



Credit Scorecards & Forecasting Default Events – A Novel Story of Non-financial Listed Companies in Pakistan

Jahanzaib Alvi¹ · Imtiaz Arif¹

Accepted: 2 September 2024

© The Author(s), under exclusive licence to Springer Japan KK, part of Springer Nature 2024

Abstract

This study innovates in credit default prediction in Pakistan by developing, calibrating, and recalibrating machine learning-based credit scorecards for non-financial listed firms, leveraging extensive financial ratio analysis. This study innovates in credit default prediction in Pakistan by developing, calibrating, and recalibrating machine learning-based credit scorecards for non-financial listed firms, leveraging extensive financial ratio analysis. Identifies 12 key financial ratios out of 71 remained vital for default prediction, with Random Forest and Artificial Neural Networks leading in scorecard performance. This marks Pakistan's first detailed scorecard approach as a potential alternative to traditional banking systems. Offers advanced risk assessment tools (credit scorecards) for improved credit risk management, aiding policymakers and finance professionals in decision-making. This research distinguishes itself through a detailed longitudinal study of non-financial Pakistani firms and a comprehensive evaluation of machine learning algorithms for default prediction. By exploiting various financial ratios to develop scorecards (an alternative of Internal Ratings-based – IRB System), it offers new insights into risk evaluation and significantly advances financial risk management. Acknowledging data limitations and variable exclusions, it sets the stage for further exploration of credit risk environment in context of Pakistan.

Keywords Credit scorecard development · Machine learning · Financial ratio analysis · Credit risk management · Non-financial firms

✉ Jahanzaib Alvi
Jahanzaib.alvi@iqra.edu.pk

Imtiaz Arif
arif.i@iuk.edu.pk

¹ Department of Business Administration, IQRA University, Karachi, Pakistan

1 Introduction

Bankruptcy risk assessment is crucial for firms, and advanced statistical models aid in predicting business failures. Credit risk management has evolved with financial innovations, necessitating the development of internal rating models aligned with regulatory guidelines (King et al., 2024; Hasnain et al., 2022). Credit risk assessment and default prediction are interconnected, with bankruptcy being a legal process sought by debtors unable to repay debts (Alhammadi et al., 2024). Machine learning techniques and statistical modeling enhance default prediction accuracy (Rahmani et al., 2024). Business failures impact the economy, prompting the need for effective default prediction models (Alvi et al., 2024). Loan portfolio models and the Internal Rating-Based approach assist in managing loan portfolios for commercial banks. The Basel III Accord and post-financial crisis challenges have led to the development of new models considering financial and macroeconomic indicators (Altman et al., 2004; Beaver et al., 2005).

Default prediction models incorporating financial and non-financial information play a crucial role in assessing credit risk for banks and financial institutions (Balasubramanian et al., 2019). Improving classification performance in credit scoring is essential for profitability (Malhotra et al., 2020). The failure of companies like WorldCom and Lehman Brothers had a significant negative impact on the economy (Kim et al., 2022). Various models have been developed to predict business failures, but contextual and methodological limitations exist (Jones 2023). Financial distress prediction models based on machine learning offer early indications of potential failure (Kato & Nakamura, 2024). Machine learning methods accurately predict asset prices and credit ratings (credit ratings means credit scoring). Developing quantitative models based on past financial data is a cost-effective solution for default prediction (Baek, 2023).

Financial distress is a state of financial difficulties faced by companies due to various factors. It involves different phases of distress, ranging from mild to severe (Florez-Lopez & Ramon-Jeronimo, 2015). The objective is to overcome the difficulties and avoid bankruptcy by analyzing the root cause and implementing a recovery plan. Monitoring and adjusting the plan as needed is crucial (Abellán & Castellano, 2017). Managing financial distress requires careful attention to the company's financial condition, performance, and actions taken at different stages (Outecheva, 2007). Corporate failure involves financial distress leading to bankruptcy, which can be preceded by warning signs (Altman, 1968). Liquidity disorder is a common cause, and insolvency occurs when the company cannot meet its obligations. Companies may recover through debt restructuring. The severity of distress determines appropriate actions (Jarrow, 2005).

Financial distress has wide-ranging effects beyond the distressed firm itself, impacting investors, competitors, and other entities the consequences can be categorized into four areas: costs borne by the defaulted firm, costs borne by investors, compensation from other entities, and costs incurred by creditors and stakeholders (Branch, 2002). Bankruptcy costs can be measured as the actual cost to the company, the impact on customers and stakeholders, compensation required

from other companies, and the overall economic impact. Financial distress can disrupt supply chains, cause job losses, and harm the economy, as seen in cases like Enron and WorldCom. Internal factors like inefficiencies and limited resources, particularly in small businesses, contribute to financial distress (Graham and Harvey, 2001.). It is crucial for companies to identify causes and take proactive measures to avoid financial ruin (Weitzel & Johnson, 1989). Factors contributing to financial distress include insufficient funding, exchange rate fluctuations, unstable management, high interest rates, declining demand, lack of regulations, competitive innovations, and poor financial practices (Jahur & Quadir, 2012).

The primary objective of this research is to bridge a significant gap in existing literature by providing an unprecedented, detailed analysis of credit risk assessment for non-financial listed companies in Pakistan, a topic seldom explored with such depth in this region as best of our knowledge. Leveraging advanced machine learning algorithms, this study meticulously selects efficient features (financial ratios) that closely mirror reality, ensuring the predictive accuracy of defaults. It rigorously tests these ratios through scenario or sensitivity analysis (proposed default definitions in methodology), evaluates the performance of various machine learning classifiers for superior accuracy and predictive power, and pioneers the development of a bespoke scorecard system for each company, this what we claim as the most novel job in context of Pakistan (Meaning deriving probability of default (PDs) events by Machine Learning algorithms – Black box Model) supported by Sigrist et al., (2023). This innovation allows for the creation of tailored rating scorecards using proposed financial ratios and weights (the beauty of this article is identifying the assigned weights along with back testing), further validated through comprehensive stress testing for model calibration and recalibration as suggested by Bequé et al., (2017). A cornerstone of this research is its establishment of a tested and efficient internal rating-based (IRB) system, marking a novel contribution to the field (this is the most genuine and authentic practical contribution of the research), any stakeholder may use our model and get real time PDs which will be very helpful for him which making investment decisions. To the best of our knowledge, no such detailed and holistic research effort has been undertaken in Pakistan before, underlining the uniqueness and potential impact of this study in advancing practical risk assessment models and providing valuable tools for investors and industries to effectively manage credit risk.

The problem addressed in this research is the lack of accurate credit default prediction models for non-financial listed companies in Pakistan. The existing deficiency in early warning signs in the non-financial sector leads to insolvency and financial distress, negatively impacting the economy and contributing to higher unemployment rates (Aggarwal et al., 2020). To fill this gap, the study aims to develop, validate, calibrate and recalibrate credit default prediction models using machine learning algorithms (Sun et al., 2021). These models will enable stakeholders, including decision-makers, management, lenders, and creditors, to anticipate and prevent business failures effectively (Chopra & Bhilare, 2018). By providing insights to regulatory bodies, commercial banks, and financial institutions, the research aims to enhance credit risk assessment and management processes,

ultimately leading to improved financial stability and reduced risk in the financial sector (Farooq et al., 2018).

This research introduces four core novelties in predicting financial defaults in Pakistan. Firstly, it pinpoints the most predictive feature for defaults. Secondly, the study showcases an algorithm with unparalleled accuracy in the realm of default prediction models. Thirdly, it underscores the profound impact of the assumptions used in labeling firms as defaulting. Lastly, making scorecards using best features extracted by ML algorithms, and aligning them with BASEL III guideline for model calibration. Using a simulation based on Basel III guidelines, the research aids in assessing the risk of non-financial listed firms in Pakistan, subsequently guiding informed loan decisions. This proposed model is instrumental for finance managers in assessing customer financial health, supporting profitable investments, and bolstering job security. By ensuring timely default predictions, stakeholders can preempt business failures. Overall, this study serves as a beacon for crafting reliable default prediction models, benefiting a wide spectrum of financial institutions and regulatory bodies while emphasizing the unique contributions of the research.

This study faces limitations in its scope, notably in not testing other advanced machine learning algorithms like CatBoost, XGBoost, and LightGBM due to time and resource constraints. Furthermore, it does not incorporate macroeconomic, corporate governance, and ESG indicators, nor additional financial ratios that could potentially increase the model's predictive power, primarily because of data accessibility issues. The research also omits analysis of listed financial institutions, unlisted companies, and SMEs in Pakistan, attributed to challenges in obtaining relevant data.

The study is divided into four sections. Section 2 explores the literature gap related to AI-based prediction models and justifies the research objectives. Section 3 explains the research methodology, including the tools and techniques used to achieve the objectives. Section 4 presents the results obtained from applying the methodology, and Sect. 5 concludes the article.

2 Literature Review

The ex-ante theory of financial distress involves predicting and analyzing the factors and decisions preceding a company's financial difficulties. It emphasizes assessing risks and returns before distress occurs, focusing on early indicators and preemptive actions (Farooq et al., 2018).

Beaver (1966) introduced discriminant theory, which quickly gained popularity for predicting failure. Altman (1968) made significant contributions by introducing the Z-score and testing model as an effective measure of assessing corporate solvency. He also highlighted the usefulness of logit and probit models in the 1980s and 1990s, which relaxed the assumptions of Multivariate discriminant analysis (MDA). Ohlson (1980) proposed an alternative approach to predicting financial failure that relaxed the assumptions of MDA. His model offered valuable insights and has influenced subsequent research in the field.

The inclusion of market variables (Aggarwal et al., 2018), financial ratios, and qualitative variables in predicting financial failures has been a common practice due to their sensitivity and relevance in intelligent models (Agarwal & Taffler, 2008; Altman, 1968; Bauer & Agarwal, 2014; Gilbert et al., 1990; Ravi Kumar & Ravi, 2007; Kaur & Aggarwal et al., 2023). Moreover, the incorporation of corporate governance indicators has proven to significantly impact firm performance (Lee & Yeh, 2004; Liang et al., 2016; Lin, Liang, & Chu, 2010).

Several prominent studies have contributed to the literature on predicting financial failures. Nam et al., (2008) developed a range of techniques including discriminant analysis, logit analysis, probit analysis, regression tree, k-nearest neighbor, rough set, and classification theory. These models have been widely recognized and utilized in the field.

Several studies have been conducted to predict credit risk default and financial distress in various contexts. Agrawal and Maheshwari (2019) focused on logistic regression-based multiple discriminant analysis in Indian firms and identified the industry index as a significant variable for predicting defaults. Obradović et al. (2018) found that models incorporating both financial and non-financial variables achieved higher prediction accuracies compared to models using only financial variables.

In terms of methodology, different approaches have been explored. Bai et al. (2019) utilized rough-set theory and fuzzy C-means clustering to determine creditworthiness in the Chinese agricultural industry, emphasizing skill-based traits and education as important variables. On the other hand, (Khoja et al., 2019; Aggarwal et al., 2019) emphasized the significance of considering local macroeconomic indicators and avoiding single-focused evaluations to reduce insolvency risk in the UK, US, and GCC, BRICS and G7 Countries (Kaur 2024).

Regarding the analysis of corporate governance and financial distress, (Fernando et al., 2020) used the panel logit model to investigate the relationship in the USA and Sri Lanka, achieving a high average accuracy for insolvency prediction in Serbia. In contrast, (Maina 2020) focused on commercial banks in Kenya and highlighted the unique role of corporate governance and supervision frameworks in determining financial distress.

Recent research in credit risk prediction and bankruptcy analysis has employed various models and techniques to enhance accuracy and effectiveness. Machine learning-based models, such as ensemble models, clustering and consensus stages, and random forest algorithms, have been shown to be prevalent and effective in predicting default and credit risk. Studies by (Grishunin et al., 2022; Tang et al., 2019) have highlighted the success of machine learning approaches using financial and macroeconomic indicators.

In addition to machine learning, other models have also demonstrated strong predictive capabilities. Research by (Lin et al., 2019) emphasized the effectiveness of wrapper-based feature selection methods, while (Zhou & Lu et al., 2015; Alaka et al., 2018) employed big data analytics to improve default risk prediction in construction firms. Furthermore, (Khemakhem & Boujelbene, 2018) highlighted the importance of financial ratios and non-financial data for accurate default prediction. The studies collectively show the importance of incorporating

different variables, methodologies, and models to gain comprehensive insights into credit risk and bankruptcy prediction.

In the field of credit risk and default prediction, various studies have been conducted to explore the effectiveness of different statistical modeling techniques. Abdou et al., (2019) investigated the use of logistic regression, discriminant analysis, probabilistic neural network, and multi-layer feedforward neural network in the Indian banking sector. Their research revealed that sophisticated credit risk scoring models like probabilistic neural networks could help mitigate 14% of credit risk defaults. Additionally, they highlighted the importance of demographic-based variables in determining default timing.

In addition to these key contributions, other studies have explored nonparametric models and techniques such as rough set theory (Yeh et al., 2010), decision tree (Chen, 2013), support vector machines, and k-nearest neighbor (West, 2000). Hybrid methods, such as advanced ensemble techniques, have also been employed to enhance the performance of existing models (Choi et al., 2018).

In Pakistan, the absence of early warning signs in non-financial firms has led to insolvency, financial distress, and an increase in unemployment (Hasnain et al., 2022). Various models, including machine learning classifiers like Neural Networks, Decision Trees, and Support Vector Machines, have been proposed to predict credit default, with potential for improved accuracy (Grishunin et al., 2022). However, determining the most accurate model remains a challenge (Halim, et al., 2022).

This research aims to identify early signs of financial problems using an appropriate machine learning model to prevent business failures (Ranawat, & Chakraborty, 2024) Financial ratios are instrumental in predicting defaults (Huang et al., 2022), and the study focuses on their use in the “Through the Cycle Methodology” of BASEL-III.

Different research studies propose machine learning and deep learning models for predicting financial distress or default (Gao & Balyan, 2022). A hybrid machine learning approach with the k-nearest neighbor model and the bagging ensemble method shows promise in credit rating and risk assessment (Lu, 2022; Lim et al., 2024).

Despite the significant advancements in predicting financial failures and credit default, there are still some literature gaps that this research aims to address. The existing studies have primarily focused on the application of machine learning models, such as Neural Networks, Decision Trees, and Support Vector Machines, in predicting credit default (Dube et al., 2021). However, there is a need to determine the most accurate model among these options (Tsai et al., 2021). This research aims to fill this gap by identifying an appropriate machine learning model that can effectively identify early signs of financial problems and prevent business failures (Zhu et al., 2023).

Furthermore, while financial ratios have been widely recognized as instrumental in predicting defaults (Alonso & Carbo, 2021), their use in conjunction with the “Through the Cycle Methodology” of BASEL-III needs further exploration. The literature has not extensively delved into the specific application of financial ratios within this methodology. This research will address this gap by focusing on the use

of financial ratios within the “Through the Cycle Methodology” of BASEL-III to predict financial problems and potential business failures.

Additionally, while machine learning and deep learning models have been proposed for predicting financial distress or default (Gao & Balyan, 2022), there is a need to explore hybrid approaches that combine different models to enhance accuracy. The research will investigate the effectiveness of a hybrid machine learning approach, specifically incorporating the k-nearest neighbor model and the bagging ensemble method, for credit rating and risk assessment (Lu, 2022).

this research aims to contribute to the existing literature by filling the gaps in determining the most accurate machine learning model for predicting financial problems, exploring the application of financial ratios within the “Through the Cycle Methodology” of BASEL-III, and investigating the effectiveness of a hybrid machine learning approach for credit rating and risk assessment. By addressing these gaps, this study seeks to enhance the understanding and prediction of financial distress and potential business failures in non-financial firms in Pakistan.

3 Methodology

The research focuses on credit default prediction and aims to develop accurate prediction models using advanced statistical and machine learning techniques. It adopts a realist ontology, a positivist epistemological approach, and a value-neutral axiology. The methodology emphasizes objective data analysis, reliable prediction model creation, and adherence to established research methodology guidelines. The study follows a deductive approach, collects financial data from reliable sources, and aims to generalize its findings for stakeholders. The methodology ensures researcher independence, impartiality, and the production of authentic and valuable results for credit default prediction.

The research utilizes a comprehensive framework of financial ratios as independent variables to predict credit defaults. These ratios are categorized into capitalization ratios, cash flow ratios, coverage ratios, leverage ratios, liquidity ratios, profitability ratios, and firm size indicators. The study acknowledges the importance of these ratios in assessing an organization’s financial health and risk of default. By including a wide range of financial ratios, the research aims to establish accurate default prediction models applicable across different scenarios, aligning with previous studies in the field.

3.1 Data, Data Preprocessing and Experimental Setup

The study utilized a comprehensive dataset obtained from the State Bank of Pakistan, Annual Reports of each firm, and a financial data service provider in Pakistan. It focused on 396 non-financial firms listed on the Pakistan Stock Exchange, covering various sectors, over a 24-year period from 2000 to 2023. The dataset consisted of 71 financial ratios categorized into seven ratio categories such as liquidity, profitability, capitalization, coverages, leverage, size and cash flows. we

used binary variable was as the dependent variable, where 0 represents non-default, and 1 represents default along with three default assumptions. Furthermore, details for independent and dependent variables may be obtained by clicking the link as given below. <https://docs.google.com/spreadsheets/d/1JuNII1q85rBd7LSf1pfg4RE8L0O0SDX/edit?usp=sharing&oid=104272322739994774143&rtpof=true&sd=true>

Data preprocessing involved eliminating insufficient history and incomplete data, as well as handling outliers using the winsorizing method (Barnett & Lewis, 1994). The resulting clean dataset comprised 396 firms and 71 financial ratios. Multivariate and univariate analysis, including correlation analysis and panel analysis, were performed to analyze the variables' effect on default prediction.

The research employed the Weight of Evidence or Information Value technique to identify the most predictive features. Various machine learning algorithms, such as Logistic Regression, K-Nearest Neighbor, Decision Tree, Random Forest, Support Vector Machine, Naive Bayes, and Artificial Neural Networks, were compared for the default prediction model. The Scikit-learn library was utilized for algorithm testing and evaluation, employing accuracy measures, logistic coefficients, and cross-validation algorithms.

3.2 Feature Selection

The author developed a two-step feature selection process. The first step involved multivariate analysis, where correlation metrics were used to identify and minimize the impact of multicollinearity among the features. This step aimed to ensure that the selected features were not highly correlated with each other (Garg and Tai, 2013).

The second step utilized univariate analysis based on the weight of evidence (WoE) and information value (IV). These parameters, suggested by Wod (1985), were employed to assess the predictive power of individual features. WoE and IV scores were calculated, and features with higher scores were considered more influential for default prediction.

By incorporating these steps, the author aimed to select features that comply with the BASEL-III guideline, which recommends the use of Information Value and Weight of Evidence in financial analysis. Figure 1 explains complete research process in a intelligent way.

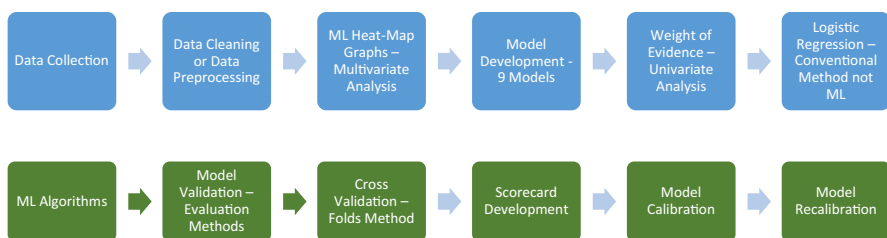


Fig. 1 Research process (Author's Illustration)

3.3 Intelligent Models

3.3.1 Logistic Regression

Logistic regression (LR) is a non-linear equation used for binary class classification, where the dependent variable is binary (either '0' or '1'). LR creates an S-curve to represent the data distribution between categories such as good or bad, default or non-default. It is commonly used in default prediction models (Myung, 2003).

LR allows the calculation of probabilities of defaults (PDs) for individual observations in the dataset based on the coefficients derived from the LR model. This distinguishes LR from other classification models in machine learning.

The author of the research utilized LR as a filtering tool for feature selection. They used LR to identify significant features and eliminate those that were not significant or had falsification in LR analysis.

3.3.2 K-Nearest Neighbor

The k-Nearest Neighbor model was established by Henley, (1996), which is an augmented form of Euclidean distance or Mahalanobis distance between the input feature vector X and the training data of $\langle X_k \rangle_k^m = 1$. It is noticeable that KNN is a supervised machine-learning model which needs the input variable. Below mentioned derivation represents the equation of the KNN classification algorithm:

3.3.3 Decision Trees

A decision tree classification algorithm was proposed by Rutkowski et al., (2012) DT is also a kind of supervised learning which needs certain input parameters or information to reach the results. There are root nodes that discriminant between default and non-default classes. The simplest example is given in the below figures.

3.3.4 Random Forests

ML algorithm for Random Forest was deliberated by Breiman (2001), which is an ensemble of decision trees. Supposedly, numbers of decision trees are constructed with bootstraps strategical sampling with a certain number of observations. Every decision DT was established based on a subset of randomly selected K features. A new feature vector will be assigned as a class by every DT. Hence, a consolidated classification. RF will assign a new feature vector a class through a voting system based on the outputs secured by DT nodes. The below figure elucidates the flow of RF working.

3.3.5 Support Vector Machines

Cortes et al., (1995) proposed model of support vector machine (SVM), which is structured on the decision boundary or Hyperplane, which discriminant classes in extensively high dimensional feature space. Linear SVM algorithm emphasis on the maximization of margins of $\|a\| = \sqrt{\sum_i^m 1a_i^2}$ between the default and non-default. Classes are placed according to the below mentioned equation of SVM.

3.3.6 Naïve Bayes

Naïve Bayes is a well-known machine learning algorithm utilized for classification tasks. It relies on Bayes' theorem and assumes that the features within a dataset are independent of each other, given the class label. This assumption simplifies the probability calculations and enhances the algorithm's computational efficiency (Dempster et al., 1977).

3.3.7 Artificial Neural Networks

Mitchell (1997) elucidated that artificial neural network (ANN) was developed based on a biological neural network system. it is quite similar to the nexus of brain neurons and its expertise to process data and make it meaningful (Thomas, 2000; Garg, & Aggarwal, 2021; Tsai et al., 2014). ANN is comprised of three layers (1) the Input Layer, (2) the Hidden Layer, and (3) Layers Tsai and (Wu, 2008), as given in the figure appended below.

3.4 Model Validation

The developed model will be rigorously validated using a range of evaluation metrics, including Precision, Recall and F1-Score. These metrics will assess the accuracy and reliability of the established model(s). To ensure robustness, cross-validation techniques such as K-Fold Cross Validation and Stratified Fold Cross Validation will be employed.

Once the model(s) have undergone comprehensive evaluation, rating scorecards will be constructed. Since ML algorithms do not provide coefficients like logistic regression, an alternative approach will be used. Feature rankings based on permutation importance, as studied by Boughaci & Alkhaldeh, (2020; Luo, (2022), will be utilized. Weights will then be assigned to each feature as given in the Table 1, reflecting their importance in the rating scorecards. This method guarantees that the selected features are appropriately considered in the credit risk assessment process.

To calculate the probability of default (PD) values for each observation, the study applied weights to the variables in each classifier. These weights were multiplied with the corresponding ratios to obtain PD values. logistic regression (LR) used

Table 1 Assigning weights for scorecard creation

Variables	ML model	Inverse	weights
S3	1	12	15.38%
LR3	2	11	14.10%
CF3	3	10	12.82%
CAP4	4	9	11.54%
LIQ10	5	8	10.26%
PR7	6	7	8.97%
CR6	7	6	7.69%
CR4	8	5	6.41%
LIQ3	9	4	5.13%
LIQ13	10	3	3.85%
PR1	11	2	2.56%
PR14	12	1	1.28%

The table provided below shows how author assigned weights to each variable based on their importance in generating credit ratings, as mentioned in previous literature (Boughaci & Alkhawaldeh, 2020; Luo, 2022)

coefficients as weights, while exponential distribution was utilized for both LR and the other six classifiers.

For LR, the exponential LR scores were derived by multiplying the ratios with the coefficients, and these scores were used to calculate Odd scores (PDs). The same treatment was applied to other financial ratios by multiplying them with assigned weights to obtain PDs. This approach ensures a fair contribution of each variable in the calculation of PD, leading to a more accurate assessment of credit risk.

The author will establish nine credit rating buckets based on the guidelines of BASEL Accord III and assign the obtained ratings to these buckets accordingly. In the final phase, the obtained ratings will be calibrated. Following the recommendations of BASEL Accord III, calibration and recalibration tests will be conducted. Calibration tests will include metrics such as area under curve (AUC), receiver operating characteristic (ROC), GINI Coefficient, KS-Stats, and Brier Score. Recalibration tests will involve the binomial distribution—traffic lights approach and the population stability index (PSI). By employing these tests, the author aims to ensure the accuracy and reliability of the obtained ratings and to align them with the standards set by BASEL Accord III.

4 Results and Analysis

In our methodology, we initiate the testing process by identifying the best features through both multivariate and univariate analyses. Initially, we employ correlation heatmaps to address multicollinearity within our dataset, resulting in the formation of three distinct panels. Consequently, this approach leads to the creation of nine models, as each panel is further examined under three different

default assumption criteria. Following this, we refine our feature selection using univariate analysis tools, as recommended by Abdullah et al. (2023); Wod (1985). After completing the univariate analysis, we proceed to narrow down our features with the aid of logistic regression, employing it as a filtering tool in its conventional form—without dividing the dataset into training and test splits—as advised by Zhou et al., (2015). Upon finalizing the feature filtering process to pinpoint the most effective and relevant features, we incorporate these selected features into our machine learning algorithms for further analysis. The outcomes of this feature selection process are accessible via the link provided below. https://docs.google.com/document/d/1764wBC_2PL7I_jdsLFUJp4Adnnk8iIFm/edit?usp=sharing&oid=104272322739994774143&rtpof=true&sd=true

Above Table 2 exhibited remarkable accuracy in the test-trained split compared to all nine models. The random forest classifier (RFC) achieved the highest accuracy rate of approximately 87.96%, with minimal false positives (FP) and false negatives (FN) compared to other classifiers. The artificial neural network (ANN) ranked second, with an accuracy rate of around 86.95%, followed by support vector machines (SVM) at 86.21%. The confusion matrix confirms the improved accuracy across all classifiers, indicating the selection of the best model.

Notably, model 6 (out of nine) demonstrated a significant increase in accuracy compared to model 5. It outperformed all other models, including model 5. Model 6 was developed with 12 features, employing a relaxed default approach with a 60% erosion of equity. The Table 3 provided below presents the relative rankings and accuracy for each classifier, accompanied by their margins of error.

Table 2 Results of machine learning algorithms

Results of ML model – model 6

Confusion matrix	True positive	False positive	False negative	True negative	Total
LR	1410 74.48%	68 3.59%	246 13.00%	169 8.93%	1893
KNN	1339 70.73%	73 3.86%	222 11.73%	259 13.68%	1893
GNB	1346 71.10%	66 3.49%	267 14.10%	214 11.30%	1893
SVM	1361 71.90%	51 2.69%	210 11.09%	271 14.32%	1893
DTC	1306 68.99%	106 5.60%	188 9.93%	293 15.48%	1893
RFC	1354 71.53%	58 3.06%	170 8.98%	311 16.43%	1893
ANN	1343 70.95%	69 3.65%	178 9.40%	303 16.01%	1893

We only report result of the most adequate model based on model accuracy. result of rest of 8 models can be upon a request. Hence, in this study model – 6 was the best amongst all and further calibration & recalibration is done based on model – 6

Table 3 Relative rankings of ML algorithms

Confusion matrix—Summary test	ML models		
	Accurate results	Type-I & Type-II Error	Ranks based on accuracy
LR	83.41%	16.59%	6
KNN	84.42%	15.58%	5
GNB	82.41%	17.59%	7
SVM	86.21%	13.79%	3
DTC	84.47%	15.53%	4
RFC	87.96%	12.04%	1
ANN	86.95%	13.05%	2

Table 4 Classification report—ML models

Classification report		Precision	Recall	F1-Score
LR	Accuracy			83%
	Macro avg	78%	68%	71%
	Weighted avg	82%	83%	82%
KNN	Accuracy			84%
	Macro avg	82%	78%	77%
	Weighted avg	84%	84%	83%
GNB	Accuracy			82%
	Macro avg	80%	70%	73%
	Weighted avg	82%	82%	81%
SVM	Accuracy			86%
	Macro avg	85%	76%	79%
	Weighted avg	86%	86%	85%
DTC	Accuracy			84%
	Macro avg	80%	77%	78%
	Weighted avg	84%	84%	84%
RFC	Accuracy			88%
	Macro avg	87%	80%	83%
	Weighted avg	88%	88%	87%
ANN	Accuracy			87%
	Macro avg	85%	79%	81%
	Weighted avg	87%	87%	86%

Table 4 illustrate accuracy scores of all ML algorithms with Model 6. The accuracy of model 6 was confirmed through validation, utilizing the classification report and complementary testing. The random forest classifier (RFC) achieved the highest accuracy of approximately 88%. It exhibited precision, recall, and F1-Score values of 88%, 88%, and 87% respectively. The Artificial Neural Network (ANN) followed closely with an accuracy of around 87% and precision, recall, and F1-Score values of 87%, 87%, and 86% respectively. The Support Vector Machine (SVM) ranked

third in accuracy, achieving an accuracy rate of approximately 86% along with a precision of 86%, recall of 86%, and F1-Score of 85%.

Additionally, it is worth noting that the decision tree classifier (DTC) and K-Nearest neighbors (KNN) showed similar performance. Both classifiers attained an accuracy rate of around 84%, precision of 84%, and F1-Score of 84%. However, DTC had a higher recall of 88% compared to KNN's 83% recall.

These results further validate the accuracy of model 6, indicating its strong performance across multiple evaluation metrics and supporting the superiority of RFC, ANN, and SVM in predicting default rates.

Table 5 represent results of model validation and cross validation. All classifiers were thoroughly validated, with the top-ranked classifiers (RFC, ANN, and SVM) maintaining their positions. Cross-validation results confirmed the excellent performance of RFC, with accurate fold values around 86%, indicating its insensitivity to data shuffling. SVM also demonstrated good performance, with fold values around 84%, suggesting its effectiveness in handling data shuffling issues.

Model 6, constructed with 12 features and a relaxed default approach (60% erosion of equity), and exhibited a significant increase in accuracy compared to model 5. It outperformed all other models, including model 5. Complementary testing validated the accuracy of model 6, with RFC achieving the highest accuracy of 88%, followed by ANN with 87% accuracy and SVM with 86% accuracy. KNN and DTC showed similar accuracies around 84%. The rankings of the top classifiers remained consistent, and all classifiers were significantly justified in the validation process.

Random forest classifier (RFC), support vector machines (SVM), and artificial neural networks (ANN) emerged as the most effective algorithms for default prediction modeling (DPM) in this study. RFC achieved the highest accuracy, followed by ANN and SVM, consistent with previous research. These algorithms are recommended for similar DPM tasks. Model 6 with RFC as the classifier showed the highest accuracy and validation scores among all nine models, making it the best scenario for DPM. The relaxed default approach in Model 6, with a 60% erosion of equity, outperformed other scenarios. ANN performed slightly better in Model 6 compared to Model 5. The study suggests further exploration of different default

Table 5 Model validation and cross validation

Validation—ML Models							
Specification	LR	KNN	GNB	SVM	DTC	RFC	ANN
FP (Type-I Error):	68	73	66	51	106	58	69
FN (Type-II Error):	246	222	267	210	188	170	178
Model Accuracy:	83.41%	84.42%	82.41%	86.21%	84.47%	87.96%	86.95%
Cross Validation—ML Models							
Specification	LR	KNN	GNB	SVM	DTC	RFC	ANN
K-Fold	81.65%	81.51%	83.39%	84.33%	79.64%	86.21%	80.74%
Stratified Fold	81.74%	81.11%	83.33%	84.14%	79.03%	86.12%	80.63%

Table 6 Selected features

Indication	Code	Group	Formulae	Expected relationship	Actual signs
0	CAP4	Capitalization	Non-current liabilities/total assets	Positive	Positive
1	CF3	Cash Flow	Log(OCF)	Negative	Negative
1	LIQ10	Liquidity	Current assets less stock/total assets	Negative	Negative
1	LIQ13	Liquidity	Stock/current assets	Negative	Negative
1	LIQ3	Liquidity	Cash/current liabilities	Negative	Negative
0	LR3	Leverage	Log(Total LT Debt + Total ST Debt)	Positive	Positive
1	PR1	Profitability	Net Sales / Total Asset	Negative	Negative
1	PR14	Profitability	Operating revenue/total assets	Negative	Negative
0	PR7	Profitability	Financial expense/current assets	Positive	Positive
1	S3	Size	log(Total Assets)	Negative	Negative
1	CR4	Coverage Ratios	EBIT / Total Liabilities	Negative	Negative
1	CR6	Coverage Ratios	FFO/Total Debt	Negative	Negative

Note: 0 means lower the better and 1 represents higher the better

conditions to identify key thresholds based on equity erosion. Overall, Model 6 with ANN as the classifier is the best scenario for predicting a firm's default/non-default status.

Table 6 represent selected features by all filtering process and machine learning testing which is basically features of model 6. The model in the study includes financial ratios from various streams, such as Capitalization, Cash Flow, Coverage Ratios, Leverage, Liquidity, Profitability, and Size. The analysis of these ratios revealed important findings. Non-current liabilities to total assets showed a positive relationship with credit default, indicating higher default probability for companies relying heavily on non-current liabilities. Operating cash flows had a negative association with default events, suggesting that higher operating cash flows reduce the likelihood of default.

Liquidity ratios, specifically LIQ10, LIQ13, and LIQ3, showed negative and significant associations with default events, indicating that higher values for these ratios indicate better financial health. The least significant variable was LIQ13, but it still contributed to the calibration of default probabilities. LR3 (Log (Total LT Debt+Total ST Debt)) had a positive relationship with defaults, while PR1 (Net Sales / Total Assets) and PR14 (Operating revenue/total assets) had negative relationships with defaults, highlighting the importance of asset turnover and cost control.

Firm size (S3) was negatively related to defaults, indicating that larger firms are less risky. Finally, CR4 (EBIT / Total Liabilities) and CR6 (FFO/Total Debt) exhibited negative relationships with defaults, emphasizing the importance of adequate EBIT and cash flows to cover liabilities and debts. The identified features in

Model 6 are justifiable for default prediction models. Financial ratios play a crucial role in default prediction, as emphasized by previous research. The capitalization ratio, cash flow, liquidity, profitability, coverage ratios, debt, and firm size are all important factors in default prediction.

Table 7 illustrate generated probability of default ranges for all seven classifiers and this will be used for scorecard development. In the study, Model 6 was determined to be the best model for default prediction. However, since ML classifiers like KNN, GNB, SVM, DTC, RFC, and ANN do not provide model coefficients, permutation importance was used to rank the features in terms of their importance. Weights were assigned to each feature based on their ranks, with higher ranks indicating higher weights. Only LR provided model coefficients, which were used to calculate odds scores for each variable. For ML classifiers, a weighted average number (Ratio * Weight) was used as the default score for each variable. An exponential distribution was then applied to standardize the distribution across all classifiers.

To assign credit ratings, 9 percentiles were used to bin the default scores into rating bands, following the guidelines of the State Bank of Pakistan. The aim was to have fewer defaults in higher credit ratings and a higher chance of default in lower credit ratings. PD ranges were developed for all 7 classifiers and used to assign credit ratings to each data point (entity for every year). These assigned ratings were then calibrated according to the guidelines of Basel Accord-III and (Bequé et al., 2017) for model calibration and monitoring.

The study involved developing credit rating bands based on default scores obtained from Model 6 and ML classifiers. These bands were calibrated and monitored according to the Basel Accord-III guidelines. Binning of rating bands are placed in link given as below. https://docs.google.com/document/d/1764wBC_2PL7L_jdsLFUJp4Adnnk8iIFm/edit?usp=sharing&oid=104272322739994774143&rtopof=true&sd=true

Table 8 represent results of stress testing or calibration or generated credit ratings (which is basically credit scorings or scorecards). In the previous section, the author focused on evaluating the discriminatory power of the RFC model, which was determined to be the best classifier based on accuracy, validation, and cross-validations. The probability of default (PD) was derived using feature rankings, and credit ratings were assigned based on SBP slabs of ratings. The percentage of defaults was calculated for each rating grade, and a cumulative accuracy profile (CAP) graph was generated.

The table above presents the results of the actual, perfect, and random models. The CAP graph in the bottom right corner indicates the discriminatory power of the RFC model. It is observed that the actual model performs well, with a concave curve lying between the perfect and random models, indicating improved discrimination. This aligns with the Basel-III guideline, which suggests that the CAP graph should exhibit such characteristics. The Area under the ROC curve (AR) and GINI values for the RFC model are both around 0.50, indicating a solid discriminatory power and these findings are backed by Abdou, Tsafack, Ntim, & Baker, (2016). This is further supported by the Kolmogorov–Smirnov (KS) test, which reveals that the majority of defaults are concentrated in the second slab, accounting for 39.32% of defaults.

Table 7 Probability of default (PDs) ranges

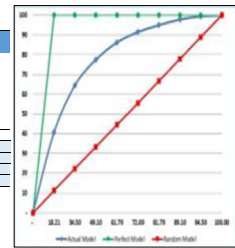
Rating band— classifiers	AAA, AA +, AA	AA, AA-, A +	A +, A, A-	A-, BBB +, BBB	BBB-, BB+, BB	BB, BB-, B +	B +, B, B-, CCC	B-, CCC, CC, C	CC, C
LR	0.00%	0.01%	0.09%	0.24%	0.45%	0.70%	1.06%	1.62%	3.00%
KNN	14.73%	17.43%	19.53%	21.07%	22.39%	23.71%	25.21%	26.86%	29.43%
GNB	1.56%	18.28%	24.30%	26.73%	28.49%	29.98%	31.43%	33.08%	35.50%
SVM	6.78%	14.12%	16.42%	18.31%	19.89%	21.42%	22.94%	24.71%	27.55%
DTC	0.70%	15.30%	22.62%	25.43%	27.06%	28.41%	29.77%	31.22%	33.45%
RFC	0.67%	15.45%	22.81%	25.70%	27.48%	28.92%	30.41%	32.03%	34.45%
ANN	13.93%	18.36%	20.77%	22.58%	23.97%	25.39%	26.88%	28.52%	31.19%

Table 8 Random forest classifier (calibration results)

Scenario # 6		Actual Model				Perfect Model			Random Model					
Grade	# of Observations	Observed Defaults	Cumulative Defaults	% of Defaults	Cumulative % of Defaults	Defaults	Cumulative Default	Cumulative % of Defaults	Defaults	Cumulative Default	Cumulative % of Defaults	Cumulative Observations	% of Cum. Observation	
Worse Grade	1	1149	608	608	40.59	1498	1498	100.00	166.44	166.44	11.11	1149	18.21	
	2	1028	388	966	23.90	64.49	0	1498	100.00	166.44	332.89	22.22	2177	34.50
	3	921	193	1159	12.88	77.37	0	1498	100.00	166.44	499.33	33.33	3098	49.10
	4	801	132	1291	8.81	86.18	0	1498	100.00	166.44	665.78	44.44	3899	61.79
	5	688	79	1370	5.27	91.46	0	1498	100.00	166.44	832.22	55.56	4887	72.69
	6	574	53	1423	3.54	94.99	0	1498	100.00	166.44	998.67	66.67	5161	81.79
	7	461	43	1466	2.87	97.36	0	1498	100.00	166.44	1165.11	77.78	5622	89.10
	8	341	24	1490	1.60	99.47	0	1498	100.00	166.44	1331.56	88.89	5963	94.50
Best Grade	9	347	8	1498	0.53	100.00	0	1498	100.00	166.44	1498.00	100.00	6310	100.00

Grade	Total Observations	Defaults	Cum. Defaults	% of T. Defaults	Cumulative % % of Defaults	# of Goods	Cumulative Goods	Cumulative % % of Goods	KS	Intervals
Worse Grade	1	1149	608	40.59%	40.59%	541	541	11.24%	29.34%	39.32%
	2	1028	388	23.90%	64%	670	1211	25%	39.3%	0
	3	921	193	12.90%	77%	728	1939	40%	37.1%	0
	4	801	132	8.8%	86%	669	2608	54%	32.0%	9
	5	688	79	5.3%	91%	609	3217	67%	24.0%	0
	6	574	53	3.5%	95%	521	3738	78%	17.3%	9
	7	461	43	2.9%	98%	418	4156	86%	11.5%	0
	8	341	24	1.6%	99%	317	4473	93%	6.5%	9
Best Grade	9	347	8	0.5%	100%	339	4812	100%	0.0%	0

Accuracy Ratio	0.4987
GINI Coefficient	0.4987
≤-AUC	0.7493
Brier Score	0.1710



The KS test highlights the discriminatory power of the RFC model, particularly in the first two slabs (worst grade), while the third slab also shows acceptable performance but requires further refinement. The RFC model demonstrates the strongest discriminatory power in predicting defaults, as evidenced by the KS statistics. Additionally, it achieves the lowest Brier score among all classifiers, including LR, indicating that it is the superior classifier for default prediction modeling (DPM).

4.1 Results of 2023

4.1.1 Random Forest Classifier

Table 9 are results of recalibration based on population stability index (PSI). Based on the analysis, it is evident that the model’s performance in 2023 has shown significant improvement, as indicated by the robust PD distribution pattern, with only a slight

Table 9 Population stability index

	Bands	Scale	Test	Trained	Test—Trained	Ln of Test of Trained	PSI
Worse Grade	CC, C	9	52.9%	52.92%	0.03%	0.00	0.000
	B-, CCC, CC, C	8	28.3%	34.82%	-6.56%	-0.21	0.014
	B+, B, B-, CCC	7	20.0%	20.96%	-0.96%	-0.05	0.000
	BB, BB-, B+	6	11.1%	16.48%	-5.37%	-0.39	0.021
	BBB-, BB+, BB	5	10.0%	11.48%	-1.48%	-0.14	0.002
	A-, BBB+, BBB	4	7.7%	9.23%	-1.54%	-0.18	0.003
	A+, A, A-	3	15.0%	9.33%	5.67%	0.48	0.027
	AA, AA-, A+	2	0.0%	7.04%	-7.04%	0.00	0.000
Best Grade	AAA, AA+, AA	1	0.0%	2.31%	-2.31%	0.00	0.000

0.07

variation between grades 3 and 4. However, this discrepancy has been accounted for and normalized through the log of test and training, resulting in a stable PSI score. Furthermore, the observed pattern of more defaults in the last rating grade and fewer defaults in the best rating grades aligns with the expected behavior.

The PSI value for the model is approximately 0.07, which is considered sufficient according to the guideline. Therefore, it can be concluded that no further adjustments are necessary for the model. To validate these findings, a binomial distribution analysis was also conducted, and the results support the research conclusions.

The performance of the 2023 model is deemed satisfactory, and based on both the PD distribution pattern and the binomial distribution analysis, recalibration is not required.

Table 10 denotes results of Binomial Distribution—Traffic Lights which is another test of recalibration prescribed by BASEL Accord III. The results demonstrate that the 2023 model performs well, with a consistent PD distribution pattern, albeit with a minor variation between rating grades 3 and 4. However, this difference has been normalized through the log of the test and training data, resulting in a stable PSI score of approximately 0.07. The lower rating grades exhibit higher default events, as expected, while the higher rating grades have fewer defaults. The binomial distribution analysis also supports the conclusion that no recalibration or changes are required.

Table 10 Binomial Distribution—Traffic Lights

Data		Tests (<i>p</i> -values)						
		r						
		10.0%						
Default rates		Trained Data	Test Data	N2023	D2023	Binomial	Normal	One-factor
Worse Grade	CC, C	52.92%	52.94%	51	27	55.5%	55.4%	50.4%
	B-, CCC, CC, C	34.82%	28.26%	46	13	86.3%	86.2%	68.9%
	B+, B, B-, CCC	20.96%	20.00%	40	8	62.1%	63.4%	48.8%
	BB, BB-, B+	16.48%	11.11%	36	4	86.6%	86.3%	71.9%
	BBB-, BB+, BB	11.48%	10.00%	30	3	68.5%	70.6%	51.8%
	A-, BBB+, BBB	9.23%	7.69%	26	2	70.6%	72.9%	53.3%
	A+, A, A-	9.33%	15.00%	20	3	28.5%	31.3%	14.3%
	AA, AA-, A+	7.04%	0.00%	15	0	NA	94.2%	NA
Best Grade	AAA, AA+, AA	2.31%	0.00%	16	0	NA	92.6%	NA

When comparing all the models, the 2022 data reveals a wide range of numbers across different classifiers. Some models report the highest default events in the 1st interval of KS stats, while others report them in the 2nd or 3rd interval. In 2022, the RFC model demonstrated strong performance in terms of AR and GINI, but lacked discriminatory power. Conversely, ANN outperformed all models with excellent predictive power and adequate discriminatory power. LR showed healthy discriminatory indicators but limited predictive power, while KNN exhibited reasonable numbers for both discriminatory and predictive power. GNB and DTC produced results that were more closely aligned.

The 2023 model’s performance is deemed satisfactory, and there is no need for further recalibration or changes. The following comprehensive comparison table summarizes the performance of all the models.

Table 11 Summary of recalibration – 2023

Default obs. / Actual			LR	KNN	GNB	SVM	DTC	RFC	ANN
Worse Grade	Bands	Scale							
	CC, C	9	62.75%	41.18%	49.02%	45.10%	45.10%	52.94%	45.10%
	B-, CCC, CC, C	8	26.09%	34.78%	28.26%	30.43%	34.78%	28.26%	28.26%
	B+, B, B-, CCC	7	5.00%	17.50%	22.50%	17.50%	17.50%	20.00%	20.00%
	BB, BB-, B+	6	11.11%	16.67%	13.89%	16.67%	19.44%	11.11%	22.22%
	BBB-, BB+, BB	5	20.00%	10.00%	6.67%	13.33%	3.33%	10.00%	3.33%
	A-, BBB+, BBB	4	7.69%	3.85%	11.54%	3.85%	15.38%	7.69%	3.85%
	A+, A, A-, AA, AA-, A+	3	5.00%	10.00%	10.00%	15.00%	5.00%	15.00%	10.00%
Best Grade	AAA, AA+, AA	1	0.00%	25.00%	0.00%	0.00%	0.00%	0.00%	6.25%
	Year—2023								
	Accuracy Ratio		0.57371	0.39174	0.48697	0.42886	0.47915	0.52735	0.41697
	GINI Coefficient		0.57371	0.39174	0.48697	0.42886	0.47915	0.52735	0.41697
	<--AUC		0.78686	0.69587	0.74348	0.71443	0.73957	0.76367	0.70848
	KS Stats		49.24%	34.39%	37.42%	34.39%	38.52%	40.76%	32.27%
	KS Interval		2	2	3	2	2	2	2
	Brier Score		0.19769	0.15899	0.16082	0.15691	0.15968	0.15882	0.15896

Results of 2016–2022 are also generated to check model’s robustness and results were consistent like 2023, therefore results of 2016–2022 can be share upon request

The Table 11 provides a comprehensive comparison of recalibration models. Most default events were concentrated in the second interval of KS stats, with GNB being the only model showing the highest defaults in the third slab, which is still acceptable. To draw a conclusion, it is necessary to consider each model's discriminatory and predictive values.

The LR model exhibited a diverse distribution and rating shuffles, resulting in decreased predictive power (highest BR Score: 0.19769). Similar observations were made for the ANN model, which had a distorted PD distribution impacting the BR Score. The remaining models showed a similar shuffle between rating grade 3 and rating grade 4. LR had strong discriminatory power with AR and GINI values around 0.57371 but slightly weak predictive power. KNN had weak discriminatory power with AR and GINI values around 0.39174 but good predictive power with a BR Score around 0.15899.

GNB, SVM, and DTC performed closely to each other, exhibiting good discriminatory power with AR and GINI values of 0.48697, 0.42886, and 0.47915, respectively. They also showed adequate predictive power with BR Scores of 0.16082, 0.15691, and 0.15968, respectively. ANN demonstrated good but not adequate performance in recalibration, with AR and GINI values around 0.41697, slightly lower than all classifiers except KNN, but good predictive power with BR scores around 0.15896.

Lastly, RFC remained the best model in recalibration and model monitoring, with AR and GINI values being the second-highest around 0.52735 (lower than LR but with better predictive power) and BR Scores around 0.15882. In conclusion, RFC outperformed all other classifiers.

5 Discussion

Default prediction is a significant concern globally, as demonstrated by high-profile bankruptcies such as WorldCom, Enron, and Lehman Brothers. Default prediction models have been developed for both mature and emerging markets, but they have faced criticism for methodological and contextual issues, particularly in defining default. Forecasting default events before they occur is crucial in corporate finance to protect stakeholders. This research addresses the gaps in default prediction modeling and serves as a benchmark, providing a comprehensive guide that incorporates recommended machine learning algorithms, validation techniques, scorecard presentation, calibration, recalibration, model stability, and monitoring. The research's novel contribution aligns with Basel Accord III, IFRS guidelines, and ESMA Standards, making it valuable to financial institutions, rating agencies, regulatory bodies, and other stakeholders. It offers real-time probability of defaults (PDs) for various machine learning algorithms, enabling stakeholders to make informed decisions on credit lending to entities in Pakistan. The research also provides the option for stakeholders to generate their own PDs using the provided coefficients or feature scores. Overall, this research is a comprehensive resource for anyone involved in default prediction models (DPM).

6 Conclusion

The research focused on developing a default prediction model using machine learning techniques. Data preprocessing involved removing variables with incomplete history and missing values. Feature selection was performed through multivariate and univariate analysis, resulting in 13 efficient features for predicting default. Seven machine learning classifiers (LR, KNN, GNB, SVM, DTC, RFC, and ANN) were evaluated, with RFC performing the best in terms of accuracy, validation, and cross-validation. The findings aligned with previous research. The third objective aimed to compare the results of all classifiers, highlighting RFC as the highest-performing model with an accuracy and validation rate of 88%. The study aimed to align with BASEL III guidelines and used the population stability index (PSI) for model stability assessment. The RFC model demonstrated superior calibration, stability, and monitoring. Probability of default (PD) scores generated by RFC were the strongest among the classifiers. Overall, the research provided valuable insights into default prediction models in corporate finance.

The first objective was achieved by identifying 13 effective features that can accurately predict default. The research findings align with the studies conducted by (Christopoulos et al., 2019; Javaid & Javid, 2018; Inam et al., 2019; Karas & Reznakova, 2020; Zhu et al., 2019; Sariev & Germano, 2020; Muñoz-Izquierdo et al., 2019; Ragab & Saleh, 2021; Petropoulos et al., 2020; Ogachi et al., 2020; Bhattacharya & Sharma, 2019; Shrivastava et al., 2020). Results of the best features were reported in the Table 6 (Selected Features) of Sect. 4 “Results and Analysis”.

6.1 Practical Implication

The research has significant implications for financial analysts, credit analysts, academic researchers, financial institutions, and credit rating companies in Pakistan. It recommends incorporating real-time probability of default (PD) ranges as a tool for assessing financial risk. The identified key features for default prediction can enhance credit risk evaluation. Academic researchers can cite this comprehensive study for future work. The statistical framework can be used in parallel with judgmental models in commercial banks’ internal rating based (IRB) systems. The research fills a gap by providing valuable statistical insights for credit rating agencies in Pakistan. Overall, it improves financial analysis, credit assessment, and understanding of default prediction models in the corporate finance domain.

6.2 Limitation of Research

The study has limitations in terms of not including all advanced machine learning and deep learning algorithms due to resource constraints. It also lacks the incorporation of macroeconomic factors, corporate governance indicators, and ESG indicators due to data accessibility issues. The study focused on BASEL-III guidelines but did not use the IFRS calibration methodology. Additional financial

ratios were not included to maintain comprehensive coverage. The study's sample was limited to certain types of firms in Pakistan, which may affect generalizability.

Acknowledgements We are very much thankful to all those who contributed their valuable input while conducting this research.

Declarations

Conflict of interest We solemnly declare that neither I 'Dr. Jahanzaib Alvi' nor my co-author 'Prof. Dr. Imtiaz Arif' have any sort of conflict of interest with this research, the work is neutral and genuine.

References

- Abdou, H. A., Mitra, S., Fry, J., & Elamer, A. A. (2019). Would two-stage scoring models alleviate bank exposure to bad debt? *Expert Systems with Applications*, *128*, 1–13. <https://doi.org/10.1016/j.eswa.2019.03.028>
- Abdou, H. A., Tsafack, M. D. D., Ntim, C. G., & Baker, R. D. (2016). Predicting creditworthiness in retail banking with limited scoring data. *Knowledge-Based Systems*, *103*, 89–103.
- Abdullah, M., Chowdhury, M. A. F., Uddin, A., & Moudud-Ul-Huq, S. (2023). Forecasting nonperforming loans using machine learning. *Journal of Forecasting*, *42*(7), 1664–1689.
- Abellán, J., & Castellano, J. G. (2017). A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Systems with Applications*, *73*, 1–10. <https://doi.org/10.1016/j.eswa.2016.12.020>
- Agarwal, V., & Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking and Finance*, *32*(8), 1541–1551. <https://doi.org/10.1016/j.jbankfin.2007.07.014>
- Aggarwal, S., Gupta, S. K., & Pathak, S. (2020). Green bonds—An instrument for financing a sustainable future. *Test Eng J*, *83*, 16459–16470.
- Aggarwal, S., & Khurana, S. (2018). Empirical examination of stock market volatility: An international comparison. *Indian Journal of Finance*, *12*(1), 47–61.
- Aggarwal, S., & Raja, A. (2019). Stock market interlinkages among the BRIC economies. *International Journal of Ethics and Systems*, *35*(1), 59–74.
- Agrawal, K., & Maheshwari, Y. (2019). Efficacy of industry factors for corporate default prediction. *IIMB Management Review*, *31*(1), 71–77. <https://doi.org/10.1016/j.iimb.2018.08.007>
- Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., & Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*, *94*, 164–184. <https://doi.org/10.1016/j.eswa.2017.10.040>
- Alhammadi, M. A. A., Fernandez, A. I., & Romero, A. V. (2024). Credit scoring and risk management in Islamic banking: The case of Al Etihad Credit Bureau. *Revista Venezolana De Gerencia: RVG*, *29*(105), 111–124.
- Alonso, A., & Carbo, J. M. (2021). Understanding the performance of machine learning models to predict credit default: A novel approach for supervisory evaluation. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3774075>
- Altman, Edward I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 189–209.
- Altman, E. I., & Rijken, H. A. (2004). How rating agencies achieve rating stability. *Journal of Banking & Finance*, *28*(11), 2679–2714.
- Alvi, J., & Arif, I. (2024). Default prediction modeling (DPM) with machine learning algorithms: case of non-financial listed companies in Pakistan. *Kybernetes*. <https://doi.org/10.1108/K-09-2023-1888>
- Baek, H. (2023). A CNN-LSTM Stock Prediction Model Based on Genetic Algorithm Optimization. *Asia-Pacific Financial Markets*, *31*(2), 205–220.
- Bai, C., Shi, B., Liu, F., & Sarkis, J. (2019). Banking credit worthiness: Evaluating the complex relationships. *Omega (United Kingdom)*, *83*, 26–38. <https://doi.org/10.1016/j.omega.2018.02.001>

- Balasubramanian, S. A., R, G. S., P, S., & Natarajan, T. (2019). Modeling corporate financial distress using financial and non-financial variables: The case of Indian listed companies. *International Journal of Law and Management*, 61(3/4), 457–484.
- Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405–417. <https://doi.org/10.1016/j.eswa.2017.04.006>
- Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (Vol. 3). New York: Wiley.
- Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking and Finance*, 40(1), 432–442. <https://doi.org/10.1016/j.jbankfin.2013.12.013>
- Beaver, W. H. (1966). Financial ratios as predictors. *Journal of Accounting Research*, 4(1966), 71–111.
- Beaver, W., McNichols, M., & Rhie, J. W. (2005). Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *In Review of Accounting Studies*, 10(1), 93–122. <https://doi.org/10.1007/s11142-004-6341-9>
- Bequé, A., Coussement, K., Gayler, R., & Lessmann, S. (2017). Approaches for credit scorecard calibration: An empirical analysis. *Knowledge-Based Systems*, 134, 213–227.
- Bhattacharya, S., & Sharma, D. (2019). Do environment, social and governance performance impact credit ratings: A study from India. *International Journal of Ethics and Systems*, 35(3), 466–484. <https://doi.org/10.1108/IJOES-09-2018-0130>
- Boughaci, D., & Alkhalwaleh, A. A. K. (2020). Appropriate machine learning techniques for credit scoring and bankruptcy prediction in banking and finance: A comparative study. *Risk and Decision Analysis*, 8(1–2), 15–24. <https://doi.org/10.3233/RDA-180051>
- Branch, B. (2002). The costs of bankruptcy: A review. *International Review of Financial Analysis*, 11(1), 39–57.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.
- Chen, M. Y. (2013). A hybrid ANFIS model for business failure prediction utilizing particle swarm optimization and subtractive clustering. *Information Sciences*, 220, 180–195.
- Choi, H., Son, H., & Kim, C. (2018). Predicting financial distress of contractors in the construction industry using ensemble learning. *Expert Systems with Applications*, 110, 1–10. <https://doi.org/10.1016/j.eswa.2018.05.026>
- Chopra, A., & Bhilare, P. (2018). Application of ensemble models in credit scoring models. *Business Perspectives and Research*, 6(2), 129–141. <https://doi.org/10.1177/2278533718765531>
- Christopoulos, A. G., Dokas, I. G., Kalantonis, P., & Koukkou, T. (2019). Investigation of financial distress with a dynamic logit based on the linkage between liquidity and profitability status of listed firms. *Journal of the Operational Research Society*, 70(10), 1817–1829. <https://doi.org/10.1080/01605682.2018.1460017>
- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273–297.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (methodological)*, 39(1), 1–22.
- Dube, F., Nzimande, N., & Muzindutsi, P.-F. (2021). Application of artificial neural networks in predicting financial distress in the JSE financial services and manufacturing companies. *Journal of Sustainable Finance & Investment*. <https://doi.org/10.1080/20430795.2021.2017257>
- Farooq, U., Jibrán Qamar, M. A., & Haque, A. (2018). A three-stage dynamic model of financial distress. *Managerial Finance*, 44(9), 1101–1116. <https://doi.org/10.1108/MF-07-2017-0244>
- Fernando, J. M. R., Li, L., & Hou, G. (2020). Financial versus non-financial information for default prediction: Evidence from Sri Lanka and the USA. *Emerging Markets Finance and Trade*, 56(3), 673–692. <https://doi.org/10.1080/1540496X.2018.1545644>
- FlorezLopez, R., & Ramon-Jeronimo, J. M. (2015). Enhancing accuracy and interpretability of ensemble strategies in credit risk assessment. A correlated-adjusted decision forest proposal. *Expert Systems with Applications*, 42(13), 5737–5753. <https://doi.org/10.1016/j.eswa.2015.02.042>
- Gao, B., & Balyan, V. (2022). Construction of a financial default risk prediction model based on the LightGBM algorithm. *Journal of Intelligent Systems*, 31(1), 767–779. <https://doi.org/10.1515/jisys-2022-0036>
- Garg, A., & Tai, K. (2013). Comparison of statistical and machine learning methods in modelling of data with multicollinearity. *International Journal of Modelling, Identification and Control*, 18(4), 295–312. <https://doi.org/10.1504/IJMIC.2013.053535>
- Garg, V., Aggarwal, S., Tiwari, P., & Chatterjee, P. (Eds.). (2021). Applications of Artificial Intelligence in Business and Finance: *Modern Trends*. CRC Press

- Gilbert, L. R., Menon, K., & Schwartz, K. B. (1990). Predicting bankruptcy for firms in financial distress. *Journal of Business Finance & Accounting*, 17(1), 161–171. <https://doi.org/10.1111/j.1468-5957.1990.tb00555.x>
- Graham, J. R., & Harvey, C. R. (2001). The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics*, 60(2–3), 187–243.
- Grishunin, S., Bukreeva, A., & Astakhova, A. (2022). Analysing the determinants of insolvency and developing the rating system for Russian insurance companies. *Procedia Computer Science*, 199, 190–197.
- Halim, Z., & Mohamed Shuhidan, S. (2022). Towards development of robust machine learning model for Malaysian corporation: A systematic review of essential aspects for corporate credit risk assessment. *Malaysian Journal of Computing (MJoC)*, 7(1), 1011–1126.
- HasnainManzoor, D., & Shafi, K. (2022). Does corporate governance and corporate social responsibility affect earnings management practices? *Journal of Positive School Psychology*, 6(11), 3229–3239.
- Henley, W. E. (1996). Nearest neighbor analysis in credit scoring pdf. *The Statistician*, 45(1), 77–95.
- Huang, S. C., Wu, C. F., Chiou, C. C., & Lin, M. C. (2022). Intelligent FinTech data mining by advanced deep learning approaches. *Computational Economics*, 59(4), 1407–1422. <https://doi.org/10.1007/s10614-021-10118-5>
- Inam, F., Inam, A., Mian, M. A., Sheikh, A. A., & Awan, H. M. (2019). Forecasting bankruptcy for organizational sustainability in Pakistan. *Journal of Economic and Administrative Sciences*, 35(3), 183–201. <https://doi.org/10.1108/jeas-05-2018-0063>
- Jahur, M. S., & Quadir, S. N. (2012). Financial distress in small and medium enterprises (SMEs) of Bangladesh: Determinants and remedial measures. *Economia. Seria Management*, 15(1), 46–61.
- Jarrow, R. A., & Purnanandam, A. K. (2005). A generalized coherent risk measure: The firm's perspective. *Finance Research Letters*, 2(1), 23–29. <https://doi.org/10.1016/j.frl.2004.11.001>
- Javid, M. I., & Javid, A. Y. (2018). Efficacy of going concern prediction model for creditor oriented regime via liquidation: A MDA approach. *Journal of Applied Accounting Research*, 19(4), 552–573. <https://doi.org/10.1108/JAAR-07-2017-0070>
- Jones, S. (2023). A literature survey of corporate failure prediction models. *Journal of Accounting Literature*, 45(2), 364–405.
- Karas, M., & Reznakova, M. (2020). Cash flows indicators in the prediction of financial distress. *Engineering Economics*, 31(5), 525–535. <https://doi.org/10.5755/j01.ee.31.5.25202>
- Kato, K., & Nakamura, N. (2024). PDE-based Bayesian inference of CEV dynamics for credit risk in stock prices. *Asia-Pacific Financial Markets*, 31(2), 389–421.
- Kaur, S., Aggarwal, S., & Arora, V. (2024). Co Movement of Stock Market of BRICS with G7 Stock Market. *Asia-Pacific Financial Markets*. <https://doi.org/10.1007/s10690-024-09455-w>
- Kaur, S., Aggarwal, S., & Sarwar, S. (2023). Trade balance, monetary supply, commodity prices, and greener energy growth: Contextual evidence from BRICS economies in the lens of sustainability. *Environmental Science and Pollution Research*, 30(29), 73928–73940.
- Khemakhem, S., Ben Said, F., & Boujelbene, Y. (2018). Credit risk assessment for unbalanced datasets based on data mining, artificial neural network and support vector machines. *Journal of Modelling in Management*, 13(4), 932–951. <https://doi.org/10.1108/JM2-01-2017-0002>
- Khoja, L., Chipulu, M., & Jayasekera, R. (2019). Analysis of financial distress cross countries: Using macroeconomic, industrial indicators and accounting data. *International Review of Financial Analysis*, 66, 101379. <https://doi.org/10.1016/j.irfa.2019.101379>
- Kim, H. J. (2022). Semi-supervised learning to predict default risk for P2P lending. *Journal of Digital Convergence*, 20(4), 185–192.
- King, S., Leaney, C., & Downing, R. (2024). The use of credit risk weights for climate-related purposes. *Reserve Bank of New Zealand Bulletin*, 87, 1–19.
- Lee, T. S., & Yeh, Y. H. (2004). Corporate governance and financial distress: Evidence from Taiwan. *Corporate Governance: An International Review*, 12(3), 378–388.
- Liang, D., Lu, C. C., Tsai, C. F., & Shih, G. A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, 252(2), 561–572. <https://doi.org/10.1016/j.ejor.2016.01.012>
- Lim, H. J., & Mali, D. (2024). Does market performance (Tobin's Q) have a negative effect on credit ratings? Evidence from South Korea. *Asia-Pacific Financial Markets*, 31(1), 53–80.
- Lin, F., Liang, D., & Chu, W. S. (2010). The role of non-financial features related to corporate governance in business crisis prediction. *Journal of Marine Science and Technology*, 18(4), 504–513. <https://doi.org/10.51400/2709-6998.1901>

- Lin, W. C., Lu, Y. H., & Tsai, C. F. (2019). Feature selection in single and ensemble learning-based bankruptcy prediction models. *Expert Systems*, 36(1), 1–8. <https://doi.org/10.1111/exsy.12335>
- Lu, T., Chen, Z., Zhang, Y., & Lu, X. (2022). A bright side of anger: A comprehensive exploration of Borrowers' emotions and credit risk management. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4131102>
- Luo, C. (2022). A comparison analysis for credit scoring using bagging ensembles. *Expert Systems*, 39(2), 1–7. <https://doi.org/10.1111/exsy.12297>
- Maina, S. N. (2020). Board characteristics and financial distress of listed commercial banks in Kenya. 1–104.
- Malhotra, D. K., Malhotra, K., & Malhotra, R. (2020). Evaluating consumer loans using machine learning techniques. *Applications of Management Science* (pp. 59–69). Emerald Publishing Limited.
- Mitchell, T. M. (1997). Does machine learning really work? *AI Magazine*, 18(3), 11.
- Muñoz-Izquierdo, N., Segovia-Vargas, M. J., & del Camacho-MiñanoMPascual-Ezama, M. D. (2019). Explaining the causes of business failure using audit report disclosures. *Journal of Business Research*, 98(July), 403–414. <https://doi.org/10.1016/j.jbusres.2018.07.024>
- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1), 90–100. [https://doi.org/10.1016/S0022-2496\(02\)00028-7](https://doi.org/10.1016/S0022-2496(02)00028-7)
- Nam, C. W., Kim, T. S., Park, N. J., & Lee, H. K. (2008). Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *Journal of Forecasting*, 27(6), 493–506. <https://doi.org/10.1002/for.985>
- Obradovic, D. B., JakaiRupi, D. I. B., & M Andri. (2018). Insolvency prediction model of the company: The case of the republic of serbia. *Economic Research-Ekonomska Istrazivanja*, 31(1), 138–157. <https://doi.org/10.1080/1331677X.2017.1421990>
- Ogachi, D., Ndege, R., Gaturu, P., & Zoltan, Z. (2020). Corporate bankruptcy prediction model, a special focus on listed companies in Kenya. *Journal of Risk and Financial Management*, 13(3), 47. <https://doi.org/10.3390/jrfm13030047>
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109–131.
- Outecheva, N. (2007). Corporate financial distress: An empirical analysis of distress risk (Doctoral dissertation, Verlag nicht ermittelbar)
- Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting*, 36(3), 1092–1113. <https://doi.org/10.1016/j.ijforecast.2019.11.005>
- Ragab, Y. M., & Saleh, M. A. (2021). Non-financial variables related to governance and financial distress prediction in SMEs-evidence from Egypt. *Journal of Applied Accounting Research*. <https://doi.org/10.1108/JAAR-02-2021-0025>
- Rahmani, R., Parola, M., & Cimino, M. G. (2024). A machine learning workflow to address credit default prediction. *arXiv preprint arXiv:2403.03785*
- Ranawat, N. S., & Chakraborty, A. (2024). The impact of third-party financial products on the consumer loan services market in the banking sector: An analysis of sales progress and consumer behavior. *Asia-Pacific Financial Markets*, 31(2), 367–387.
- Ravi Kumar, P., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review. *European Journal of Operational Research*, 180(1), 1–28. <https://doi.org/10.1016/j.ejor.2006.08.043>
- Rutkowski, L., Pietruczuk, L., Duda, P., & Jaworski, M. (2012). Decision trees for mining data streams based on the McDiarmid's bound. *IEEE Transactions on Knowledge and Data Engineering*, 25(6), 1272–1279.
- Sariev, E., & Germano, G. (2020). Bayesian regularized artificial neural networks for the estimation of the probability of default. *Quantitative Finance*, 20(2), 311–328. <https://doi.org/10.1080/14697688.2019.1633014>
- Shrivastava, A., Kumar, N., Kumar, K., & Gupta, S. (2020). Corporate distress prediction using random forest and tree net for india. *Journal of Management and Science*, 10(1), 1–11.
- Sigrist, F., & Leuenberger, N. (2023). Machine learning for corporate default risk: Multi-period prediction, frailty correlation, loan portfolios, and tail probabilities. *European Journal of Operational Research*, 305(3), 1390–1406.
- Stopher, P. R., MeyburgAShrivastava, S., Jeyanthi, P. M., & Singh, S. (2020). Failure prediction of Indian Banks using SMOTE, Lasso regression, bagging and boosting. *Cogent Economics and Finance*. <https://doi.org/10.1080/23322039.2020.1729569>

- Sun, J., Fujita, H., Zheng, Y., & Ai, W. (2021). Multi-class financial distress prediction based on support vector machines integrated with the decomposition and fusion methods. *Information Sciences*, 559, 153–170. <https://doi.org/10.1016/j.ins.2021.01.059>
- Tang, L., Cai, F., & Ouyang, Y. (2019). Applying a nonparametric random forest algorithm to assess the credit risk of the energy industry in China. *Technological Forecasting and Social Change*, 144, 563–572. <https://doi.org/10.1016/j.techfore.2018.03.007>
- Thomas, J. D., & Sycara, K. (2000). Integrating genetic algorithms and text learning for financial prediction. *Data Mining with Evolutionary Algorithms*, 72–75
- Tsai, C.-F., Sue, K.-L., Hu, Y.-H., & Chiu, A. (2021). Combining feature selection, instance selection, and ensemble classification techniques for improved financial distress prediction. *Journal of Business Research*, 130(300), 200–209. <https://doi.org/10.1016/j.jbusres.2021.03.018>
- Tsai, K. L., Tomomori-Sato, C., Sato, S., Conaway, R. C., Conaway, J. W., & Asturias, F. J. (2014). Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. *Cell*, 157(6), 1430–1444. <https://doi.org/10.1016/j.cell.2014.05.015>
- Weitzel, W., & Jonsson, E. (1989). Decline in organizations: A literature integration and extension. *Administrative Science Quarterly*, 34(1), 91. <https://doi.org/10.2307/2392987>
- West, D. (2000). Neural network credit scoring models. *Computers and Operations Research*, 27(11–12), 1131–1152. [https://doi.org/10.1016/S0305-0548\(99\)00149-5](https://doi.org/10.1016/S0305-0548(99)00149-5)
- Wod, I. J. (1985). Weight of evidence: A brief survey. *Bayesian Statistics*, 2, 249–270.
- Wu, D., & Olson, D. L. (2008). Supply chain risk, simulation, and vendor selection. *International Journal of Production Economics*, 114(2), 646–655. <https://doi.org/10.1016/j.ijpe.2008.02.013>
- Yeh, K. H., Su, C., Lo, N. W., Li, Y., & Hung, Y. X. (2010). Two robust remote user authentication protocols using smart cards. *Journal of Systems and Software*, 83(12), 2556–2565. <https://doi.org/10.1016/j.jss.2010.07.062>
- Zhou, L., Lu, D., & Fujita, H. (2015). The performance of corporate financial distress prediction models with features selection guided by domain knowledge and data mining approaches. *Knowledge-Based Systems*, 85, 52–61.
- Zhu, Y., Zhou, L., Xie, C., Wang, G. J., & Nguyen, T. V. (2019). Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. *International Journal of Production Economics*, 211, 22–33. <https://doi.org/10.1016/j.ijpe.2019.01.032>
- Zhu, X., Chu, Q., Song, X., Hu, P., & Peng, L. (2023). Explainable prediction of loan default based on machine learning models. *Data Science and Management*, 6(3), 123–133.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.