	-0.87	23.24	-237	-6.61	1.77	7.32	53.88
Asset Turnover	241	0.75	0.46	0.00	0.43	0.68	0.93	2.46
Inventory Turnover	241	5.71	8.02	0.00	2.54	3.97	6.19	86.33
Receivables Turnover	241	11.04	35.85	0.00	4.18	5.34	7.81	432.33
Financial Distress	241	-	-	-	-	-	-	-

Note: Financial Distress is a binary variable indicating distressed (1) and non-distressed (0) firms.

Activity ratios also vary substantially across firms, particularly Receivables Turnover, which exhibits a highly skewed distribution with extreme upper values. Overall, the descriptive statistics highlight the presence of strong cross-sectional variation in financial conditions, underscoring the suitability of machine learning methods capable of capturing nonlinear relationships and heterogeneous firm characteristics in financial distress prediction.

The distribution of the financial distress variable indicates that 132 observations (54.8%) are classified as non-distressed firms, while 109 observations (45.2%) correspond to financially distressed firms, yielding a total of 241 observations. The class distribution is close to balanced, suggesting that class imbalance is unlikely to bias the model estimation or evaluation results.

Model Specification

To predict financial distress, three different classification models are employed: Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF). These models are selected to represent distinct modeling paradigms, allowing for a comprehensive comparison between linear, kernel-based, and tree-based approaches.

LR

LR is used as the baseline model due to its simplicity and interpretability. It models the probability of financial distress as a logistic function of the explanatory variables and assumes a linear relationship between financial ratios and the log-odds of distress. Despite its restrictive assumptions, LR is widely used in financial distress and bankruptcy prediction studies and provides a transparent benchmark for evaluating more complex models.

LR is a probabilistic classification method grounded in the generalized linear modeling framework. It estimates the conditional probability of a binary outcome by modeling the log-odds of the dependent variable as a linear combination of the explanatory variables (Cox, 1958; Hosmer et al., 2013). Formally, the model assumes that the probability of financial distress follows a logistic distribution, ensuring that predicted values are bounded between zero and one.

From the perspective of financial distress prediction, LR is particularly attractive due to its interpretability and solid statistical foundations. The estimated coefficients directly reflect the marginal impact of financial ratios on the likelihood of distress, allowing for economically meaningful interpretations. A seminal contribution in this field is provided by Ohlson (1980), who demonstrates that financial ratios can be effectively used within a logistic framework to probabilistically predict corporate bankruptcy. Since then, LR has been widely adopted in early warning systems and distress prediction studies as a benchmark model.

Despite its advantages, LR relies on the assumption of linearity in the log-odds and may fail to capture complex nonlinear relationships or interaction effects among financial ratios. Nevertheless, due to its transparency and robustness, it remains a fundamental reference model in the financial distress literature and serves as a natural baseline against which more flexible machine learning approaches can be evaluated (Altman, 1968; Hosmer et al., 2013).

SVM

SVM is employed to capture potential nonlinear relationships between financial ratios and financial distress. By mapping the original feature space into a higher-dimensional space through a kernel function, SVM constructs an optimal separating hyperplane that maximizes the margin between distressed and non-distressed firms. In this study, SVM is combined with Principal Component Analysis (PCA) to mitigate multicollinearity among financial ratios and to improve generalization performance.

SVM is a supervised learning algorithm grounded in statistical learning theory and structural risk minimization. The core objective of SVM is to identify an optimal separating hyperplane that maximizes the margin between different classes, thereby improving generalization performance on unseen data (Vapnik, 1995, 1998). When the original feature space is not linearly separable, SVM employs kernel functions to project the data into a higher-dimensional space, where linear separation becomes feasible.

In financial distress prediction, SVM is particularly useful for capturing nonlinear relationships between financial ratios and distress outcomes. Financial indicators often exhibit complex interactions and nonlinearity that cannot be adequately modeled using linear classifiers. By utilizing kernel-based transformations, SVM can flexibly adapt to such patterns while maintaining strong theoretical guarantees regarding generalization error (Cortes & Vapnik, 1995). However, SVM performance may be adversely affected by multicollinearity and high-dimensional feature spaces, which are common characteristics of financial ratio datasets. To address this issue, SVM is frequently combined with dimensionality reduction techniques such as PCA. PCA reduces feature redundancy and noise by transforming correlated financial ratios into a smaller set of orthogonal components, thereby enhancing numerical stability and reducing overfitting risk (Jolliffe, 2002). Consequently, the integration of PCA and SVM provides a robust framework for financial distress classification by balancing model flexibility and generalization capability.

RF

RF is applied as an ensemble learning method based on a large number of decision trees. Each tree is trained on a bootstrapped sample of the data with random feature selection at each split, and the final classification is obtained through majority voting. This structure enables RF to effectively capture nonlinear patterns and complex interactions among financial ratios while remaining robust to multicollinearity and outliers.

RF is an ensemble learning method based on the aggregation of multiple decision trees and is grounded in the principles of bagging and random feature selection. Introduced by Breiman (2001), the RF algorithm constructs a large number of decision trees using bootstrap samples of the training data, while randomly selecting a subset of features at each split. The final classification decision is obtained through majority voting across individual trees, which substantially reduces variance and improves predictive stability.

From a theoretical standpoint, RF mitigates the overfitting problem commonly associated with single decision trees by decorrelating individual learners and averaging their predictions. This ensemble structure allows the model to capture complex nonlinear relationships and higher-order interactions among explanatory variables without imposing restrictive parametric assumptions (Breiman, 2001). Moreover, RF is inherently robust to multicollinearity, outliers, and noise, making it particularly suitable for financial ratio datasets where strong correlations among variables are prevalent.

From the perspective of financial distress prediction, RF offers an important methodological advantage by accommodating heterogeneous firm characteristics and nonlinear decision boundaries. Additionally, the model provides variable importance measures derived from the mean decrease in impurity or permutation-based criteria, enabling an assessment of the relative contribution of financial ratios to distress prediction. Despite its reduced interpretability compared to linear models, RF has been widely adopted in financial classification problems due to its strong predictive performance and robustness (Hastie et al., 2009). In this study, the use of these three complementary models enables a systematic assessment of how different modeling assumptions and levels of complexity affect financial distress prediction performance. To ensure a comprehensive and unbiased evaluation, multiple performance metrics are employed, reflecting both classification accuracy and class-specific predictive capability.

Evaluation Metrics and Validation Strategy

Model performance is assessed using accuracy, macro F1-score, recall for the financially distressed class, and the area under the receiver operating characteristic curve (ROC–AUC). While accuracy provides a general measure of correct classification, the macro F1-score is emphasized to account for potential class imbalance and to ensure equal weighting of distressed and non-distressed firms. Recall for the distressed class is particularly important from the perspective of financial distress prediction, as failing to identify distressed firms may lead to substantial economic consequences.

To evaluate model robustness and generalization ability, stratified cross-validation is applied throughout the analysis. This validation strategy preserves the original class distribution within each fold and mitigates the risk of performance bias arising from random sample splits. The combined use of test-set evaluation and cross-validated performance metrics allows for a reliable comparison of alternative

International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 282-303

models and supports the selection of the most effective approach for financial distress classification.

4. Results

Based on the evaluation metrics and validation strategy outlined in the previous section, this section presents the empirical findings of the financial distress prediction models. The results are reported and compared across alternative modeling approaches to assess their predictive performance, robustness and and generalization ability.

Data Split and Class Distribution

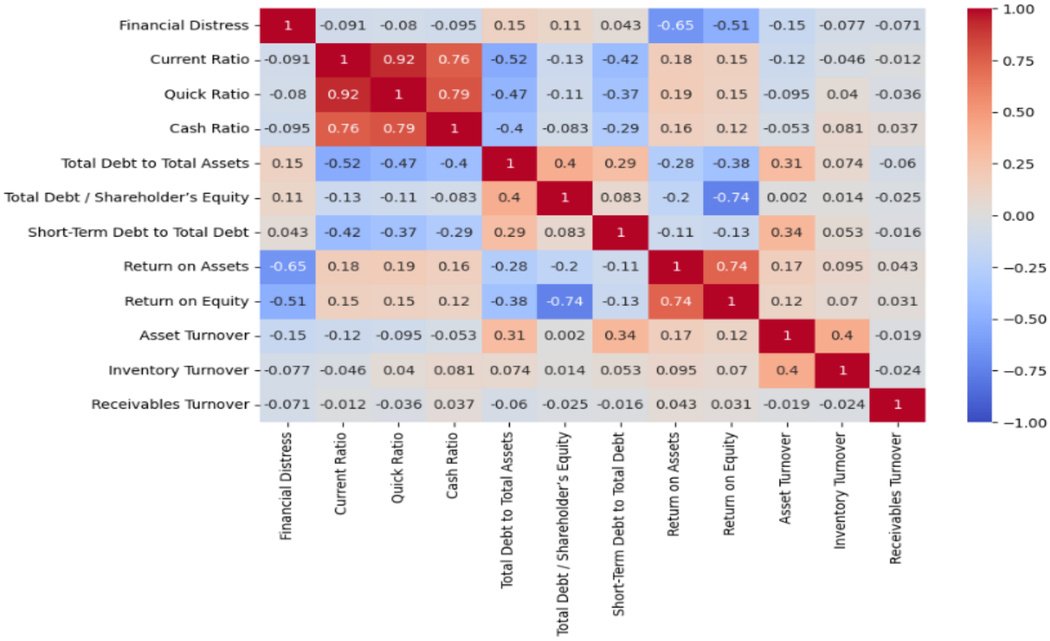
The dataset was divided into training and test sets (20%) using a stratified sampling strategy to preserve the original class distribution. In the training set, 105 observations belonged to the non-distressed class (0) and 87 to the distressed class (1). Similarly, the test set contained 27 non-distressed and 22 distressed observations. This stratified split ensures that class proportions are consistently represented across both subsets, thereby enhancing the reliability of the evaluation results and reducing the likelihood that the observed performance is driven by random sampling effects.

Multicollinearity Assessment

As shown in Figure 1, strong correlations among liquidity ratios (Current-Quick: 0.92; Quick-Cash: 0.79; Current-Cash: 0.76), profitability (RoA-RoE: 0.74), and leverage ratios (Total Debt/Shareholder's Equity-RoE: -0.74; Total Debt/Shareholder's Equity-Total Debt/Total Assets: 0.40) may indicate potential multicollinearity. Accordingly, dimensionality reduction techniques, such as PCA, are applied prior to the SVM classification stage.

As shown in Table 3, Variance Inflation Factor (VIF) analysis was conducted to assess the presence of multicollinearity among the explanatory variables. The highest VIF values were observed for RoE (9.63), Current Ratio (9.19), Quick Ratio (8.95), and RoA (7.12). Importantly, none of the variables exceeded the commonly accepted threshold of VIF > 10, indicating that severe multicollinearity is not present in the model (Kutner et al., 2005; Hair et al., 2010). Although relatively high correlations are observed between profitability measures and among liquidity ratios, such relationships are expected given the financial nature of these indicators. Moreover, the use of L2 regularization in logistic regression mitigates the potential adverse effects of moderate multicollinearity (Hoerl & Kennard, 1970; Hastie et al., 2009).

Figure 1. Correlation structure of financial ratios used in the analysis



Therefore, no variables were excluded from the baseline model, preserving interpretability while maintaining model stability. However, due to the structural characteristics of the SVM algorithm and its sensitivity to multicollinearity and high-dimensional feature spaces, the SVM model was implemented in conjunction with PCA.

Table 3. VIF analysis of financial ratios

Variable	VIF
Return on Equity (RoE)	9.63
Current Ratio	9.19
Quick Ratio	8.95
Return on Assets (RoA)	7.12
Total Debt / Shareholder's Equity	5.83
Total Debt to Total Assets	4.53
Cash Ratio	2.78
Asset Turnover	1.81
Short-Term Debt to Total Debt	1.54
Inventory Turnover	1.41
Receivables Turnover	1.04

Note: VIF values are computed using the training sample.

International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 282-303

Dimensionality Reduction Results

PCA was applied with the objective of retaining 95% of the total variance in the data. Under this criterion, the original feature space consisting of 11 financial ratios was reduced to 8 principal components (Table 4). This level of dimensionality reduction indicates that PCA effectively condensed the information content of the original variables without being overly aggressive. The retained components preserve most of the explanatory power of the original feature set, while simultaneously mitigating multicollinearity and reducing redundancy among highly correlated financial ratios. As a result, the transformed feature space becomes more suitable for subsequent classification using kernel-based methods such as SVM.

Table 4. Principal components (loadings)

PC Loadings	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Current Ratio	0.440	-0.222	0.207	0.042	0.035	-0.176	0.092	-0.235
Quick Ratio	0.426	-0.205	0.282	0.003	0.014	-0.193	-0.020	-0.122
Cash Ratio	0.386	-0.185	0.311	0.085	-0.015	-0.162	-0.211	0.589
Total Debt to Total Assets	-0.406	-0.064	0.235	-0.177	0.339	-0.104	-0.014	0.545
Total Debt / Shareholder's Equity	-0.310	-0.262	0.324	-0.148	0.509	-0.153	-0.215	-0.344

An examination of the PCA loadings provides insight into the financial meaning of the extracted components. For example, the first principal component (PC1), as reported in Table 4, is primarily driven by liquidity and leverage-related variables. High positive loadings are observed for the Current Ratio, Quick Ratio, and Cash Ratio, along with leverage indicators such as Total Debt to Total Assets and Debt to Equity. This component can therefore be interpreted as a liquidity - leverage dimension, capturing firms' short-term solvency conditions in conjunction with their overall debt burden.

Overall, the extracted principal components correspond to economically meaningful financial dimensions - namely liquidity and leverage, profitability, and operational efficiency. These components provide a compact and interpretable representation of the original financial ratios, facilitating more stable and efficient classification in the subsequent SVM analysis.

Model Performance Comparison

LR without dimensionality reduction was employed as the baseline model to provide a transparent benchmark for comparison. The performance of the baseline LR model on the test set indicates strong and well-balanced classification capability. The confusion matrix reveals that 23 out of 27 non-distressed firms and 18 out of 22 distressed firms were correctly classified, corresponding to classification accuracies of approximately 85% and 82% for classes 0 and 1, respectively, as reported in Table 4.

In SVM, hyperparameter tuning was performed using GridSearchCV with the macro F1-score as the optimization criterion. The optimal configuration selected by the inner cross-validation procedure retained 95% of the total variance through PCA (pca__n_components = 0.95) and employed an RBF-kernel SVM with regularization parameter C = 1 and kernel coefficient gamma = 0.01. Under this configuration, the model achieved a best cross-validated macro F1-score of 0.89. The selected value of the regularization parameter indicates that the model achieves a favorable balance between bias and variance, avoiding excessive model complexity while maintaining strong generalization capability. Similarly, the relatively small gamma value results in a smoother decision boundary, preventing overfitting and contributing to a more stable classification structure across different folds.

The performance of the PCA-enhanced SVM model on the test set demonstrates balanced and robust classification capability. The confusion matrix indicates that 22 out of 27 non-distressed firms were correctly classified, while 18 out of 22 distressed firms were accurately identified. This corresponds to five misclassifications for the non-distressed class and four for the distressed class, as reported in Table 4.

In addition to linear and kernel-based models, a tree-based ensemble method was employed to capture potential nonlinear interactions among financial ratios. The confusion matrix of the RF model indicates strong and well-balanced classification performance on the test set. Specifically, 25 out of 27 non-distressed firms were correctly classified, with only 2 misclassifications in this class. Similarly, 19 out of 22 financially distressed firms were accurately identified, while 3 distressed firms were incorrectly classified as non-distressed, as reported in Table 4.

Table 4. Comparison of confusion matrices across classification models

Model		**LR**			**PCA-enhanced SVM**			**RF**	
		Predict			*Predict*			*Predict*	
		0	**1**		**0**	**1**		**0**	**1**
Actual	**0**	23	4	**0**	22	5	**0**	25	2
	1	4	18	**1**	4	18	**1**	3	19

Note: Classification values are reported for the test set.

According to Table 5, the baseline LR model achieved strong overall performance metrics. The model achieved an accuracy of 0.837 and a macro F1-score of 0.835, reflecting balanced predictive performance across both classes. Notably, the F1-score for the distressed class (Financial Distress = 1) was 0.818, demonstrating the model's effectiveness in identifying financially distressed firms. Furthermore, the ROC–AUC value of 0.943 indicates excellent discriminatory power between distressed and non-distressed firms. To further assess the generalization ability of the baseline logistic regression model, a 5-fold cross-validation (CV) procedure was conducted using the macro F1-score as the evaluation metric. The macro F1-scores obtained across the folds were 0.9175, 0.8748, 0.8714, 0.7707, and 0.9161, yielding an average macro F1-score of 0.871 with a standard deviation of 0.053.

In terms of evaluation metrics, the SVM model achieved an accuracy of approximately 0.816 and a macro F1-score of 0.815, reflecting well-balanced performance across both classes. Importantly, the recall for the financially distressed class was approximately 0.818, indicating that the model is effective in identifying firms experiencing financial distress and does not excessively miss high-risk cases. Furthermore, the ROC–AUC value of approximately 0.941 suggests excellent discriminatory power between distressed and non-distressed firms. To further evaluate the generalization performance of the PCA-enhanced SVM model, an outer-like 5-fold CV procedure was conducted using the macro F1-score as the evaluation metric. The resulting macro F1-scores across the folds were 0.9155, 0.8322, 0.8510, 0.8118, and 0.9368, yielding an average macro F1-score of approximately 0.869 with a standard deviation of 0.048.

When compared to the baseline LR model, the PCA-enhanced SVM model exhibits marginally lower test-set performance. However, the difference between the two models is relatively small, indicating comparable predictive capability. Overall, these results suggest that while the baseline logistic regression model performs slightly better on the test set, the PCA-enhanced SVM approach remains a competitive and reliable alternative with strong discriminative ability. The high mean macro F1-score, combined with the relatively low standard deviation, indicates that the model exhibits stable and consistent performance across different data splits. Importantly, this average cross-validated performance is nearly identical to that of the baseline LR model suggesting comparable generalization capabilities between the two approaches.

The RF classifier demonstrates the strongest predictive performance among all evaluated models. Using 300 trees with no depth restriction and a minimum of two observations per leaf, the model achieved a test accuracy of 0.898 and a macro F1-score of 0.896. The recall for the financially distressed class reached 0.864, indicating a high ability to correctly identify firms experiencing financial distress. In terms of discriminatory power, the RF model achieved a ROC–AUC value of 0.965 on the test set, reflecting excellent separation between distressed and non-distressed firms. Furthermore, 5-fold CV results confirm the robustness of the model, with an average macro F1-score of 0.932 and a low standard deviation of 0.028, suggesting strong generalization performance and minimal overfitting.

Table 5. Performance comparison of classification models

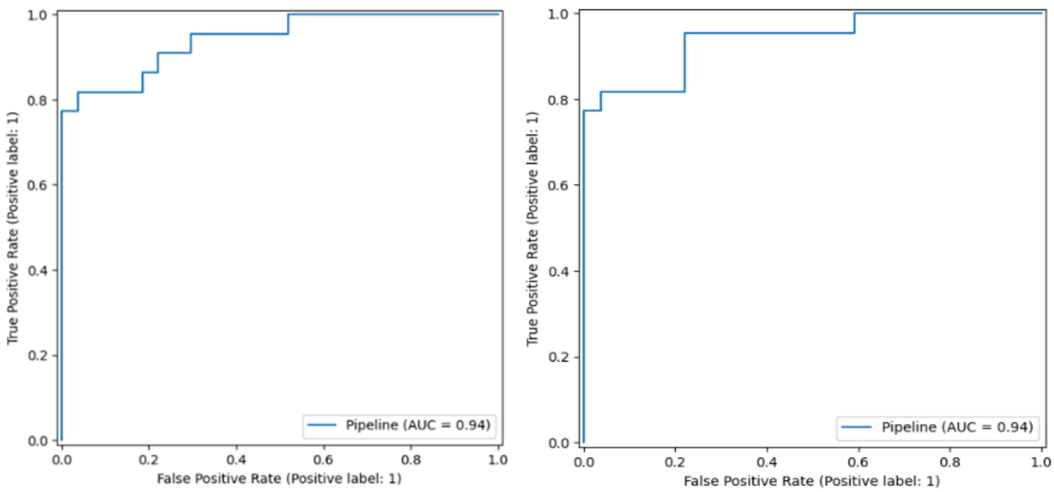
Model	Model Specialty	Accuracy (Test)	MacroF1 (Test)	Recall (FD = 1)	AUC (Test)	CV Macro F1 (Mean ± Std)
LR	Linear (Baseline)	0.837	0.835	0.818	0.943	0.871 ± 0.053
PCA-enhanced SVM	Kernel-based (RBF)	0.816	0.815	0.818	0.941	0.869 ± 0.048
RF	Tree-based and nonlinear	0.898	0.896	0.864	0.965	0.932 ± 0.028

Notes: Accuracy, Macro F1, Recall, and ROC–AUC values are reported for the test set. Cross-validation (CV) results correspond to 5-fold stratified cross-validation. FD denotes Financial Distress.

The RF model outperforms both the baseline LR and PCA-enhanced SVM models across all key evaluation metrics, as shown in Table 5.

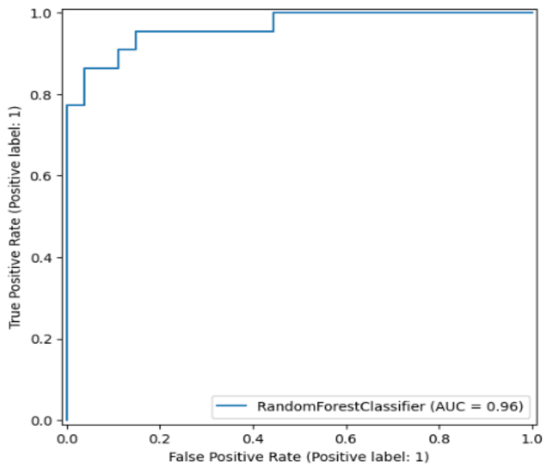
Figure 2. Comparison of ROC curves across classification models

LR PCA-enhanced SVM



RF

International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 282-303



Note: The figure illustrates the discriminatory performance of LR, PCA-enhanced SVM, and RF models.

The ROC curves in Figure 2 demonstrate the strong classification performance of the RF model. The AUC was approximately 0.96, indicating excellent discriminatory ability between financially distressed and non-distressed firms. An AUC of this magnitude implies that, when a distressed firm and a non-distressed firm are randomly selected, the model correctly assigns a higher distress probability to the distressed firm with a 96% likelihood. Within the financial distress prediction literature, this level of performance is considered both highly competitive and methodologically plausible, reflecting a robust separation between the two classes.

Overall, the ROC–AUC results corroborate the findings obtained from accuracy and F1-based metrics, confirming the strong classification quality of the proposed RF model. Cross-validation results further confirmed the robustness and generalizability of the model.

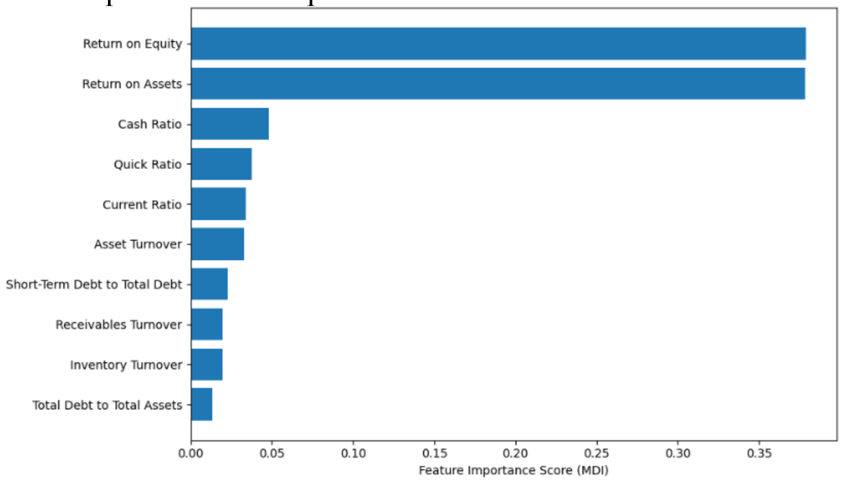
Feature Importance Analysis

The feature importance analysis indicates that profitability measures play a dominant role in financial distress prediction. In particular, RoE and RoA emerge as the most influential predictors, jointly accounting for approximately 70–75% of the total feature importance. This finding underscores the central role of firms' profitability in determining financial stability. Figure 3 presents the feature importance scores derived from the RF model.

Liquidity indicators, including the Cash Ratio, Quick Ratio, and Current Ratio, exhibit moderate importance, suggesting that short-term solvency conditions provide additional explanatory power beyond profitability measures. In contrast, leverage-related and activity-based ratios contribute relatively less to the model, indicating that debt structure and operational efficiency play a secondary role in financial distress prediction within the analyzed sample.

Overall, the results demonstrate that financial distress is primarily driven by firms' ability to generate profits and maintain adequate liquidity. The RF model effectively captures nonlinear interactions among financial ratios and provides a robust and accurate framework for identifying the key financial determinants of distress.

Figure 3. Top 10 Feature importances derived from the RF model



Note: Importance scores are derived from the mean decrease in impurity (MDI) measure of the RF model and sum to one.

Model Selection

The results provide clear evidence regarding the relative performance of the evaluated models. The baseline LR model demonstrates strong predictive capability; however, its performance remains inherently limited by its linear structure. The PCA-enhanced SVM model performs competitively, indicating that nonlinear decision boundaries can capture additional structure in the data, yet it does not yield a substantial improvement over the baseline approach. In contrast, the RF model consistently outperforms both alternative models across all key evaluation metrics and exhibits the highest level of stability, as reflected by its superior cross-validated macro F1-score and low performance variability. This indicates that the ensemble-based, tree-driven structure of RF is particularly effective in capturing complex and nonlinear interactions among financial ratios.

LR is retained as a benchmark due to its interpretability, while the PCA-enhanced SVM model serves as a robustness check to assess the contribution of kernel-based nonlinear modeling. When compared with the baseline LR and PCA-enhanced SVM models, the Random Forest classifier exhibits superior performance across all key evaluation metrics, including macro F1-score, ROC–AUC, and cross-validated performance. While the LR model remains attractive due to its interpretability, the RF model offers a substantial improvement in predictive accuracy and robustness. Accordingly, the RF model is selected as the final model in this study, as it provides the best balance between predictive performance and generalization ability for financial distress classification.

IJCEAS
International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 282-303

5. Conclusion

This study investigates the effectiveness of various machine learning approaches in predicting financial distress using firm-level financial ratios. Three distinct modeling strategies- Logistic Regression, PCA-enhanced Support Vector Machine, and Random Forest- were evaluated to assess the trade-off between model interpretability, complexity, and predictive performance.

The empirical results demonstrate that while the baseline Logistic Regression model provides strong and interpretable performance, ensemble-based tree models offer substantial improvements in predictive accuracy. In particular, the Random Forest classifier consistently outperforms alternative models across key evaluation metrics, including macro F1-score, recall for the distressed class, ROC–AUC, and cross-validated performance. These findings indicate that nonlinear interactions and higher-order relationships among financial ratios play a critical role in explaining financial distress.

Feature importance analysis reveals that profitability indicators, especially RoE and RoA, are the most influential determinants of financial distress, followed by liquidity measures. This highlights the central role of firms' ability to generate sustainable profits and maintain adequate short-term solvency in avoiding financial distress.

From a practical perspective, the results offer important policy and managerial implications. Regulators and financial institutions may benefit from incorporating ensemble-based machine learning models into early warning systems to enhance the timely identification of financially vulnerable firms. For firm managers, the findings underscore the importance of maintaining profitability and liquidity as primary safeguards against financial distress. Overall, the study demonstrates that Random Forest models provide a powerful and robust framework for financial distress prediction, balancing predictive performance and generalization capability.

6. Limitations and Future Research

While this study makes important contributions, several limitations should be noted. First, the analysis relies on a single-industry dataset focused on manufacturing firms with a limited number of observations, which may constrain the generalizability of the results across different industries, countries, or macroeconomic environments. Future research could extend the analysis by employing larger and more diverse datasets, including cross-country or multi-sector samples.

Second, the study relies exclusively on financial ratios as explanatory variables. While these indicators capture important aspects of firm performance, incorporating non-financial variables -such as corporate governance characteristics, market-based indicators, or macroeconomic conditions- may further improve predictive accuracy. Finally, although Random Forest demonstrated superior performance in this study, other advanced ensemble and boosting techniques, such as Gradient Boosting Machines or XGBoost, may yield additional gains. Comparative analyses involving these methods represent promising avenues for future research.

REFERENCES

Ackuayi, S. M. (2025). Comparative analysis of financial distress prediction models: Evidence from African industries. Trent University (Canada) ProQuest Dissertations & Theses, 2025. 32003029.

Aksoy, B., & Boztosun, D. (2021). Finansal başarısızlık tahmininde makine öğrenmesi yöntemlerinin sınıflandırma performansının karşılaştırılması: Borsa Istanbul örneği. Anadolu Üniversitesi Sosyal Bilimler Dergisi, 21(4), pp. 237–268. https://doi.org/: 10.17218/hititsbd.880658

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, 23(4), pp. 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-score model. Journal of International Financial Management & Accounting, 28(2), pp. 131–171. https://doi.org/10.1111/jifm.12053

Arslan, M., & Tutkavul, K. (2024). Yakın izleme pazarında işlem gören işletmelerin finansal başarısızlık risklerinin tahmin edilmesi: BİST'te bir uygulama. MANAS Sosyal Bilimler Dergisi, 13(4), pp. 1299–1319. https://doi.org/10.33206/mjss.1402555

Aslan, T. (2024). Altman Z-skor modeli ile finansal başarısızlık tahmini: Borsa Istanbul'da işlem gören elektrik üretim şirketleri üzerine bir uygulama. Üçüncü Sektör Sosyal Ekonomi Dergisi, 59(2), pp. 876–889. https://doi.org/ 10.15659/3.sektor-sosyal-ekonomi.24.06.2360

Balcaen, S., & Ooghe, H. (2006). 35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems. The British Accounting Review, 38(1), pp. 63–93. https://doi.org/ 10.1016/j.bar.2005.09.001

Barboza, F., Kimura, H., & Altman, E. I. (2017). Machine learning models and bankruptcy prediction. Expert Systems with Applications, 83, pp. 405–417. https://doi.org/10.1016/j.eswa.2017.04.006

Beaver, W. H. (1966). Financial ratios as predictors of failure. Journal of Accounting Research, 4, 71–111. https://doi.org/10.2307/2490171

Breiman, L. (2001). Random forests. Machine Learning, 45(1), pp. 5–32. https://doi.org/10.1023/A:1010933404324

Charitou, A., Neophytou, E., & Charalambous, C. (2004). Predicting corporate failure: Empirical evidence for the UK. European Accounting Review, 13(3), pp. 465–497. https://doi.org/10.1080/0963818042000216811

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), pp. 273–297. https://doi.org/10.1007/BF00994018

Cox, D. R. (1958). The regression analysis of binary sequences. Journal of the Royal Statistical Society: Series B, 20(2), pp. 215–242. https://doi.org/10.1111/j.2517-6161.1958.tb00292.x

Du Jardin, P. (2021). *Failure pattern-based ensembles applied to bankruptcy forecasting.* Decision Support Systems, 143, pp. 113496. https://doi.org/10.1016/j.dss.2021.113496

Duan, J., Sun, J., & Wang, S. (2021). Financial distress prediction using machine learning techniques: A systematic review. Technological Forecasting and Social Change, 165, 120521. https://doi.org/10.1016/j.techfore.2021.120521

Engin, U., & Durer, S. (2023). Financial distress prediction from time series data using XGBoost: BIST100 of Borsa Istanbul. Doğuş Üniversitesi Dergisi, 24(2), pp. 589–604. https://doi.org/10.31671/doujournal.1238432

Finnet Database, www.finnet.com.tr (Accessed in November, 2025).

Gülal, Ö. S., Seçme, G., & Köse, E. (2023). Predicting financial distress in the BIST Industrials Index: Evaluating traditional models and clustering techniques. Ekonomi Politika ve Finans Araştırmaları Dergisi, 8(4), pp. 660–680. https://doi.org/10.30784/epfad.1370893

Güler, S., & Özçalık, M. (2018). Hisse getirisi, faiz oranı ve dolar kuru ilişkisi: BİST'te bir uygulama. Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi, 16(4), 291–308. https://doi.org/10.18026/cbayarsos.505983

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate data analysis (7th ed.). Pearson.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer, New York.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), pp.55–67. https://doi.org/10.2307/1267351

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (3rd ed.). Wiley, New York.

Jolliffe, I. T. (2002). Principal component analysis (2nd ed.). Springer, New York.

Kou, G., Peng, Y., & Wang, G. (2021). Evaluation of clustering algorithms for financial risk analysis using MCDM methods. Information Sciences, 538, pp.132–147. https://doi.org/10.1016/j.ins.2020.05.037

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). Applied linear statistical models (5th ed.). McGraw-Hill.

Mai, F., Tian, S., Lee, C., & Ma, L. (2018). Deep learning models for bankruptcy prediction using textual disclosures. European Journal of Operational Research, 274(2), pp. 743–758. https://doi.org/10.1016/j.ejor.2018.10.024

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research, 18(1), pp. 109–131. https://doi.org/10.2307/2490395

Özarı, Ç. (2023). Finansal başarısızlık açısından Borsa Istanbul Ticaret Endeksi'inde işlem gören işletmelerin VIKOR yöntemi ile değerlendirilmesi. Nişantaşı Üniversitesi Sosyal Bilimler Dergisi, 11(2), pp. 655–673. https://doi.org/10.52122/nisantasisbd.1386786

Serdarer Kuzu, B., & Giray Yakut, S. (2020). Destek vektör makineleri yardımıyla imalat sanayisinde mali başarısızlık tahminlerinin teknoloji yoğunluğuna göre incelenmesi. Osmaniye Korkut Ata Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 4(2), 36-54.

Sun, J., Li, H., & Lang, J. (2019). Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. Knowledge-Based Systems, 173, pp 33–50. https://doi.org/10.1016/j.knosys.2019.02.006

Terzi, S. (2024). Examining the relationship between financial statement comparability with financial distress: Evidence from Borsa Istanbul. Dokuz Eylül Üniversitesi İşletme Fakültesi Dergisi, 25(2), pp. 40–51. https://doi.org/10.24889/ifede.1467119

Toraman, C., & Karaca, S. S. (2016). Finansal oranlar yardımıyla finansal başarısızlığın tahmini: Borsa Istanbul imalat sanayi işletmeleri üzerine bir uygulama. Muhasebe ve Finansman Dergisi, (72), pp. 175–194. https://doi.org/10.25095/mufad.396683

Vapnik, V. N. (1995). The nature of statistical learning theory. Springer, New York.

Vapnik, V. N. (1998). Statistical learning theory. Wiley, New York.

Yıldırım, M. Ç., & Özekenci, S. Y. (2025). Finansal başarısızlık tahmin modellerinin karşılaştırılması: BİST gıda ve içecek sektöründe bir uygulama. Üçüncü Sektör Sosyal Ekonomi Dergisi, 60(1), pp. 982–1004. https://doi.org/ 10.15659/3.sektor-sosyal-ekonomi.25.03.2539.