

Article

Comparative Analysis of Machine Learning Models for Bankruptcy Prediction in the Context of Pakistani Companies

Domicián Máté ^{1,2,*} , Hassan Raza ³  and Ishtiaq Ahmad ⁴

¹ Department of Engineering Management and Enterprise, Faculty of Engineering, University of Debrecen, 4028 Debrecen, Hungary

² Department of Higher Education and Training—National Research Foundation, South African Research Chairs Initiative in Entrepreneurship Education, Department of Business Management, University of Johannesburg, Johannesburg 2006, South Africa

³ Department of Management Sciences, Shaheed Zulfikar Ali Bhutto Institute of Science & Technology University, Islamabad 44000, Pakistan; dr.hassan@szabist-isb.edu.pk

⁴ Department of Management Sciences, National University of Modern Languages University, Islamabad 44000, Pakistan; iahmed@numl.edu.pk

* Correspondence: mate.domician@eng.unideb.hu; Tel.: +36-(20)-9915258

Abstract: This article presents a comparative analysis of machine learning models for business failure prediction. Bankruptcy prediction is crucial in assessing financial risks and making informed decisions for investors and regulatory bodies. Since machine learning techniques have advanced, there has been much interest in predicting bankruptcy due to their capacity to handle complex data patterns and boost prediction accuracy. In this study, we evaluated the performance of various machine learning algorithms. We collect comprehensive data comprising financial indicators and company-specific attributes relevant to the Pakistani business landscape from 2016 through 2021. The analysis includes AdaBoost, decision trees, gradient boosting, logistic regressions, naive Bayes, random forests, and support vector machines. This comparative analysis provides insights into the most suitable model for accurate bankruptcy prediction in Pakistani companies. The results contribute to the financial literature by comparing machine learning models tailored to anticipate Pakistani stock market insolvency. These findings can assist financial institutions, regulatory bodies, and investors in making more informed decisions and effectively mitigating financial risks.

Keywords: bankruptcy prediction; machine learning models; comparative analysis; Pakistani companies; financial risk assessment



Citation: Máté, Domicián, Hassan Raza, and Ishtiaq Ahmad. 2023. Comparative Analysis of Machine Learning Models for Bankruptcy Prediction in the Context of Pakistani Companies. *Risks* 11: 176. <https://doi.org/10.3390/risks11100176>

Academic Editors: Mogens Steffensen and Anatoliy Swishchuk

Received: 10 August 2023
Revised: 17 September 2023
Accepted: 7 October 2023
Published: 10 October 2023



Copyright: © 2023 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) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Business failure prediction is an essential area of finance that helps identify the probability of organizations failing and eventually going bankrupt. The failure of a business can cause significant losses for creditors and stockholders; the principal investors in any company. As a result, many users of financial statements place a high value on their capacity to anticipate insolvency. These users include, for instance, businesses, investors, credit rating organizations, auditors, and regulators. Using a model to predict bankruptcy and identify early warning signs becomes increasingly essential during a financial and economic crisis. Many studies have been conducted over the years, developing various statistical and machine learning models to predict bankruptcy (Qu et al. 2019; Tunio et al. 2021).

In emerging economies like Pakistan, predicting whether a company will fail financially is a significant challenge for stock market investors. In the past, traditional statistical models were used to predict bankruptcy. However, they often have limitations, such as assuming a linear relationship between variables and relying on a subjective selection of variables. Meanwhile, machine learning models have gained popularity for this task due to their ability to analyze large amounts of data and identify complex patterns that may be

difficult for humans to detect (Islam et al. 2022). Machine learning models are increasingly being used because they can continuously improve accuracy by learning from and adapting to new data in real-time. Machine learning models can handle non-linear correlations and automatically select the most significant predictor variables. Decision trees, random forests, neural networks, and support vector machines are some widely used machine learning models that can predict whether a business will fail. Using machine learning models for predicting bankruptcies can result in more accurate and reliable predictions when applied extensively.

Different fundamentals apply to emerging economies like Pakistan, such as the limited availability of historical data, lax bankruptcy laws, volatile stock markets, and unstable political and economic environments. By utilizing advanced algorithms to effectively analyze extensive data, machine learning models can assist in overcoming some of these difficulties. These algorithms can spot patterns and connections in the data that previous statistical analyses like multiple discriminant analysis (MDA) might miss, giving information about potential hazards and opportunities in the financial world. Hence, machine learning models provide an essential tool for predicting financial failure in developing nations like Pakistan, allowing investors and related financial institutions to make better judgments and more successfully manage risk against the financial failures of the companies.

For several reasons, understanding the intricacies of bankruptcy prediction in emerging markets is crucial for a global readership. Firstly, with the increasing interconnectedness of global financial systems, comprehending risks in one market can have broader implications for international investors, financial institutions, and policymakers. Additionally, international investors often seek diversification by investing in emerging markets, and insights from studies in this context can aid in better risk assessments and informed decision-making (Li et al. 2021). Furthermore, extending the application of established models to emerging markets enriches the academic discourse, either by strengthening the general applicability of the models or highlighting their limitations (Patel et al. 2022). Lastly, while the focus may be on Pakistani companies, the challenges faced in emerging markets are often similar, making the findings potentially transferable to other emerging or frontier markets (Wang et al. 2021).

By juxtaposing machine learning models, which are widely accepted in developed contexts, against the backdrop of a quintessential emerging market like Pakistan, our study aims to bridge a significant gap in the literature. The insights, challenges, and lessons drawn from this analysis are relevant to local stakeholders and resonate with a global audience seeking a comprehensive understanding of global financial landscapes (Khan et al. 2021). This research contributes to the existing body of knowledge and offers valuable insights for academia, practitioners, and policymakers alike.

Prior research has extensively evaluated bankruptcy prediction, but mainly in the context of established economies, potentially rendering them less applicable to emerging markets (Papan and Spyridou 2020). Predicting bankruptcy in such environments is complicated due to data paucity, unpredictable stock markets, and a fluid political landscape (Kliestik et al. 2020). The research leverages advanced machine learning to decipher intricate data patterns, eclipsing traditional statistical methods. The goal is to discern the most efficient model and relevant financial ratios tailored to the Pakistani backdrop, providing invaluable insights to investors and lenders. This work innovates by holistically assessing machine learning models for bankruptcy predictions within Pakistan, aiming to bolster informed decision-making and fortify the nation's financial milieu (Kanapickienė et al. 2023).

This study aims to identify the most effective method for predicting business failure in Pakistani non-financial firms by applying multiple machine learning models to 36 financial ratios to answer two critical questions: the suitability of model selection and the selection of the most appropriate ratios. Before investing or lending money, stakeholders such as shareholders, managers, banks, and creditors must evaluate a company's financial condition. This study seeks to contribute to the literature on business failure prediction by

shedding light on the most effective methods for predicting business failure in Pakistan. This study also has practical implications for investors, lenders, and regulators, who can use the findings to make informed judgments regarding investing in or lending money to non-financial companies in Pakistan.

2. Literature Review

Bankruptcy studies began in the 1930s with ratio analyses to predict future bankruptcy and continued through the mid-1960s using single factor/ratio analyses for comparison purposes. [Beaver \(1966\)](#) introduced a univariate analysis in the late 1960s, which provided the first statistical justification for the ability of financial ratios to account for defaults. It employed MDA techniques to estimate the likelihood of bankruptcy in a sample of enterprises. In the early phases of bankruptcy prediction, [Altman](#) utilized discriminant analysis in 1968, a widely used method for model development ([Altman 1968](#)). Following this, various bankruptcy prediction studies were encountered, each with unique models and factors in their quantity and variation. [Altman's \(1968\)](#) original model is a five-factor multivariate discriminant analysis model, whereas [Jo et al.'s \(1997\)](#) model has 57 factors. In other models, the number of factors considered ranges from one to 57. Since then, creditors, tribunals, auditors, accountants, and researchers have all come to adopt the Z-score methodology ([Deakin 1972](#); [Edmister 1972](#); [Altman et al. 1977](#); [Laitinen 1991](#); [Grice and Ingram 2001](#)). However, the multinormality hypothesis was ultimately refuted in favor of the hypothesis that explanatory variables have distinct distributions. In order to anticipate bankruptcy, the logit ([Ohlson 1980](#)) and probit models were frequently used ([Zmijewski 1984](#)).

In the 1990s, neural networks ([Lennox 1999](#)) and the genetic algorithm ([Shin and Lee 2002](#)) from the machine learning subfield of artificial intelligence were introduced. They generated compelling forecasting results without requiring statistical restrictions. Using data from 1985 to 2013, [Barboza et al. \(2017\)](#) compared the accuracy of five machine learning models for predicting bankruptcy to more established statistical methods (discriminant analysis and logistic regression). Comparing the new machine learning techniques to traditional ones significantly increased the accuracy of bankruptcy forecasts and provided greater precision than the statistical methods ([Aziz and Dar 2006](#)).

The model and its financial ratios must be appropriately chosen to predict bankruptcy accurately ([Tang and Chi 2005](#)). Statisticians have devised numerous techniques for selecting relevant predictor variables, such as principal components analysis (PCA), MDA, and the least absolute shrinkage and selection operator (LASSO) technique ([Pompe and Bilderbeek 2005](#)). The initial list of explanatory variables may include up to 50 ratios derived from detailed information obtained from balance accounts and income statements. However, typically, only 5 to 10 ratios are chosen for the model ([Tian et al. 2015](#)). Variable selection procedures may differ depending on the data used, such as annual or quarterly financial data or ratios averaging several years before the bankruptcy ([Fan and Li 2001](#)). The effects of model accuracy during periods of economic decline have been studied, and bankruptcy prediction models for SMEs and publicly traded companies have been developed ([Du Jardin 2015](#)). However, these models need more access to the necessary data of some businesses ([Karas and Režňáková 2014](#); [Ciampi 2015](#)).

[Shi and Li \(2019\)](#) show that logit and neural network models are the most popular and extensively researched methods for predicting bankruptcy. [Mai et al. \(2019\)](#) evaluated conventional learning machine models with convolutional neural networks on an extensive database of public corporations and discovered that the simplified models performed reliably. [Hosaka \(2019\)](#) discovered that convolutional neural networks provide more accurate predictions. However, there has yet to be an agreement on how to use convolutional neural networks to predict bankruptcy. In recent years, artificial intelligence algorithms and machine learning models have demonstrated promising results in predicting business failure without requiring statistical assumptions. Numerous researchers have compared the accuracy of traditional statistical models to machine learning techniques and discovered

that machine learning techniques function more effectively. However, consensus on the most influential business failure prediction model is still needed. This paper seeks to investigate the existing models and techniques used to predict business failure and to determine the most effective approach.

Logistic regression has been recognized as a straightforward and comprehensible model that has demonstrated strong performance in the context of binary classification tasks (Mood 2010). Nevertheless, it is worth noting that the model may encounter challenges in dealing with intricate non-linear relationships and can be influenced by the exclusion of factors, even if these variables are not directly related to the independent variables in the model. Random forests (RFs) have been demonstrated to be beneficial in managing high-dimensional data and large datasets while also exhibiting resilience against overfitting. Nevertheless, they compromise a certain degree of interpretability. The simplicity and efficiency of naive Bayes have been demonstrated, particularly in text classification.

Shetty et al. (2022) conducted a comprehensive study comparing the bankruptcy prediction power of five machine learning models with traditional statistical techniques. Using North American firms' data from 1985 to 2013, they found that machine-learning models outperformed discriminant analysis and logistic regression in accuracy. Their results demonstrated the potential of machine learning techniques in enhancing bankruptcy prediction.

Another study by Kitowski et al. (2022) focused on identifying symptoms of bankruptcy risk based on bankruptcy prediction models in Poland. They employed various machine learning techniques, including extreme gradient boosting (XGBoost), support vector machines (SVMs), and deep neural networks. By utilizing easily obtainable financial ratios, they achieved 82–83% global accuracy in predicting bankruptcies for Polish enterprises. Their model proved simple yet accurate, providing a user-friendly tool for discriminating between bankrupt and non-bankrupt firms. In the agricultural sector, they explored bankruptcy risk prediction to ensure the sustainable operation of agriculture companies. They applied different Z-score models and calculated bankruptcy probabilities on a sample of agricultural companies listed on the Belgrade Stock Exchange. Their research highlighted the importance of bankruptcy prediction in maintaining the sustainability of agricultural businesses.

Lombardo et al. (2022) developed a dataset and benchmarks for bankruptcy prediction in the context of the American stock market. Their study focused on machine learning techniques and their application in predicting bankruptcy in the American stock market. They investigated the design and application of different machine learning models for estimating survival probabilities over time and default prediction using time-series accounting data. The dataset used in their experiments included 8262 different public companies listed on the American stock market between 1999 and 2018.

Furthermore, Kainth and Wahlstrøm (2021) investigated the impact of International Financial Reporting Standards (IFRS) on bankruptcy prediction for privately held Swedish and Norwegian companies. Their study examined the transparency promoted by IFRS and its influence on bankruptcy prediction.

Nevertheless, the method assumes naive independence among characteristics. Decision trees provide transparency and versatility (Liang et al. 2016). Nevertheless, it is essential to acknowledge that these models are susceptible to overfitting and instability. The performance of machine learning algorithms, such as AdaBoost and GBT, was enhanced with the integration of weak learners, as demonstrated by Bühlmann and Hothorn (2007). Nevertheless, these models exhibit sensitivity to noisy data and necessitate hyperparameter adjustment.

3. Data Design and Methodology

The Pakistan Stock Exchange (PSX) provided the data for this investigation. The imperative to comprehend the nuanced fluctuations in the economic landscape is not only an exercise in intellectual curiosity but also a matter of practical necessity. This research

dives into the financial health of firms by leveraging a dataset from the PSX spanning the years 2016 to 2021. This period was crucial, enabling us to discern patterns that indicate either an ascent or descent in the broader economic context of Pakistan.

For our sample constitution, rigorous criteria were indispensable. We began with a focus on non-financial firms with a continuous listing on the PSX, leading to the inclusion of 385 publicly traded entities. These companies, diverse in their economic sectors, were the bedrock of our analysis. Essential to our approach was the extraction of 36 financial ratios from their financial disclosures. These ratios, when integrated, functioned as the independent variables in our machine learning algorithms. Moreover, to uphold the integrity of our study, firms with ambiguous or incomplete financial data within the selected duration were systematically excluded.

Our methodological rigor was further enhanced using a paired sampling technique. This method entailed contrasting firms boasting positive cash flows or operational profits over five years with those manifesting a hostile trajectory. Such an approach was strategic, especially considering the overarching economic disturbances, notably the disruptions caused by the COVID-19 pandemic.

Guided by the pivotal work of [Piatt and Piatt \(2002\)](#), our study embraced a binary classification approach. A company's financial vulnerability was characterized by its incapacity to address fiscal responsibilities, significantly when marred by negative operating income, which is often a precursor to bankruptcy or insolvency. These financial tribulations could emanate from internal oversights or shifting external market dynamics, such as regulatory changes, amplified competition, or other externalities. Based on our criteria, firms with a negative trajectory in net operating income and operating cash flows for three successive years were identified as financially unstable.

In widening our analytical lens, [Nehrebecka \(2021\)](#) guided us to consider scenarios influenced by external shocks. While insights were gleaned from the financial outlines of dominant market entities as of March 2020, our methodology consciously bypassed firms from sectors profoundly affected by the reverberations of COVID-19. This crisis was crucial to ensure our analysis centered on companies whose trajectories were more influenced by intrinsic determinants than sweeping externalities like the pandemic.

For a thorough investigation, we meticulously extracted 36 financial ratios from the financial statements of these companies. The ratios were categorized into six classes, each fulfilling a particular analytical objective.

The initial course, profitability measures (Class 1), comprises a set of eight measures specifically formulated to assess the profitability and operational effectiveness of the organization. The statistics encompass net profit margin, asset turnover, return on assets, financial leverage, return on equity, gross profit margin, operating return on assets, and return on capital employed.

In the second class (Class 2), liquidity ratios, the emphasis is placed on evaluating a corporation's short-term financial well-being by examining three primary indicators: current ratio, quick ratio, and cash current liabilities ratio.

Cash flow ratios, specifically those discussed in (Class 3), encompass a collection of five ratios: cash flow from operations to sales, cash return on assets, cash return on equity, cash to income, and debt coverage ratio. These ratios offer valuable insights into the company's cash flow management and capacity to maintain long-term financial stability.

Class 4, which focuses on activity ratios, encompasses a comprehensive set of eight ratios. These ratios include the inventory turnover ratio, number of days in inventory, receivables turnover ratio, number of days in receivables, payable turnover ratio, number of days in payable, working capital turnover, and cash conversion cycle. These measurements provide insights into the effectiveness of the company's operational actions.

The fifth category (Class 5), valuation variables, encompasses seven variables essential for evaluating a company's market and investment appeal. The ratios encompass the paid-up value of shares, market price per share, basic earnings per share, price-earnings ratio, dividend payout ratio, cash dividend per share, and book value per share.

Finally, Class 6 encompasses four fundamental solvency ratios: the debt equity ratio, debt-to-assets ratio, debt-to-capital ratio, and interest cover ratio. These ratios provide valuable insights into the company's capacity to fulfil its long-term financial commitments.

The utilization of a systematic classification of financial ratios offers a methodical framework for assessing many aspects of firms' financial performance and stability. This process serves as the fundamental basis for our comprehensive examination, facilitating the derivation of significant conclusions and insights from the collected data.

Ensuring that models are resilient and generalizable is paramount in the expansive realm of modelling. A meticulous methodology was employed in building a predictive model for discerning between bankrupt and non-bankrupt entities.

Central to our strategy was the k-fold cross-validation technique. This method subdivided the dataset into 'k' distinct subsets or 'folds'. Each fold, in turn, was systematically designated for validation, with the model being trained on the remaining $k - 1$ folds. The rigorous iterative cycle was repeated k times, guaranteeing that every fold was used a validation set once. The culmination of these iterations was the computation of an aggregate performance metric, such as accuracy or F1-score, derived from the average across all the validation iterations. The advantage of this approach is its inclusivity: every data point is used both in a validation set once and in a training set $k - 1$ times. This comprehensive evaluation starkly contrasts the traditional train-test division, offering a broader and more nuanced assessment of model performance.

Our adherence to rigor continued. The selection of explanatory variables was executed with unparalleled precision, ensuring a foundation built on theoretical and empirical robustness. Critical variables were spotlighted by harnessing code capabilities to perform feature selection based on the relative mean differences between the bankrupt and non-bankrupt firms. This discernment was bolstered through Monte Carlo hypothesis testing, ensuring the statistical significance of the observed mean differences. Drawing inspiration from time-tested methodologies, such as the Altman Z-score, we included only specific independent variables in our model, aligning with the best practices from prior research.

Given the criticality of addressing overfitting, multiple strategies were employed. Regularization techniques, applicable to algorithms like logistic regression and SVMs, were integral in thwarting the development of overly intricate models. For decision trees, pruning was our tool of choice to reduce the model's complexity. Ensemble methods like random forests and gradient boosting became our bulwark against overfitting, offering the strength of multiple base estimators.

Furthermore, a grid search and cross-validation fusion guided us to the most suitable hyperparameters. Such meticulous tuning was vital not just for model performance but also as a bulwark against overfitting. Iterative models, specifically gradient boosting, were endowed with early stopping; a mechanism to halt training once no further improvements in validation error were observed. Recognizing the importance of data volume, we endeavored to incorporate a rich dataset, and for model simplification, only the most pertinent features were retained.

A holistic analysis was conducted, with hyperparameter selection being a pivotal aspect. This endeavor allowed the model complexity to be precisely calibrated against accuracy, ensuring impeccable tailoring to our specific objective. Our exploration was thorough, examining various hyperparameters across multiple machine learning algorithms. To further strengthen our model, techniques like Synthetic Minority Oversampling Technique (SMOTE) were used to address data imbalance issues, and PCA was employed for dimensionality reduction.

Overall, our commitment to a judicious approach, rooted in both empirical evidence and theoretical robustness, culminated in a predictive model of exceptional reliability, resilience, and relevance.

3.1. Support Vector Machine Model

Support Vector Machines (SVMs) are supervised machine learning algorithms used for classification analysis. SVMs function by locating the optimal hyperplane that distinguishes classes in a dataset. The goal of SVMs is to find a hyperplane that maximizes the margin between the two classes in a binary classification problem given a set of training data x_i, y_i , where x_i is the input vector and y_i is the corresponding binary output label (+1 or -1). The margin separates the nearest data points for each class and the hyperplane.

The equation of a hyperplane in a SVM is given by:

$$w^T x + b = 0, \quad (1)$$

where w is a vector perpendicular to the hyperplane, b is the bias term, and x is the input vector. The distance between a point x and the hyperplane is given by:

$$d = \frac{|w^T x + b|}{\|w\|}, \quad (2)$$

where the denominator is the norm of the vector w .

SVMs are used in quadratic optimization problems to identify the best hyperplane. The objective function is given by:

$$\min 1/2 \|w\|^2 + C \sum \xi_i, \quad (3)$$

subject to:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad (4)$$

where C regulates the trade-off between maximizing the margin and other factors and minimizing the classification. ξ_i is the slack variable that allows for some misclassification, and $y_i(w^T x_i + b)$ is the decision function that divides the two classes.

Lagrange multipliers can be used to solve the optimization problem, and the solution can be described in terms of the support vectors, which are the data points that lie closest to the hyperplane. One possible way to express the decision function is as follows:

$$f(x) = \text{sgn}\left(\sum \alpha_i y_i K(x_i, x) + b\right), \quad (5)$$

where α_i are the Lagrange multipliers, $K(x_i, x)$ is the kernel function that maps the input vectors to a higher-dimensional feature space, and sgn is the sign function that returns +1 or -1 depending on the sign of the argument. This function transfers the input vectors to a higher-dimensional feature space.

SVMs excel in high-dimensional spaces and are effective in datasets with abundant features, but they have limitations that need to be considered. Their training time increases significantly with larger datasets, and they do not provide direct probability estimates. The choice of the kernel function is also critical and requires expertise in the specific domain. Researchers and practitioners should carefully evaluate these factors when using SVMs in their applications.

3.2. Logistic Regression Model

Logistic regression is renowned for its unique capability to offer a transparent probabilistic interpretation of its outputs, enabling straightforward adjustments of decision thresholds. Its design inherently incorporates regularization techniques, which serve as a protective measure against overfitting. Moreover, logistic regression is adaptable when integrating new data, mainly through techniques like stochastic gradient descent. However, its limitation lies in assuming a linear decision boundary, rendering it unsuitable for handling non-linear complexities. This linear constraint often results in inferior performance compared to more advanced algorithms that handle intricate patterns.

Logistic regression (LR) a practical tool for predicting a company's financial failure using binary data. It estimates the likelihood of an outcome using one or more prediction factors. Bankruptcy prediction uses financial ratios or other criteria to predict bankruptcy. LR uses a logistic function to convert the linear combination of predictor variables into an outcome probability.

Logistic functions are S-shaped curves that reflect event probability from 0 to 1. The logistic equation is as follows:

$$P(Y = 1) = \frac{1}{(1 + e^{(-z)})}, \quad (6)$$

where $P(Y = 1)$ is the likelihood that a bankruptcy will occur, z is the linear combination of predictor variables, and e is the base of the natural logarithm.

An expression for the linear combination of predictor variables is:

$$z = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p, \quad (7)$$

where b_0 is the intercept and b_1, b_2, \dots, b_p are the coefficients of the predictor variables X_1, X_2, \dots, X_p .

After training, the model can predict firm bankruptcy based on financial measures and other indicators. A threshold probability can predict a company's bankruptcy. If the likelihood of bankruptcy exceeds 0.5, the firm will likely fail. The threshold can be changed to balance false positives and negatives.

3.3. Random Forest Model

Random forests are highly regarded for their versatility and ability to handle large datasets with high dimensionality effectively. One of their notable strengths is their robustness in managing missing values, ensuring that accurate predictions can still be made despite significant data gaps. However, these advantages come with inevitable trade-offs. The complex structure of the random forest model can result in slower evaluation speeds, which may be a consideration in time-sensitive applications. Additionally, while random forests are generally adaptable, there is a risk of overfitting when dealing with particularly noisy datasets. It is essential to carefully consider these factors when utilizing random forests in practical applications.

Random forest (RF) is a machine learning approach that may be utilized for classification and regression problems. It is a type of strategy known as an ensemble that builds several different decision trees and then adds up all their predictions to produce a single result.

Given below is a representation of the statistical equation for the RF model:

$$h(x) = \sum w_j I_{tree}(x, \theta_j), \quad (8)$$

for each split node j .

3.4. Naive Bayes Model

The naive Bayes classifier is lauded for its computational efficiency and straightforwardness, enabling effective parameter estimations even with limited training data and demonstrating aptitude in multi-class prediction scenarios. However, its foundational presumption of feature independence often misaligns with real-world complexities, potentially compromising its predictive accuracy. Furthermore, the model's propensity to assign a zero probability to previously unobserved categories during testing presents a noteworthy limitation in dynamic contexts.

The naive Bayes (NB) classifier is a probabilistic algorithm that employs Bayes' theorem to forecast the probability of a specific event based on prior knowledge of the conditions connected to that event. This probability prediction is accomplished using Bayes' theorem

to analyze the conditions associated with the event. In the context of bankruptcy prediction, NB can compute the probability that a firm will declare bankruptcy based on its financial ratios. This can be done by comparing the company in question to its peers.

The following statistical equations are utilized when applying the NB method:

$$P\left(\frac{Y}{X}\right) = P\left(\frac{X}{Y}\right) \times \frac{P(Y)}{P(X)}, \quad (9)$$

where Y represents the class variable (“bankrupt” or “non-bankrupt”), X represents the feature vector (“financial ratios”), $P(Y|X)$ represents the probability of Y given X , $P(X|Y)$ denotes the likelihood of X given Y , $P(Y)$ embodies the prior probability of Y , and $P(X)$ characterizes the prior probability of X .

3.5. Decision Tree Model

Decision trees are esteemed for their transparent structure, enabling lucid visualization and interpretation of the data-driven decision-making process. Their inherent versatility allows for nominal data preprocessing, adeptly managing a mix of numerical and categorical variables. Nonetheless, these models display a proclivity for overfitting in the absence of judicious tuning, mainly when training on constrained datasets. Their sensitivity to minuscule data perturbations can manifest in pronounced structural divergences, and empirically, their predictive prowess can be eclipsed by more sophisticated algorithms.

The decision tree (DT) model is a well-known approach in the field of machine learning that may be applied to problems involving classification and regression. The decision tree approach can be used in the context of bankruptcy prediction to classify enterprises as either bankrupt or not bankrupt, depending on the values of some predictor criteria. The decision tree determines this classification. The decision tree algorithm is based on recursively splitting the data based on the predictor variables to generate a decision tree-like structure that predicts the outcome variable. The core principle behind the algorithm is described in the next paragraph.

The equation that is used to train the decision tree model in the manner described below is as follows:

$$h(x) = T(x, \theta), \quad (10)$$

where $h(x)$ is the model’s prediction for data point x , T represents the decision tree structure, and θ encompasses the parameters, including the thresholds at each node and the chosen feature for each split. Each node might represent a specific financial indicator for bankruptcy prediction, such as the liquidity ratio, debt ratio, or profit margin. The thresholds at each node are optimized to discern potential bankruptcies. For instance, a node might inquire if the liquidity ratio is below a specific critical value, directing firms with lower liquidity to a branch as more indicative of bankruptcy. The leaf nodes represent the final predictions (either ‘bankrupt’ or ‘solvent’).

3.6. Adaptive Boosting Model

The AdaBoost algorithm is distinguished by its adeptness at delineating complex decision boundaries, stemming from its capacity to amalgamate multiple weak predictors into a potent predictive ensemble, inherently resisting overfitting. Nevertheless, its susceptibility to noisy data and anomalies presents potential pitfalls in accuracy. Furthermore, achieving optimal performance necessitates rigorous hyper-parameter tuning, potentially elongating the model’s training duration.

Adaptive boosting (AdaBoost) is a boosting algorithm that combines numerous weak learners into one powerful learner. In AdaBoost, weak learners are often decision trees that only have one possible branch, and these trees are referred to as decision stumps. This approach functions by repeatedly training weak learners on the dataset, primarily focusing on the examples of misclassification made by earlier iterations of weak learners. The result is a weighted total of the weak learners, each given a weight based on their performance in

the previous tests. The AdaBoost method is a robust one that has been demonstrated to be successful in a wide variety of different situations.

Training the AdaBoost model can be accomplished using the following equation:

$$f(x) = \text{sign} \sum \alpha_j h_j(x), \quad (11)$$

For bankruptcy prediction, the sign will determine whether a firm is predicted to go bankrupt (negative) or not (positive).

3.7. Gradient Boosting Model

Gradient boosting machines (GBMs) are heralded in the machine learning domain for their exemplary precision across diverse applications. Exhibiting profound adaptability, they seamlessly accommodate an assortment of predictor variables and facilitate customization concerning various loss functions. Nonetheless, GBMs are full of challenges. A pertinent concern is their predisposition towards overfitting, particularly in the absence of judicious hyper-parameter optimization. Furthermore, the training phase, characterized by its sequential nature, can be computationally onerous. Additionally, data noise can compromise the model's efficacy in some scenarios. It is imperative to recognize that these general observations on GBMs' strengths and potential pitfalls should be contextualized within any given application's specifics and associated dataset.

The gradient boosting (GB) algorithm is an example of an ensemble learning algorithm. It takes several weak learners and combines them into a single powerful learner. It constructs the model in stages, each consisting of the following steps: first, a weak learner is trained to minimize the loss function using the gradient descent method; next, the model is updated by adding this weak learner; and finally, the model is evaluated. The procedure is repeated iteratively, with each consecutive learner concentrating on fixing the mistakes committed by the prior learner.

The gradient boosting model should be trained using the equation below:

$$f(x) = \sum \gamma_j h_j(x) \quad (12)$$

Here, $h_j(x)$ denotes the weak learner, and γ_j represents the associated weight. This equation underscores the cumulative nature of GB, where each component learner contributes to the overarching predictive model. Given the volatility and dynamics of Pakistan's corporate sector, utilizing GB's strengths could offer more precise insights into impending bankruptcies, aiding stakeholders in preemptive decision-making.

4. Results of the Comparative Analysis

4.1. Descriptive Analysis

Based on the descriptive statistics (e.g., mean and standard deviation), the accompanying tables (Tables 1 and 2) reflect the financial ratios under examination. The failing businesses' had a significantly lower return on assets (ROA) compared to the successful businesses, suggesting that the failed businesses needed to be more effective at turning their assets into earnings. Companies with a high return on investment (ROI) outperformed those with a low ROI. The operating return on assets (OROA) of the failing enterprises was much lower than that of the successful businesses, showing that the failing businesses needed to be more efficient at converting their assets into profits. The successful enterprises had significantly higher OROAs than the unsuccessful ones.

The debt coverage ratios of the failed enterprises were much lower than those of the successful firms, demonstrating that they could not pay their debts. The debt coverage ratios of successful businesses were far more significant than those of the failing businesses. The successful businesses had a much lower debt-to-assets ratio than the failed businesses. The cash return on assets (CROA) of the failing enterprises was significantly lower than that of successful ones, indicating that they were less efficient at converting their assets into cash flow. Profitable businesses have much greater CROAs than those that fail. The businesses

that failed had considerably lower quick ratios than the successful ones, demonstrating a reduced capacity to satisfy short-term financial obligations.

Table 1. Descriptive statistics for bankrupt firms.

Statistics	ROA	Operating ROA	Debt to Assets	Cash ROA	Profit per Share	Debt Coverage	Asset Turnover	Quick Ratio
Mean	−8.21	−4.93	0.84	−1.70	−11.21	−0.08	0.70	0.27
Std. Dev.	12.33	11.87	0.69	11.03	21.14	0.78	0.60	0.39
Min	−160.30	−143.08	0.04	−71.34	−176.60	−14.58	0.00	0.00
Max	0.00	11.52	5.77	58.64	0.00	0.48	4.23	3.01

Source: authors' estimations.

Table 2. Descriptive statistics for non-bankrupt firms.

Statistics	ROA	Operating ROA	Debt to Assets	Cash ROA	Profit per Share	Debt Coverage	Asset Turnover	Quick Ratio
Mean	8.10	12.80	0.54	7.32	23.50	0.31	1.14	0.72
Std. Dev.	11.85	13.06	0.27	13.71	57.87	0.29	0.63	0.96
Min	−20.48	−16.00	0.04	−49.47	0.00	−0.14	0.01	0.00
Max	337.92	343.05	5.54	221.18	811.55	2.78	4.82	17.32

Source: authors' estimations.

In contrast to the enterprises that ultimately failed, the successful ones had a much higher quick liquidity ratio. The strength of this study resides in the fact that it provides empirical support for the hypothesis that specific financial indicators are related to a company's propensity to file for bankruptcy. According to the study, there were significant differences between the firms that ultimately failed and those that did not with regards to their primary financial measures. These differences could be early indicators of imminent financial difficulties.

4.2. Correlation Matrix

Table 3 presents the association between the grouping variable and the final selected ratios, e.g., cash return on assets (CROA), Debt coverage ratio, quick ratio, asset turnover, return on assets (ROA), operational return on assets (OROA), debt-to-assets ratio, and profits per share. The correlation coefficients range from −1 to 1, with −1 being a perfect negative correlation, 0 denoting no association, and 1 reflecting a perfect positive correlation.

Table 3. Correlation matrix between financial ratios of bankrupt and non-bankrupt firms.

Ratios	ROA	Operating ROA	Debt to Assets	Cash ROA	Profit per Share	Debt Coverage	Asset Turnover	Quick Ratio	
ROA	1	0.763	−0.143	−0.048	0.384	0.658	0.090	0.025	
Operating ROA	0.777	1	−0.113	−0.052	0.352	0.667	0.150	0.024	
Debt to assets	0.191	0.199	1	−0.170	0.065	0.068	−0.081	−0.274	
Cash ROA	0.597	0.613	−0.019	1	−0.053	−0.077	−0.079	0.141	
Profit per share	0.317	0.347	−0.002	0.308	1	0.167	−0.119	−0.038	
Debt coverage	0.485	0.500	−0.445	0.357	0.279	1	0.053	−0.119	
Asset turnover	0.166	0.235	0.156	0.146	0.214	0.066	1	0.101	
Quick ratio	0.120	0.105	−0.398	0.098	0.087	0.544	0.0003	1	

Source: authors' estimations.

The ROA (0.366) and operational return on assets (0.373) had the most extensive positive correlations in the bankrupt group, indicating that these variables may predict bankruptcy. These correlations were highest for the bankrupt group. Furthermore, there was a strong correlation between the debt coverage ratio (0.348) and basic earnings per share (0.318). The debt-to-assets ratio displayed a slightest positive connection (0.852) among all the variables. The CROA (0.387), the only variable with a stronger positive correlation in

the bankrupt group than the non-bankrupt group, and debt-to-assets ratio (0.044), which also had the weakest correlation with the non-bankrupt group, were the variables that had the weakest correlations with the grouping (bankruptcy) variable. Additionally, the groups not filing for bankruptcy had the weakest relationships with both characteristics.

4.3. Results of the Machine Learning Models

The following table (Table 4) presents the performance of various classifiers of machine learning models on the given dataset. A measure of the classifier's performance is given in each table column, and each row in the table represents a separate classifier. The applicable name of the classification model is shown in the first column. The classifier represents the number of features that the model utilizes in the next column. The first block selects the top 10 columns from the training and test sets. It assigns them to the corresponding variables, X_train_10 and X_test_10. The second block chooses the first 50 columns from the training and test sets and puts them into the corresponding variables X_train_50 and X_test_50. The third block chooses the first 70 columns from the training and test sets and assigns them to the corresponding variables X_train_70 and X_test_70. Reducing the dimensionality of the data and enhancing the performance of the machine learning models are the goals of choosing a subset of features. The accuracy is the proportion of instances that are accurately divided into bankrupt and non-bankrupt enterprises from all the instances in the dataset. The F1-measure is the harmonic mean of precision and recall, which measures the balance between precision and recall. Recall is a different metric that measures the proportion of positive events in the dataset that were accurately identified as positive. The percentage of correctly identified predicted positive cases among all the predicted positive instances in the dataset is shown in the last column of the table, labeled 'Precision'.

Table 4. Model-wise classification.

Model	Classifier	Accuracy	F1-Measure	Recall	Precision
SVM	SVM_10	89%	89%	87%	91%
	SVM_50	99%	99%	97%	100%
	SVM_70	99%	99%	97%	100%
LR	LR_10	92%	92%	92%	91%
	LR_50	99%	99%	99%	100%
	LR_70	99%	99%	99%	100%
RF	RF_10	100%	100%	100%	100%
	RF_50	100%	100%	100%	100%
	RF_70	100%	100%	100%	100%
NB	NB_10	58%	70%	96%	55%
	NB_50	70%	76%	96%	64%
	NB_70	70%	76%	96%	64%
DT	DT_10	99%	99%	99%	100%
	DT_50	100%	100%	100%	100%
	DT_70	100%	100%	100%	100%
AdaBoost	AdaBoost_10	100%	100%	100%	100%
	AdaBoost_50	100%	100%	100%	100%
	AdaBoost_70	100%	100%	100%	100%
GBT	GBT_10	100%	100%	99%	100%
	GBT_50	100%	100%	100%	100%
	GBT_70	100%	100%	100%	100%

Source: authors' estimations.

Table 4 shows that all the models achieved a high level of accuracy, with most models achieving 100% accuracy. The models with the highest accuracy were the decision tree, AdaBoost, and gradient boosting ones. Depending on how many features were utilized, the SVM and logistic regression models also did well, obtaining accuracy rates of 89 to

99 percent. The accuracy of the naive Bayes model, which varied from 58 to 70 percent depending on the number of characteristics utilized, was the lowest. The models' overall high precisions, recalls, and F1-measure scores show that they performed exceptionally well at correctly identifying the bankrupt and non-bankrupt enterprises in the dataset.

The receiver operating characteristics (ROC) curve (Figure 1) reveals the sensitivity–specificity balance at different thresholds. The ROC analysis can assess a model's discriminatory capacity and determine the ideal operating point based on the user's goals and tolerance for false positives and negatives. Most models have a 100% accuracy, indicating exceptional precision. The decision tree, AdaBoost, and gradient boosting models were the most accurate. The SVM and logistic regression models performed well depending on the feature count. The naive Bayes was less accurate than the other models.

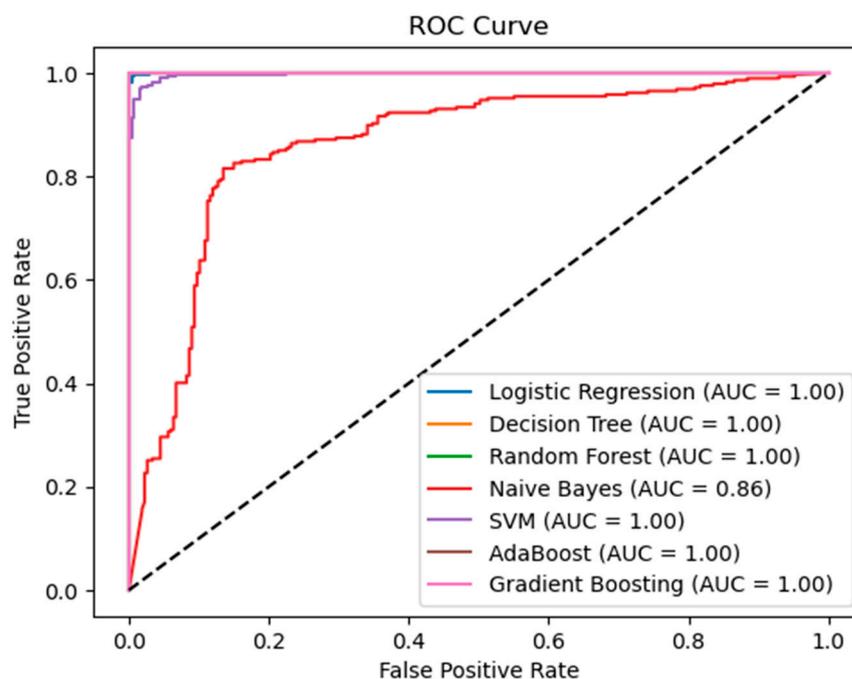


Figure 1. Receiver operating characteristics (ROC) curve.

5. Discussion

This study offers a comprehensive evaluation of several machine learning models for bankruptcy prediction, and the outcomes garnered pave the way for profound insights. The results of this study involved consistently high levels of accuracy across models, elucidating their potential for effective bankruptcy prediction. Distinctly, the decision tree, AdaBoost, and gradient boosting classifiers reached an impressive pinnacle, with an accuracy of 100%. Such a feat emphasizes their robustness and sets a benchmark for future studies in similar domains.

While these models achieved perfect accuracy, the SVM and logistic regression models also garnered commendable results. Their accuracy fluctuated between 89% and 99%, depending on the feature count employed. It is noteworthy that as the number of features increased, the accuracy of these models tended to edge closer to perfection. This trend underscores the pivotal role of feature selection, corroborating earlier findings by [Shetty et al. \(2022\)](#) that advocate for the judicious choice of features to optimize model performance.

In stark contrast, though consistent across different feature counts, the naive Bayes model's performance remained suboptimal compared to its counterparts. Its accuracy, ranging from 58% to 70%, raises questions regarding its suitability for such datasets, emphasizing the need for further exploration. However, it is imperative to note that the high recall scores for the naive Bayes model suggest its strength in identifying bankrupt enterprises, even if its overall accuracy is less.

Comparatively, traditional models like Altman's Z-Score and Taffler's models have been widely recognized in the financial realm (Altman 1968). For instance, Altman's model demonstrated an 82–94% accuracy rate. However, our study reveals that specific machine learning models can surpass even these commendable benchmarks. One primary distinction between the two approaches is their adaptability. Due to their inherent design, machine learning models are inherently adaptable to changing financial landscapes. In contrast, while transparent and interpretable, traditional models like the Z-Score model possess fixed characters and require periodic modifications, as evidenced by Altman's subsequent revisions.

Our study bridges the gap between theory and practice, drawing from empirical results to provide stakeholders with actionable insights. The high accuracy rates, especially among models like decision trees, AdaBoost, and gradient boosting, signify not only the potential of these tools for reliable bankruptcy prediction but also emphasize the importance of machine learning in contemporary financial forecasting. Furthermore, juxtaposed against traditional models, our findings underscore the evolving nature of bankruptcy prediction tools and the need for continuous innovation in this domain.

Conclusively, while the tested models demonstrated a promising avenue for bankruptcy prediction, it is vital to approach their deployment with a nuanced understanding of their strengths and limitations. Further research, building upon our findings, can aid in honing these tools for even more precise predictions in the dynamic landscape of Pakistani enterprises.

6. Conclusions

This study aimed to evaluate Pakistani enterprises using machine learning techniques to predict their financial difficulties and likelihood of bankruptcy. The results of this study indicate that several financial ratios, such as return on assets, operating return on assets, debt coverage ratio, asset turnover, earnings per share, debt-to-assets ratio, Cash return on assets, and quick ratio, can be used to predict whether or not a company will file for bankruptcy.

Overall, this study on bankruptcy prediction using machine learning techniques and financial ratios has the potential to contribute to the field of financial risk management by improving risk assessment, serving as an early warning system, enhancing risk management practices, informing regulatory considerations, and inspiring future research and innovation.

According to the empirical evidence supporting the study, financial ratios can be used to predict insolvency in Pakistani enterprises. The machine learning models' results show the best options for predicting financial distress and insolvency in Pakistani enterprises. This study also underlines the need to use financial indicators to predict bankruptcy, which can help financial analysts, investors, and regulators make more informed decisions.

The results of these models are essential to financial analysts, investors, and stakeholders who want accurate bankruptcy predictions. The decision tree, AdaBoost, and gradient boosting models performed well, obtaining a 100% accuracy. The SVM and logistic regression models showed exceptional flexibility in feature selection settings, with 89 to 99% accuracy rates depending on the selected characteristics. The Naive Bayes model performed poorly, with 58% to 70% accuracy. However, its utility for specific feature sets must be considered. These models' high precision, recall, and F1-measure scores show their ability to distinguish bankrupt from non-bankrupt enterprises, making them useful for industry experts who need accurate and fast bankruptcy identification.

The results of this study may have significant effects on Pakistan's non-financial sector. Policymakers and regulatory authorities may find the study's insights helpful in creating and working to achieve effective frameworks and laws to reduce systemic risks in the financial sector. By identifying the financial ratios that contribute to bankruptcy prediction, regulators can establish thresholds or guidelines for monitoring the financial health of companies and enforcing appropriate measures when necessary. Financial institutions can

utilize bankruptcy prediction models to proactively manage their exposure to potentially risky borrowers, leading to a more resilient banking system. Additionally, businesses can use these findings to monitor their financial health and make the required modifications to prevent financial bankruptcy. This study also emphasizes the value of applying machine learning techniques to bankruptcy forecasting, which can help to increase prediction accuracy and lower the risks associated with financial investments.

Future studies may need to be revised to select and implement machine learning algorithms used in the comparative analysis. The choice of algorithms may influence the performance and outcomes of the bankruptcy prediction models. The models' efficacies and capacities to forecast bankruptcy in Pakistani enterprises may be constrained by biases in the data utilized for training and evaluation. The findings and conclusions of this study may be specific to the context of Pakistani companies, limiting their generalizability to other regions or industries.

Author Contributions: Conceptualization, H.R. and I.A.; methodology, H.R.; software, H.R.; validation, H.R., I.A. and D.M.; formal analysis, H.R.; writing—original draft preparation, H.R.; writing—review and editing, D.M.; visualization, H.R.; supervision, I.A.; funding acquisition, D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is publicly available at [https://www.sbp.org.pk/departments/stats/FSA\(Non\).pdf](https://www.sbp.org.pk/departments/stats/FSA(Non).pdf) (accessed on 17 September 2023).

Acknowledgments: We appreciate the support of the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

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

References

- Altman, Edward I. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance* 23: 589–609. [CrossRef]
- Altman, Edward I., Robert G. Haldeman, and P. Narayanan. 1977. ZETATM Analysis A New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking & Finance* 1: 29–54. [CrossRef]
- Aziz, M. Adnan, and Humayon A. Dar. 2006. Predicting Corporate Bankruptcy: Where We Stand? *Corporate Governance* 6: 18–33. [CrossRef]
- Barboza, Flavio, Herbert Kimura, and Edward Altman. 2017. Machine Learning Models and Bankruptcy Prediction. *Expert Systems with Applications* 83: 405–17. [CrossRef]
- Beaver, William H. 1966. Financial Ratios as Predictors of Failure. *Journal of Accounting Research* 4: 71–111. [CrossRef]
- Bühlmann, Peter, and Torsten Hothorn. 2007. Boosting Algorithms: Regularization, Prediction and Model Fitting. *Statistical Science* 22: 477–505. [CrossRef]
- Ciampi, Francesco. 2015. Corporate Governance Characteristics and Default Prediction Modeling for Small Enterprises. An Empirical Analysis of Italian Firms. *Journal of Business Research* 68: 1012–25. [CrossRef]
- Deakin, Edward B. 1972. A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research* 10: 167. [CrossRef]
- Edmister, Robert O. 1972. An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. *Journal of Financial and Quantitative Analysis* 7: 1477–93. [CrossRef]
- Fan, Jianqing, and Runze Li. 2001. Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association* 96: 1348–60. [CrossRef]
- Grice, John Stephen, and Robert W. Ingram. 2001. Tests of the Generalizability of Altman's Bankruptcy Prediction Model. *Journal of Business Research* 54: 53–61. [CrossRef]
- Hosaka, Tadaaki. 2019. Bankruptcy Prediction Using Imaged Financial Ratios and Convolutional Neural Networks. *Expert Systems with Applications* 117: 287–99. [CrossRef]
- Islam, Ummana, Nishat Anjum, Abdul Kadar Muhammad Masum, and Abm Yasir Arafat. 2022. Forecasting of Bank Performance Using Hybrid Machine Learning Techniques. Paper presented at 2022 International Conference on Innovations in Science, Engineering and Technology, Chattogram, Bangladesh, February 25–28; pp. 433–38. [CrossRef]

- Jardin, Philippe Du. 2015. Bankruptcy Prediction Using Terminal Failure Processes. *European Journal of Operational Research* 242: 286–303. [\[CrossRef\]](#)
- Jo, Hongkyu, Ingoo Han, and Hoonyoung Lee. 1997. Bankruptcy Prediction Using Case-Based Reasoning, Neural Networks, and Discriminant Analysis. *Expert Systems with Applications* 13: 97–108. [\[CrossRef\]](#)
- Kainth, Akarsh, and Ranik Raaen Wahlstrøm. 2021. Do IFRS Promote Transparency? Evidence from the Bankruptcy Prediction of Privately Held Swedish and Norwegian Companies. *Journal of Risk and Financial Management* 14: 123. [\[CrossRef\]](#)
- Kanapickienė, Rasa, Tomas Kanapickas, and Audrius Nečiūnas. 2023. Bankruptcy Prediction for Micro and Small Enterprises Using Financial, Non-Financial, Business Sector and Macroeconomic Variables: The Case of the Lithuanian Construction Sector. *Risks* 11: 97. [\[CrossRef\]](#)
- Karas, Michal, and Mária Režňáková. 2014. To What Degree Is the Accuracy of a Bankruptcy Prediction Model Affected by the Environment? The Case of the Baltic States and the Czech Republic. *Procedia—Social and Behavioral Sciences* 156: 564–68. [\[CrossRef\]](#)
- Khan, Shahzad, Sahibzada Tasleem Rasool, and Syed Imran Ahmed. 2021. Role of Cardiac Biomarkers in COVID-19: What Recent Investigations Tell Us? *Current Problems in Cardiology* 46: 100842. [\[CrossRef\]](#)
- Kitowski, Jerzy, Anna Kowal-Pawul, and Wojciech Lichota. 2022. Identifying Symptoms of Bankruptcy Risk Based on Bankruptcy Prediction Models—A Case Study of Poland. *Sustainability* 14: 1416. [\[CrossRef\]](#)
- Kliestik, Tomas, Katarina Valaskova, George Lazaroiu, Maria Kovacova, and Jaromir Vrbka. 2020. Remaining Financially Healthy and Competitive: The Role of Financial Predictors. *Journal of Competitiveness* 12: 74–92. [\[CrossRef\]](#)
- Laitinen, Erkki K. 1991. Financial Ratios and Different Failure Processes. *Journal of Business Finance & Accounting* 18: 649–73. [\[CrossRef\]](#)
- Lennox, Clive S. 1999. The Accuracy and Incremental Information Content of Audit Reports in Predicting Bankruptcy. *Journal of Business Finance & Accounting* 26: 757–78. [\[CrossRef\]](#)
- Li, Guofa, Yifan Yang, Tingru Zhang, Xingda Qu, Dongpu Cao, Bo Cheng, and Keqiang Li. 2021. Risk Assessment Based Collision Avoidance Decision-Making for Autonomous Vehicles in Multi-Scenarios. *Transportation Research Part C: Emerging Technologies* 122: 102820. [\[CrossRef\]](#)
- Liang, Deron, Chia Chi Lu, Chih Fong Tsai, and Guan An Shih. 2016. Financial Ratios and Corporate Governance Indicators in Bankruptcy Prediction: A Comprehensive Study. *European Journal of Operational Research* 252: 561–72. [\[CrossRef\]](#)
- Lombardo, Gianfranco, Mattia Pellegrino, George Adosoglou, Stefano Cagnoni, Panos M. Pardalos, and Agostino Poggi. 2022. Machine Learning for Bankruptcy Prediction in the American Stock Market: Dataset and Benchmarks. *Future Internet* 14: 244. [\[CrossRef\]](#)
- Mai, Feng, Shaonan Tian, Chihoon Lee, and Ling Ma. 2019. Deep Learning Models for Bankruptcy Prediction Using Textual Disclosures. *European Journal of Operational Research* 274: 743–58. [\[CrossRef\]](#)
- Mood, Carina. 2010. Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do about It. *European Sociological Review* 26: 67–82. [\[CrossRef\]](#)
- Nehrebecka, Natalia. 2021. COVID-19: Stress-Testing Non-Financial Companies: A Macroprudential Perspective. The Experience of Poland. *Eurasian Economic Review* 11: 283–319. [\[CrossRef\]](#)
- Ohlson, James A. 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research* 18: 109. [\[CrossRef\]](#)
- Papana, Angeliki, and Anastasia Spyridou. 2020. Bankruptcy Prediction: The Case of the Greek Market. *Forecasting* 2: 505–25. [\[CrossRef\]](#)
- Patel, Charmi, Dominic Pilon, Deepshikhar Gupta, Laura Morrison, Marie Hélène Lafeuille, Patrick Lefebvre, and Carmela Benson. 2022. National and Regional Description of Healthcare Measures among Adult Medicaid Beneficiaries with Schizophrenia within the United States. *Journal of Medical Economics* 25: 792–807. [\[CrossRef\]](#) [\[PubMed\]](#)
- Piatt, Harlan D., and Marjorie B. Piatt. 2002. Predicting Corporate Financial Distress: Reflections on Choice-Based Sample Bias. *Journal of Economics and Finance* 26: 184–99. [\[CrossRef\]](#)
- Pompe, Paul P. M., and Jan Bilderbeek. 2005. The Prediction of Bankruptcy of Small- and Medium-Sized Industrial Firms. *Journal of Business Venturing* 20: 847–68. [\[CrossRef\]](#)
- Qu, Yi, Pei Quan, Minglong Lei, and Yong Shi. 2019. Review of Bankruptcy Prediction Using Machine Learning and Deep Learning Techniques. *Procedia Computer Science* 162: 895–99. [\[CrossRef\]](#)
- Shetty, Shekar, Mohamed Musa, and Xavier Brédart. 2022. Bankruptcy Prediction Using Machine Learning Techniques. *Journal of Risk and Financial Management* 15: 35. [\[CrossRef\]](#)
- Shi, Yin, and Xiaoni Li. 2019. An Overview of Bankruptcy Prediction Models for Corporate Firms: A Systematic Literature Review. *Intangible Capital* 15: 114–27. [\[CrossRef\]](#)
- Shin, Kyung Shik, and Yong Joo Lee. 2002. A Genetic Algorithm Application in Bankruptcy Prediction Modeling. *Expert Systems with Applications* 23: 321–28. [\[CrossRef\]](#)
- Tang, Tseng Chung, and Li Chiu Chi. 2005. Predicting Multilateral Trade Credit Risks: Comparisons of Logit and Fuzzy Logic Models Using ROC Curve Analysis. *Expert Systems with Applications* 28: 547–56. [\[CrossRef\]](#)
- Tian, Shaonan, Yan Yu, and Hui Guo. 2015. Variable Selection and Corporate Bankruptcy Forecasts. *Journal of Banking & Finance* 52: 89–100. [\[CrossRef\]](#)
- Tunio, Fayaz Hussain, Yi Ding, Amad Nabi Agha, Kinza Agha, and Hafeez Ur Rehman Zubair Panhwar. 2021. Financial Distress Prediction Using Adaboost and Bagging in Pakistan Stock Exchange. *The Journal of Asian Finance, Economics and Business* 8: 665–73. [\[CrossRef\]](#)

Wang, Na, Juanwen Chen, Mankin Tai, and Jingyuan Zhang. 2021. Blended Learning for Chinese University EFL Learners: Learning Environment and Learner Perceptions. *Computer Assisted Language Learning* 34: 297–323. [[CrossRef](#)]

Zmijewski, Mark E. 1984. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research* 22: 59–82. [[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.