

Article

Advancing Credit Rating Prediction: The Role of Machine Learning in Corporate Credit Rating Assessment

Nazário Augusto de Oliveira *  and Leonardo Fernando Cruz Basso 

Department of Business Administration—Strategic Finance, Mackenzie Presbyterian University (UPM), São Paulo 01302-907, Brazil; leonardofernando.basso@mackenzie.br

* Correspondence: nazario.oliveira@mackenzista.com.br; Tel.: +55-11-95040-5870

Abstract: Accurate corporate credit ratings are essential for financial risk assessment; yet, traditional methodologies relying on manual evaluation and basic statistical models often fall short in dynamic economic conditions. This study investigated the potential of machine-learning (ML) algorithms as a more precise and adaptable alternative for credit rating predictions. Using a seven-year dataset from S&P Capital IQ Pro, corporate credit ratings across 20 countries were analyzed, leveraging 51 financial and business risk variables. The study evaluated multiple ML models, including Logistic Regression, Support Vector Machines, Decision Trees, Random Forest, Gradient Boosting (GB), and Neural Networks, using rigorous data pre-processing, feature selection, and validation techniques. Results indicate that Artificial Neural Networks (ANN) and GB consistently outperform traditional models, particularly in capturing non-linear relationships and complex interactions among predictive factors. This study advances financial risk management by demonstrating the efficacy of ML-driven credit rating systems, offering a more accurate, efficient, and scalable solution. Additionally, it provides practical insights for financial institutions aiming to enhance their risk assessment frameworks. Future research should explore alternative data sources, real-time analytics, and model explainability to facilitate regulatory adoption.

Keywords: machine learning; credit ratings; predictive modeling; financial risk assessment; risk management



Academic Editor: Zhifang He

Received: 15 April 2025

Revised: 6 June 2025

Accepted: 11 June 2025

Published: 17 June 2025

Citation: de Oliveira, Nazário Augusto, and Leonardo Fernando Cruz Basso. 2025. Advancing Credit Rating Prediction: The Role of Machine Learning in Corporate Credit Rating Assessment. *Risks* 13: 116. <https://doi.org/10.3390/risks13060116>

Copyright: © 2025 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

Credit ratings play a pivotal role in financial markets by providing standardized assessments of a firm's creditworthiness, shaping investment decisions, influencing risk management strategies, and determining regulatory compliance obligations (Kou et al. 2019). These ratings directly affect borrowing costs, bond yields, and market perceptions, thereby reinforcing overall market stability.

Traditionally, credit rating agencies (CRAs) have relied on a combination of financial ratio analysis and qualitative judgment when assigning ratings. This conventional approach typically involves assessing key financial indicators—such as leverage, liquidity, and profitability—alongside non-financial factors, including management quality and macroeconomic trends (Egan-Jones Ratings 2024; Investment Grade Capital 2024). However, these methodologies exhibit inherent limitations, notably the assumption of linear relationships between variables and the potential to overlook complex, non-linear dynamics prevalent in modern financial markets. Furthermore, regulatory inconsistencies and enforcement challenges continue to impede the effective supervision of CRAs, particularly in emerging markets (Rabinowitz et al. 2024).

In response to these challenges, machine learning (ML) and artificial intelligence (AI) have emerged as transformative tools in credit risk assessment. ML algorithms excel at analyzing vast and complex datasets, enabling the detection of intricate patterns that traditional statistical models may fail to capture (Kou et al. 2019). This capability has led to substantial improvements in predictive performance across various financial applications, such as bankruptcy prediction (Barboza et al. 2017) and credit card default forecasting (Chang et al. 2024).

The incorporation of alternative data sources has further enhanced the predictive capabilities of credit risk models. Integrating non-traditional information enables more nuanced and timely credit evaluations, providing a multidimensional view of borrower risk that extends beyond conventional financial statements (Lu et al. 2019).

Despite these advancements, critical gaps persist in the literature. Notably, there is a scarcity of comprehensive studies systematically comparing the performance and robustness of different ML algorithms across diverse market conditions and industry sectors (Kou et al. 2019). Moreover, concerns regarding the explainability and fairness of AI-driven models are increasingly prominent. Regulatory authorities mandate that financial institutions ensure transparency, accountability, and governance in their credit risk models, particularly when employing complex ML algorithms (European Banking Authority 2021).

Recent scholarly work has further underscored these concerns. Enhancing both the accuracy and fairness of AI applications in credit risk assessment is crucial for promoting trust and stakeholder acceptance (Z. Wang 2024). Additionally, interpretable models should be preferred over black-box approaches in high-stakes decision-making contexts, such as credit rating, to mitigate risks and ensure regulatory compliance (Rudin 2019).

To address these challenges, this study conducts a rigorous comparative evaluation of multiple ML algorithms applied to corporate credit rating prediction. Specifically, it aims to assess the predictive performance of widely used models—including Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVMs), Neural Networks (NN), and Gradient Boosting Machines (GBM)—against traditional credit rating approaches. Moreover, the study seeks to identify and interpret the most influential factors affecting credit ratings through advanced feature importance analysis.

The robustness of these models will be tested across various economic contexts to evaluate their practical applicability and generalizability. This consideration is particularly important given that a deeper understanding of generalization remains critical for the reliable deployment of deep-learning models (Zhang et al. 2021).

To support these objectives, the study utilizes a comprehensive dataset covering seven years of corporate credit ratings across 20 countries sourced from S&P Capital IQ Pro. The dataset comprises 51 variables, with 43 related to financial risk and 8 to business risk. Rigorous data pre-processing procedures—including outlier detection, normalization, and treatment of missing values—are applied to enhance data quality. Feature selection follows established best practices, notably the recursive feature elimination method, which has proven effective in isolating the most predictive variables in high-dimensional datasets (Guyon et al. 2002).

Each selected ML algorithm is trained using stratified k-fold cross-validation to prevent overfitting and ensure model generalizability. Performance evaluation is conducted across a suite of metrics—accuracy, precision, recall, F1 score, and ROC-AUC—aligned with recommended evaluation standards for ML model assessments (Biecek and Burzykowski 2021).

This research makes several key contributions to the field. First, it offers an extensive comparative analysis of ML algorithms for corporate credit rating prediction, directly addressing gaps identified in the literature (Kou et al. 2019). Second, it incorporates both

traditional financial indicators and alternative data sources, providing new insights into the determinants of corporate creditworthiness. Finally, it responds to regulatory and ethical imperatives by embedding explainability and fairness considerations into model selection and evaluation (European Banking Authority 2021; Rudin 2019; Y. Wang 2024).

The remainder of this paper presents a review of the literature on ML applications in credit rating prediction, followed by the research methodology detailing model selection and evaluation metrics. Subsequently, the empirical findings and their implications are discussed, concluding with recommendations for financial practitioners, regulators, and future research.

2. Literature Review

Credit rating prediction has long been a critical area of financial research due to its role in assessing borrowers' creditworthiness. Traditional approaches relied heavily on statistical models, which laid the foundation for systematic risk assessment. A seminal contribution is Altman's (1968) Z-score model, which used multiple discriminant analysis (MDA) to combine financial ratios into a singular risk metric for bankruptcy prediction. Merton's (1974) structural model further advanced the field by applying option pricing theory, conceptualizing a company's equity as a call option on its assets to estimate default probabilities.

However, traditional credit rating models exhibit notable limitations. Their dependence on historical financial data and accounting-based ratios constrains their ability to capture complex, non-linear relationships and adapt to real-time shifts in a borrower's financial condition, undermining predictive accuracy (Cheng et al. 2024; Umeaduma and Adedapo 2025). Additionally, these models often rely on simplifying assumptions about the distribution of financial ratios and market conditions—assumptions that can fail during periods of economic turbulence or firm-specific distress (Addy et al. 2024). A further limitation lies in their predominantly binary classification framework, which focuses solely on default versus non-default outcomes, neglecting nuanced credit rating migrations essential for risk-sensitive applications (Addy et al. 2024; Jagtiani and Lemieux 2019).

To address these shortcomings, enhanced statistical methods emerged. Linear regression models established relationships between credit ratings and variables such as financial ratios and economic indicators. For instance, Kaplan and Urwitz (1979) demonstrated the effectiveness of linear regression in predicting bond ratings. Subsequently, LR became a staple for binary classification in credit risk contexts, notably through Ohlson's (1980) pioneering application for bankruptcy prediction. Nevertheless, LR is limited by its assumption of linearity in the logit function and susceptibility to multicollinearity among independent variables.

Recognizing these constraints, CRAs began integrating quantitative data with expert judgment. Qualitative factors—such as management quality, governance practices, and industry dynamics—play a critical role in bond ratings, contributing to more comprehensive and forward-looking assessments. However, this integration introduces risks of subjectivity and potential inconsistencies that can compromise model objectivity (Seetharaman et al. 2017).

In recent years, ML and AI have significantly transformed credit risk modeling by offering advanced tools for data analysis and prediction. ML algorithms excel at identifying complex, non-linear patterns in large datasets, thereby addressing many limitations inherent in traditional models (Talaie Khoei and Kaabouch 2023). Talaie Khoei and Kaabouch (2023) provided a comprehensive review of various ML methods—including NN, SVMs, and ensemble learning techniques—highlighting their advantages in enhancing predictive performance and robustness across financial applications. Further exemplifying these advancements, Chang et al. (2024) demonstrated that deep-learning architectures such

as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) outperform conventional models in credit card default prediction, particularly in handling sequential and structured financial data.

The growing integration of advanced ML techniques into credit rating prediction reflects a broader trend toward more sophisticated analytical tools. For example, [Tran and Tham \(2025\)](#) showed that boosting ensembles and SVMs outperform traditional models like LR in financial risk evaluation owing to their superior capacity for modeling complex, non-linear relationships. Similarly, [Mokheleli and Museba \(2023\)](#) highlighted how NNs and ensemble methods enhance the predictive accuracy of credit score models by capturing intricate patterns within financial datasets.

Despite their superior predictive capabilities, ML models often face criticism for their “black-box” nature, which complicates interpretability and transparency—particularly in high-stakes financial decision-making. In response, recent research has prioritized the development of Explainable Artificial Intelligence (XAI) techniques, such as Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME). These methods improve the interpretability of model predictions and are essential for compliance with emerging regulatory frameworks, including the European Union’s AI Act and the Basel Committee’s guidelines on model risk management ([Hassija et al. 2024](#); [Sabharwal et al. 2024](#)).

[Y. Wang \(2024\)](#) emphasized that ensuring fairness and transparency in ML-based credit risk assessments is vital for promoting stakeholder trust and meeting regulatory requirements. Complementing this perspective, [Hassija et al. \(2024\)](#) reviewed a range of methods for interpreting black-box models, underscoring their growing relevance in financial contexts. Moreover, [Weber et al. \(2024\)](#) conducted a systematic review that highlighted the increasing importance of XAI in finance, particularly for enhancing model accountability and mitigating risks associated with opaque decision-making processes.

In addition to advances in ML techniques, recent research has emphasized the importance of non-financial and strategic variables in credit risk and default prediction—particularly for SMEs. Traditional financial indicators, while foundational, may be insufficient to capture the nuanced factors that influence a firm’s creditworthiness and recovery potential.

For instance, [Altman et al. \(2022, 2023\)](#) introduced the Omega Score, a revised predictive model that integrates both financial and qualitative variables for SME default prediction, significantly improving model performance and robustness. These studies stress the value of including variables related to management quality, market positioning, and strategic behavior.

This perspective is further reinforced in sector-specific contexts. [Srhoj et al. \(2024\)](#) demonstrate that in the tourism industry, predictors of SME default differ from general models, underscoring the need for industry-tailored frameworks. Their findings highlight the unique operational and demand-side dynamics that impact firms in this sector.

Furthermore, recovery-oriented research by [Altman et al. \(2024\)](#) explores the drivers of turnaround success, finding that qualitative factors—such as managerial flexibility and strategic refocusing—are crucial for explaining why some distressed firms rebound while others do not. This work suggests a growing consensus around the integration of strategic, behavioral, and sector-specific dimensions in predictive modeling.

Parallel to advancements in interpretability, the incorporation of alternative data sources has emerged as a significant trend in credit risk modeling. Traditional financial indicators are now frequently supplemented by non-traditional data such as environmental, social, and governance (ESG) metrics, supply chain information, and social media sentiment. [Lu et al. \(2019\)](#) provided empirical evidence demonstrating that alternative data

significantly improves the predictive accuracy of credit risk models by offering a more comprehensive and multidimensional perspective on borrower risk.

The influence of fintech and digital footprints in credit scoring is further evidenced by studies such as those by [Jagtiani and Lemieux \(2019\)](#) and [Berg et al. \(2020\)](#). Their findings reveal that alternative data enhances model performance by capturing forward-looking indicators typically overlooked by traditional models. However, the integration of such data also raises complex challenges related to privacy, ethical considerations, and model validation, necessitating ongoing research and robust regulatory oversight ([Owoade et al. 2024](#)). In line with this, the [Basel Committee on Banking Supervision \(2023\)](#) has issued principles for managing climate-related financial risks, reinforcing the need to incorporate ESG-related alternative data in credit risk assessments.

The impact of ML extends beyond credit rating prediction to domains such as fraud detection and portfolio management. Fraud detection systems have evolved from rule-based mechanisms to sophisticated ML-driven models. Recent studies highlight the effectiveness of advanced ML approaches in identifying complex fraud patterns, with ensemble methods and synthetic sampling techniques contributing to improved classification accuracy ([Bello et al. 2023](#); [Ileberi et al. 2021](#)).

In portfolio management, ML has enhanced traditional models such as Modern Portfolio Theory ([Markowitz 1952](#)) and the Capital Asset Pricing Model ([Sharpe 1964](#)). For instance, [Krauss et al. \(2017\)](#) and [Gu et al. \(2020\)](#) applied deep NNs and gradient-boosted trees for stock return prediction, achieving significant improvements over conventional time-series forecasting methods.

Illustrating the broader applicability of ML techniques, the study “On the analytical study of the service quality of Indian Railways under soft-computing paradigm” demonstrates the versatility of approaches such as rough set theory, extra trees classifiers, and SVMs beyond financial contexts, reinforcing the value of these tools across diverse domains ([Majumder et al. 2024](#)).

While substantial progress has been made, future research should focus on enhancing the interpretability of ML models, integrating alternative data sources more effectively, and ensuring alignment with evolving regulatory frameworks. Additionally, systematic comparative studies evaluating the performance of different ML algorithms across various economic contexts and industry sectors remain relatively scarce ([Kou et al. 2019](#)).

The convergence of ML techniques with domain expertise and traditional financial theories offers significant potential for developing more robust and reliable credit risk models. Furthermore, advancing responsible AI principles—such as fairness, accountability, and transparency—is essential to harmonize technological innovation with societal and regulatory expectations. Recent contributions by [Barocas et al. \(2023\)](#) underscore the growing emphasis on fairness and causal approaches in credit scoring, providing valuable guidance for future research agendas.

2.1. Theoretical Framework

2.1.1. Credit Risk Theory

Credit risk theory forms the backbone of credit rating assessments, emphasizing the potential risk of loss resulting from a borrower’s failure to repay a loan. Robert Merton’s foundational works (1974) laid the groundwork for modern credit risk evaluation through the Merton model, which applies option pricing theory to corporate debt valuation. Merton’s model fundamentally reshaped the understanding of credit risk by treating corporate debt as a risk-free bond combined with a short put option on the firm’s assets, reflecting the probability of default.

Before [Merton \(1974\)](#) introduced his structural model, [Altman \(1968\)](#) developed the Z-score model, a statistical tool for predicting bankruptcy. Altman's model utilized financial ratios and discriminant analysis to estimate the likelihood of default, providing a practical and widely adopted framework for credit risk assessment. While Merton's approach focuses on market-based measures, Altman's model emphasizes accounting data, highlighting a divergence in methodology but a convergence in the goal of assessing credit risk.

[Jarrow and Turnbull \(1995\)](#) significantly advanced credit risk modeling by developing a reduced-form approach that incorporates the term structure of interest rates. Their model provides a market-based, dynamic method for estimating default probabilities and credit spreads, improving upon static accounting-based measures like the Altman Z-score. This approach underscores the importance of market-sensitive frameworks in credit risk assessment.

2.1.2. Machine Learning Theory

ML theories have significantly transformed credit rating prediction by introducing advanced algorithms capable of managing large-scale datasets and capturing complex, non-linear patterns. Foundational techniques such as DT models ([Breiman et al. 1984](#)) offer intuitive and interpretable structures but are inherently prone to overfitting. To mitigate these limitations, contemporary ensemble methods such as RF ([Breiman 2001](#)) and GBM ([Friedman 2001](#)) have been developed, followed by more sophisticated variants like XGBoost ([Chen and Guestrin 2016](#)), LightGBM ([Ke et al. 2017](#)), and CatBoost ([Prokhorenkova et al. 2018](#)), which deliver enhanced scalability and predictive accuracy.

The backpropagation algorithm ([Rumelhart et al. 1986](#)) remains a cornerstone of NN training, underpinning modern deep-learning architectures that adeptly model complex financial relationships. Recent empirical studies, such as that of [Tran and Tham \(2025\)](#), have demonstrated the efficacy of deep NNs in processing extensive historical credit data, thereby significantly improving predictive performance.

Despite these technological advances, the “black-box” nature of many ML models has raised concerns about transparency and interpretability, particularly in high-stakes financial decision-making contexts. This has catalyzed extensive research on XAI methodologies, notably SHAP ([Lundberg and Lee 2017](#)) and LIME, to ensure model interpretability and regulatory compliance. Regulatory frameworks, including the European Union's Artificial Intelligence Act ([European Commission 2023](#)) and the Basel Committee's guidance on model risk management, now underscore the imperative for interpretable models in financial applications.

Benchmarking studies focused on credit scoring and rating prediction have consistently validated the superiority of ensemble-based models over traditional statistical approaches ([Lessmann et al. 2015](#)), reinforcing their growing prominence in contemporary credit risk analytics. Moreover, recent comprehensive reviews on ensemble deep learning have emphasized their advantages in improving predictive performance across various domains, further highlighting their methodological robustness and broad applicability ([Mohammed and Kora 2023](#)).

2.1.3. Behavioral Finance Theory

Behavioral finance theory examines how psychological factors and biases influence financial decision-making, which can impact credit ratings. [Tversky and Kahneman's \(1974\)](#) prospect theory highlights how cognitive biases, particularly loss aversion, shape risk perceptions. Other behavioral biases, such as overconfidence and herding behavior, also influence investor decisions. These biases contribute to irrational market behavior,

influencing credit risk assessment through changes in market sentiment, risk perception, and investor-driven fluctuations in credit spreads, which can ultimately affect credit ratings.

Shiller (2003) explored the role of market sentiments and speculative bubbles in financial markets, underscoring the influence of behavioral factors on asset pricing and risk perception. Behavioral finance theories suggest that traditional models may overlook these irrational elements, leading to mispricing and inaccurate credit ratings.

ML can address these complexities by incorporating behavioral indicators into predictive models. By analyzing patterns in investor behavior and sentiment, ML algorithms can adjust for biases and improve the accuracy of credit risk predictions. This integration of behavioral finance with ML approaches represents a convergence of theories aimed at enhancing the robustness of credit rating methodologies.

2.1.4. Gaps in the Literature

Applying ML algorithms in credit rating prediction has gained significant attention recently. However, despite these advancements, the black-box nature of many ML models poses challenges for interpretability and regulatory acceptance. Researchers have developed XAI techniques to address these concerns that provide insights into model predictions. Ribeiro et al. (2016) introduced LIME, a method designed to explain the predictions of any classifier, enhancing the transparency of ML models. Similarly, Lundberg and Lee (2017) developed SHAP, a unified approach to interpreting model predictions, gaining traction in financial applications. These techniques have significantly improved interpretability, yet challenges remain in effectively integrating them into real-world financial decision-making.

Despite these advancements, several limitations persist in the literature. First, most studies have focused on static models that do not account for temporal changes in financial data. Second, data quality and availability remain significant challenges. Many studies rely on proprietary datasets that are not publicly accessible, limiting the reproducibility and generalizability of findings. Finally, as Doshi-Velez and Kim (2017) emphasized, while interest in XAI techniques continues to grow, achieving a balance between model accuracy and interpretability remains an ongoing challenge. Addressing these issues is crucial for adopting ML-based credit rating models in regulatory and financial environments.

Given these gaps, the current study aims to advance the field by developing a dynamic machine-learning framework for credit rating prediction that accounts for temporal changes in financial data. By utilizing publicly available datasets, this study seeks to enhance the reproducibility and generalizability of findings.

In addition to these technical limitations, recent empirical work has drawn attention to the absence of strategic and industry-specific considerations in many credit risk prediction models. For example, Altman et al. (2022, 2023) highlight the relevance of non-financial predictors—such as governance and competitive advantage—in improving SME default prediction. Likewise, Srhoj et al. (2024) show that sector-specific models yield better predictive accuracy by accounting for structural differences across industries. Altman et al. (2024) go a step further by identifying the qualitative and strategic determinants of firm recovery, advocating for a broader conceptualization of creditworthiness that includes the likelihood of turnaround.

By engaging with these perspectives, the present study aims to not only enhance methodological rigor through ML but also set a foundation for integrating qualitative and sector-sensitive factors into future predictive frameworks.

3. Methodology

3.1. Materials and Methods

This study focuses on credit rating prediction, a well-established research topic in financial risk assessment. Credit ratings categorize a client's financial condition into predefined classes, such as "A," "B," "C," or "D." These classifications are assigned based on financial and qualitative indicators using ML techniques.

A comprehensive dataset must be compiled from various sources to develop a predictive model. Relevant financial and qualitative features are extracted as predictors in a classification model. Credit rating prediction is framed as a classification problem, where a supervised learning model, trained on labeled data, assigns a credit rating to new clients based on their financial characteristics. Key financial indicators, such as the debt-to-income ratio, history of defaults, and credit inquiries, play a significant role in determining whether a client receives a "good" or "poor" rating.

The proposed methodology follows a structured approach, illustrated in Figure 1. First, financial and qualitative data are collected and processed to build a robust dataset, ensuring each feature represents a meaningful aspect of the company's risk profile. Since not all financial and qualitative attributes contribute equally to prediction accuracy, a two-stage feature refinement process is applied: (i) correlation analysis to remove highly correlated and redundant features and (ii) feature selection to exclude variables that may negatively impact classification performance. Finally, a ML classification model is trained on the selected features and used to predict credit ratings.

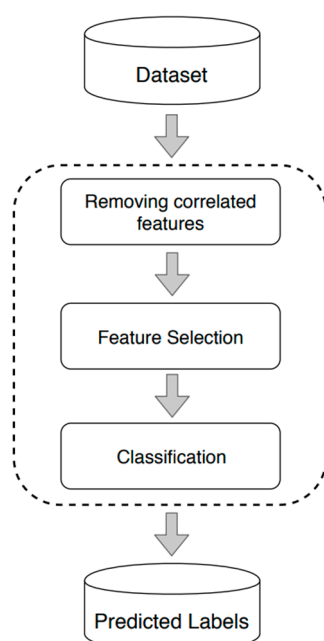


Figure 1. Methodological approach.

3.2. Removing Correlated Features

The previous section outlined the process of collecting, processing, and structuring data—comprising features and labels—as input for the classification step. However, certain features may exhibit high correlation, leading to redundant information that does not contribute unique predictive value. Removing such redundancy is essential for enhancing model interpretability, reducing overfitting, and improving classification accuracy.

In statistical analysis and ML, correlation is a fundamental metric for assessing feature relevance and redundancy. The linear correlation coefficient (ρ) is the most widely used measure to quantify the strength and direction of the relationship between two variables, X

and Y . It ranges from -1 to 1 , where $\rho = 1$ signifies a perfect positive correlation—indicating that as one variable increases, the other increases proportionally. Conversely, $\rho = -1$ represents a perfect negative correlation, where an increase in one variable corresponds to a proportional decrease in the other. A value of $\rho = 0$ suggests no linear relationship between the variables, implying that changes in one do not systematically correspond to changes in the other.

In the context of credit rating prediction, financial variables such as net income and EBITDA or current ratio and quick ratio are often highly correlated. Retaining both in the dataset may introduce redundancy, inflating the feature space without improving predictive performance. To mitigate this issue, correlation analysis is applied to identify and eliminate features with excessive collinearity, ensuring that only the most informative predictors are used in the classification model.

3.3. Feature Selection

The subsequent step in the proposed methodology (see Figure 1) is feature selection, which aims to identify and eliminate irrelevant and redundant information. This process reduces the dimensionality of the dataset, thereby improving computational efficiency and enhancing the performance of ML algorithms by focusing on the most informative features (Cheng 2024). From an ML perspective, feature selection in classification tasks offers several key benefits. It enhances model accuracy by removing superfluous features, leading to more compact and computationally efficient models. Furthermore, it facilitates knowledge discovery by isolating the most influential variables, thereby improving interpretability and providing deeper insights into the underlying data structure (Büyükeğeci and Okur 2023).

Feature selection techniques are typically categorized into three classes based on their interaction with the classification model: filter methods, wrapper methods, and embedded methods (Al-shalif et al. 2024). Filter methods operate independently of the learning algorithm, selecting features based on intrinsic data characteristics during a pre-processing phase. Conversely, wrapper and embedded methods incorporate feature selection within the model training process. Wrapper methods evaluate various feature subsets using a search strategy informed by the model's predictive performance, whereas embedded methods perform feature selection intrinsically as part of the model's parameter optimization during training (Cheng 2024).

This study employed wrapper methods due to their balance of interpretability and simplicity compared to embedded techniques. A commonly adopted search strategy within wrapper methods is greedy hill climbing, which iteratively modifies the current feature subset by either adding or removing one feature at a time. Variants of this strategy include forward selection, focusing exclusively on adding features, and backward elimination, which sequentially removes features. To overcome the limitations associated with these traditional approaches, recent advancements such as meta-heuristic algorithms (e.g., genetic algorithms, particle swarm optimization) and floating search methods have been introduced, enabling more flexible and dynamic inclusion and exclusion of features during the selection process (Al-shalif et al. 2024; Nemati et al. 2024). Moreover, hybrid approaches that combine multiple feature selection strategies have demonstrated promising results, achieving an optimal trade-off between computational efficiency and model performance (Nemati et al. 2024).

3.4. Classification

The proposed approach's final and most critical stage involves using classifiers to develop a classification model and assessing its performance through standard evaluation metrics, as explained in the next section.

3.4.1. Classification in Credit Rating Prediction

Classification is the process of predicting the target class of a given instance based on a set of input features. In financial applications, classification models are crucial in assessing credit risk, particularly in predicting corporate credit ratings, default probabilities, and borrower creditworthiness. For example, using financial information about a company, a classification model can predict the company's credit rating (the target class). In this context, the company is labeled with the value of the target class (e.g., credit rating). The term "label" will be used throughout this paper to refer to the target class, following standard ML terminology.

Building a classification model involves learning a mapping from input features to output labels. The training data used to develop the model contains labels corresponding to each training company, and the model's objective is to predict the label for input values corresponding to companies not included in the training data. A predefined set of features, such as profitability ratios, leverage, liquidity, and macroeconomic indicators, characterizes each company. However, financial classification tasks introduce unique challenges, including class imbalance, where highly rated firms significantly outnumber defaulted firms, and temporal dependencies, as creditworthiness evolves (Altman and Saunders 1998; Duffie and Singleton 2003).

Numerous classification algorithms are currently available, including traditional statistical methods such as LR and more advanced ML techniques like SVMs, NN, DT, GBM, and RF. However, it is impossible to definitively state that any algorithm is superior, as its effectiveness depends on the specific application and the nature of the dataset being used.

3.4.2. Evaluation Criteria

Evaluating classification models in credit risk applications requires the use of performance metrics tailored to financial decision-making. Common metrics include Area Under the ROC Curve (AUC-ROC), Gini coefficient, and F1 score, which help assess model discrimination power and balance between precision and recall. Furthermore, given the regulatory requirements in financial markets, model explainability remains a significant concern, particularly for ML techniques that operate as black-box models (Kuiper et al. 2022).

Despite advancements in ML, challenges remain in applying classification techniques to financial risk modeling. Regulatory compliance, explainability requirements, and the dynamic nature of financial markets necessitate careful feature selection and robustness testing.

4. Experiments and Results

In Section 3.1, we describe our methodology (summarized in Figure 1), which leverages financial and business risk features to construct a training dataset. This dataset is then used to train a classification model designed to predict the credit rating of companies. To enhance classification performance, we apply data analysis techniques, such as feature correlation analysis and feature selection.

Section 4.1 provides a detailed overview of the training dataset, focusing on the specific features we collected. In Section 4.2, we discuss the classification models employed in our study, emphasizing their role in credit rating prediction. The results of our methodology are presented in Section 4.3, where we analyze feature correlations within the dataset and identify the most important features contributing to classification performance.

4.1. Training Dataset

The dataset used in this study, presented in Table 1, was obtained from S&P Capital IQ Pro and consists of 51 features, including 43 financial features and 8 business features, covering rated companies across 20 countries. We collected financial and qualitative variables

from Capital IQ Pro and Bloomberg for 3453 companies over the period 2018–2024. The feature selection was based on data availability and relevance to credit rating assessment, ensuring a comprehensive representation of financial and business characteristics.

Table 1. Financial features used in the dataset for credit rating prediction. These variables capture key aspects of company performance, such as liquidity, profitability, leverage, and cash flow, and were used as input for the machine-learning models.

Financial Features	
Accessible Cash	EBITDA Margin (%)
Adjusted Capital	Equity
ARO Debt Adjustment	FFO Cash Interest Coverage Ratio (x)
Capex/Revenues (%)	FFO/Debt (%)
Capital Expenditures	FOCF/Debt (%)
Capital Structure	Free Operating Cash Flow
Cash	Funds from Operations (FFO)
Cash Flow from Operations	Gross Margin
Cash Interest Paid	Interest Expense
Cash Taxes Paid	Liquidity
CFO/Debt	Net Income
DCF/Debt	Pension and Postretirement Debt
Debt	R&D/Revenues
Debt/Debt and Equity	Reported Lease Liabilities
Debt/EBITDA	Return on Capital
Discretionary Cash Flow	Revenues
Dividends	SG&A/Revenue
EBIT	Share Repurchases
EBIT Interest Coverage	Total Operating Expense Before D&A
EBIT Margin	Volatility to Cash Flow/Leverage
EBITDA	Working Capital
EBITDA Interest Coverage	
Business features	
Distinctive Advantage	Governance
Country Risk	Operating Efficiency
Diversification	Size, Breadth, and Variety
Industry Risk	Volatility of Profitability

Source: Capital IQ Pro.

In this study, we utilize the full range of S&P Global credit rating grades, which includes 23 ordered levels, ranging from AAA (highest creditworthiness) to D (default). Market participants often group ratings of CCC+ and below due to their elevated default risk and financial vulnerability, reflecting a heightened probability of credit distress.

When dealing with categorical dependent variables such as credit ratings in ML models, each algorithm employs distinct techniques for processing these features. The following section provides a detailed analysis of how LR, SVMs, NN, DT, Gradient Boosting (GB), and RF handle credit ratings as a categorical target variable.

4.2. Classifiers

In this study, the target variable consists of multiple credit rating categories, necessitating the use of classifiers capable of handling multiclass classification. To ensure methodological diversity and capture different aspects of credit rating prediction, we selected six classifiers spanning linear models, tree-based models, and NNs.

Linear models, such as LR and SVMs, were chosen for their interpretability and ability to model direct relationships between financial variables and risk.

Tree-based models, including DT, RF, and GB, were selected for their ability to capture non-linear interactions, provide feature-importance insights, and maintain robustness when handling high-dimensional data.

Finally, NNs were incorporated due to their capacity to model highly complex, non-linear relationships without assuming a predefined decision boundary, making them well-suited for capturing intricate financial patterns.

The following section provides a discussion of each classification method, emphasizing their capability to handle multiclass classification.

4.2.1. Logistic Regression

LR, commonly used for binary classification, can be extended to multiclass prediction tasks, such as forecasting multiple credit ratings, through approaches like One-vs-Rest (OvR) or One-vs-One (OvO). However, when the categories exhibit a natural ranking (e.g., AAA > AA > A > BBB), ordinal LR is a more suitable choice.

As demonstrated by [Goldmann et al. \(2024\)](#), ordinal LR explicitly accounts for the ordered structure of the dependent variable, thereby improving interpretability and producing more meaningful coefficient estimates compared to standard multinomial LR. This modeling approach ensures that the predicted probabilities adhere to the inherent ranking of credit ratings, leading to more accurate assessments in credit risk modeling.

4.2.2. Support Vector Machines

SVMs, originally designed for binary classification, have been effectively extended to handle multiclass classification through strategies such as OvO and OvR. These methods enable SVMs to manage complex rating systems involving multiple classes, ensuring scalability and reliable classification performance. [Kurbakov and Sulimova \(2024\)](#) provide a comprehensive comparative study on multiclass extensions for SVMs, demonstrating their effectiveness and practical applicability across diverse classification tasks. Their proposed Dual-Layer Smart Sampling SVM (DLSS-SVM) method enhances scalability and efficiency without compromising accuracy, reflecting contemporary advancements in multiclass SVM methodologies.

4.2.3. Neural Networks

NNs offer substantial flexibility in handling both binary and multiclass categorical outcomes, such as credit ratings. In multiclass classification, the SoftMax activation function is commonly applied in the output layer to produce a probability distribution over possible classes. Categorical variables are typically represented using one-hot encoding, where each category is converted into a binary vector. However, for high-cardinality categorical variables, embedding layers provide a more efficient alternative by mapping categories into continuous vector spaces. As highlighted by [Cao \(2024\)](#), embedding models efficiently transform sparse, high-dimensional inputs into dense, low-dimensional representations, thereby enhancing both memory efficiency and learning performance.

4.2.4. Decision Trees

DT inherently support multiclass classification as they recursively split data at each node, accommodating multiple class labels. Their flexibility in handling both categorical and continuous features make them useful in applications like credit rating prediction. [Mienye and Jere \(2024\)](#) emphasize that DT are a versatile and foundational method for classification, demonstrating significant effectiveness in managing complex decision-making tasks across various domains.

4.2.5. Gradient Boosting

Gradient Boosting (GB) can be extended to multiclass classification by adapting boosting algorithms to optimize loss functions for multiple target categories. [Friedman \(2001\)](#) introduced GBM and discussed their potential extensions beyond binary classification. This theoretical framework has been successfully operationalized in modern implementations such as XGBoost ([Chen and Guestrin 2016](#)), which efficiently supports multiclass classification tasks.

4.2.6. Random Forests

RF are an effective approach for multiclass classification, leveraging an ensemble of DT trained on random subsets of data and features. Their ensemble nature enhances robustness and predictive accuracy across classification tasks. [Breiman \(2001\)](#) introduced RF as a method that generalizes well across various classification settings, demonstrating their effectiveness, including in multiclass problems. This has been further exemplified in recent applications within intrusion detection systems ([Alharthi et al. 2025](#)).

4.3. Results and Discussion

As described in Section 3.1, not all features are relevant when constructing a model since sometimes they represent redundant information. Hence, two experiments were carried out: in Section 4.3.1, the model was tested after removing features that were highly correlated (redundant), whereas in Section 4.3.2, the model was tested using only a selected subset of the original features. This experiment aimed to determine whether a smaller but more relevant set of features could still provide good results.

4.3.1. Feature Correlation Analysis

To enhance model performance and reduce redundancy, we conducted a correlation analysis to identify highly correlated features. High correlation among independent variables can introduce multicollinearity, leading to unstable model estimates, increased computational complexity, and a higher risk of overfitting.

To mitigate these issues, we identified and excluded features with high pairwise correlation, ensuring that only the most informative variables were retained. This feature selection step aims to improve the model's generalization ability while maintaining computational efficiency.

Identification of Highly Correlated Features

The Pearson correlation coefficient was employed to assess the linear relationships between features and identify potential redundancy.

We used a three-step feature correlation analysis to assess and mitigate redundancy in the dataset,

1. **Identification of Top 10 Correlated Features**—For each numerical feature, we identified the ten features with the highest absolute correlation values, considering both positive and negative correlations. This process enabled the selection of the most strongly related feature pairs, helping to pinpoint variables that contributed to potential multicollinearity.
2. **Aggregation of Highly Correlated Features**—From the identified sets of highly correlated features, we determined which variables consistently appeared among the top correlated features across multiple other variables. For example, if a given feature was ranked among the top 10 correlated features for at least 90% of other variables, it was flagged as a candidate for exclusion to reduce redundancy.

3. **Computation of Mean Correlation Scores**—For the aggregated set of high-correlation candidates, we computed the mean absolute correlation score across all pairwise correlations. This metric provided a quantitative measure of each feature’s overall redundancy within the dataset, supporting an objective feature selection process.

The analysis produced a refined set of features that exhibited consistently high correlations with other variables, indicating potential redundancy. Table 2 presents the most frequently correlated features along with their mean absolute correlation scores, providing a quantitative assessment of their degree of redundancy.

Table 2. Features with the highest mean absolute correlations, indicating potential redundancy in the dataset.

Feature	Mean Absolute Correlation
EBITDA	0.81
Interest Expense	0.74
Funds from operations (FFO)	0.73
Net Income	0.69
Working capital	0.69
Cash flow from operations	0.65
Revenues	0.63
Competitive Advantage	0.56
Free operating cash flow	0.53
Capital expenditures	0.47

Source: Own authorship.

Following the statistical analysis, domain experts evaluated the candidate features to ensure alignment with the dataset’s business objectives. To enhance the predictive model’s efficiency and interpretability, redundant features were identified and eliminated through Pearson correlation analysis. Pearson’s correlation coefficient quantifies the linear association between two variables, with values ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation). Features exhibiting a strong absolute correlation (e.g., $|r|$ between 0.5 and 0.7) were considered for removal due to their potential to introduce multicollinearity, increase computational complexity, and reduce model generalizability by conveying overlapping information. However, the selection of an appropriate correlation threshold for feature elimination must be context-dependent, as moderate correlation does not always imply redundancy, particularly in complex predictive models where interactions between variables may hold informational value. Thus, domain expertise and additional statistical assessments, such as variance inflation factor (VIF) analysis, were employed to ensure that the feature selection process preserved essential explanatory power while mitigating multicollinearity concerns.

Consequently, Table 3 presents features that were excluded due to having a VIF greater than 10:

Table 3. Variance inflation factor (VIF) values showing multicollinearity among selected features.

Variable	VIF
EBITDA	49.5
Funds from operations (FFO)	46.2
Net Income	10.2
Free operating cash flow	14.1
Capital expenditures	20.7

Source: Own authorship.

By integrating both statistical correlation analysis and domain-specific expertise, the feature selection process was refined to enhance model robustness. The elimination of highly correlated variables mitigates the risk of overfitting, improves model interpretability, and enhances computational efficiency.

4.3.2. Credit Rating Analysis

In this experiment, we analyze the impact of feature selection on model performance. While Section 4.3.1 focused on eliminating highly correlated features, this section evaluates how removing low-importance features affects classification accuracy.

The dataset used for credit rating prediction, presented in Table 4, comprises 27,664 instances collected from 20 countries. However, the distribution of observations is highly imbalanced, with the United States contributing 16,374 instances (59% of the dataset), while Switzerland accounts for only 257 instances (1%). This imbalance extends to the distribution of credit ratings across countries, leading to significant variations in rating proportions. Such disparities pose challenges for model generalization and predictive performance.

To mitigate these issues and improve model robustness, a country-specific modeling approach was adopted. This strategy accounts for regional differences in economic conditions, regulatory environments, and credit risk factors, thereby enhancing the model's ability to capture country-specific patterns in credit ratings.

Rating Categorization and Balance Improvement

The original credit ratings were highly detailed, leading to an uneven distribution of data across rating categories. This imbalance could negatively impact the analysis by reducing statistical reliability. To address this, we grouped the ratings into four broader categories, ensuring a more balanced distribution while preserving the essential distinctions needed for meaningful interpretation. Table 5 presents the grouped ratings by category.

Figure 2 presents a stacked bar chart illustrating the distribution of Class A, B, C, and D ratings across various countries, showing the relative proportion of each rating within each country.

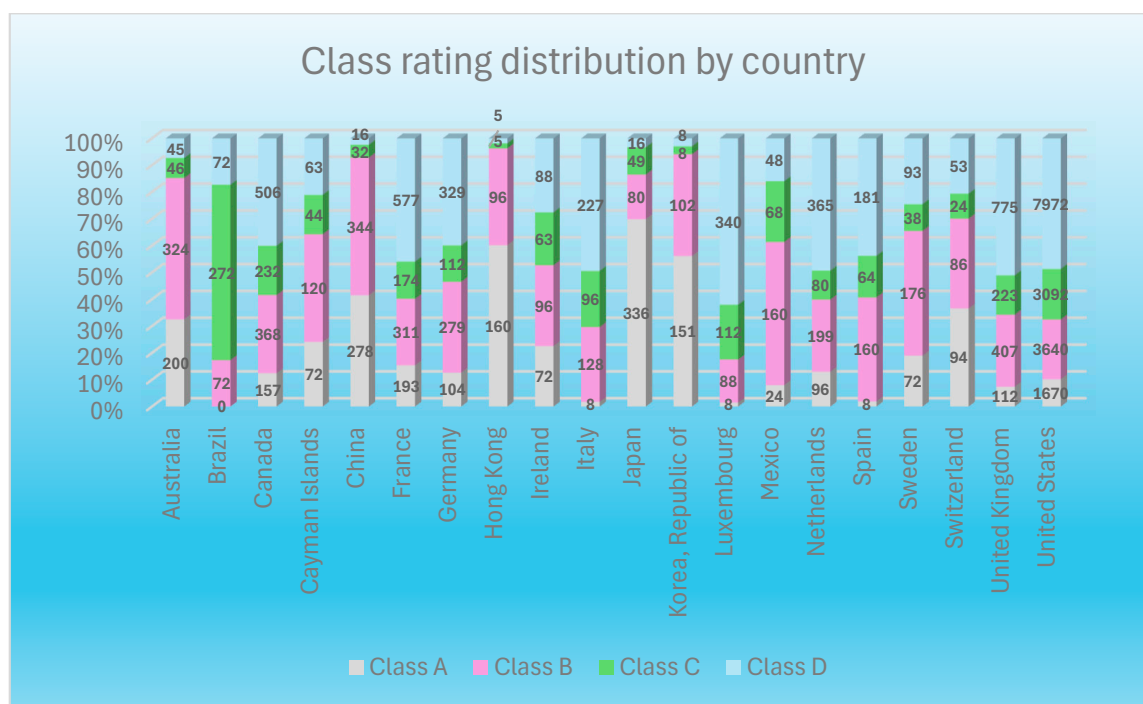


Figure 2. Class rating distribution by country.

Table 4. Distribution of credit rating observations by country and S&P rating categories from AAA to D.

S&P Ratings Grid	A	A-	A+	AA	AA-	AA+	AAA	B	B-	B+	BB	BB-	BB+	BBB	BBB-	BBB+	CC	CCC	CCC-	CCC+	D	SD	Total
Australia	32	144	8			8	8	14	23	8	14	8	24	136	56	132							615
Brazil								8	16	8	64	144	64	8	64						24		416
Canada	45	96	8		8			191	101	118	48	96	88	128	152	88		8		8	72	16	1263
Cayman Islands	24	32	16					8	31	16		36	8	40	32	48					8		299
China	71	135	72					8		8	8		24	152	96	96							670
France	48	105	8	16	16			275	182	88	39	63	72	72	112	127		8			24		1255
Germany	48	32	16		8			139	70	69	16	40	56	104	95	80	8	8			27	8	824
Hong Kong	32	56	40		8	24				5		5		16	40	40							266
Ireland	24	32	8		8			24	24	32	31	16	16	24	32	40					8		319
Italy		8						112	77	22	24	24	48	64	32	32					16		459
Japan	144	64	104		24					8	9	8	32	32		48					8		481
Korea, Republic of		31	8	104	8			8			8			32	32	38							269
Luxembourg		8						150	104	39	40	32	40	24	24	40	8	16			16	7	548
Mexico		24						24	8	8	8	24	36	72	56	32					8		300
Netherlands	24	40	16		8	8		174	39	56	24	40	16	88	55	56		16		8	72		740
Spain		8						69	49	31	16	32	16	56	80	24					16	16	413
Sweden	16	40	16					61	8	16	16	14	8	48	64	64					8		379
Switzerland	8	38	24	8	16			24	8		8	8	8	22	16	48					21		257
United Kingdom	40	48	24					239	286	122	88	72	63	175	103	129		40	8	56	16	8	1517
United States	400	896	206	16	120	16	16	2228	2787	1433	1047	1137	908	1520	864	1256		331	120	1008	24	41	16,374

Source: Own authorship.

Table 5. Grouped S&P credit ratings: Class A (highest), B and C (intermediate), and D (lowest/default).

Credit Ratings Classes	S&P Global's Rating Grade
Class A:	AAA, AA+, AA, AA-, A+, A, A-
Class B:	BBB+, BBB, BBB-
Class C:	BB+, BB, BB-
Class D:	B+, B, B-, CCC+, CCC, CCC-, CC, C, SD, D

Source: [S&P Global \(2024\)](#).

One of the most striking observations is that the United States has the highest number of Class D ratings, 7972, indicating that a large share of rated entities in the country fall into the lowest rating category. On the other hand, countries such as France, Germany, and the United Kingdom exhibit a strong presence of Class A ratings, highlighting a concentration of highly rated entities. Similarly, Japan and Canada also have a significant proportion of Class A ratings, reinforcing their position as countries with a higher credit quality profile.

Meanwhile, countries such as Brazil, the Cayman Islands, and Mexico have a greater proportion of Class B and Class C ratings, indicating a mid-tier credit profile. In contrast, Germany, China, and Hong Kong demonstrate a more balanced distribution across Classes A, B, C, and D, suggesting a diversified risk profile.

Class D ratings are also notably significant in several other countries besides the United States. Countries such as Switzerland, Spain, and Italy exhibit a visible portion of Class D ratings, suggesting a notable presence of lower-rated entities. Likewise, the United Kingdom also has a considerable number of Class D ratings, 775, indicating that while some entities are highly rated, a significant portion falls within the lower classification.

Justification for Country-Specific Models

Given the heterogeneity of economic environments and the varying distribution of credit ratings across countries, training a single global model may result in suboptimal performance. A universal model could struggle to capture the complex, country-specific relationships embedded within the data. To mitigate this limitation, we implemented a country-specific modeling approach, developing separate models for each of the 20 countries in the dataset.

This methodology offers several advantages:

- **Enhanced Data Homogeneity:** Training models on country-specific datasets ensures greater uniformity in economic conditions, reducing variability introduced by cross-country differences and improving model reliability.
- **Improved Predictive Accuracy:** By tailoring models to individual economic contexts, this approach enhances predictive performance by mitigating confounding effects inherent in aggregated datasets.
- **Scalability and Adaptability:** Independent models enable efficient updates and fine-tuning as new country-specific data becomes available, supporting both scalability and continuous improvement.

Each model was trained using the respective country's data subset, underscoring the value of localized insights in credit rating prediction.

4.3.3. Credit Rating Prediction

To assess the predictive performance of our models, we utilized the Scikit-learn library in Python 3.10 and evaluated five ML algorithms: LR, RF, GB, SVMs, and Artificial Neural Networks (ANN). Each algorithm was initially implemented using default hyperparameter configurations. Model evaluation was performed through five repetitions of 10-fold

cross-validation, a widely adopted technique to ensure robust and reliable performance assessment while mitigating potential biases due to data partitioning.

DT were initially included in the model evaluation process. However, they consistently exhibited lower accuracy and generalization ability compared to ensemble methods such as RF and GB. Given their comparatively weaker performance, they were excluded from the final results table for brevity.

The results, summarized in Table 6, report the mean accuracy scores for each algorithm across the 20 country-specific datasets. In most cases, the differences in model performance were marginal. However, notable variations were observed in the datasets from the United States, United Kingdom, France, and Australia, where ANN, RF, and GB consistently outperformed the other algorithms. In contrast, LR and SVMs demonstrated comparable performance but generally yielded lower accuracy in these datasets relative to the more complex models.

Table 6. Mean accuracy of five machine-learning algorithms—ANN, GB, Logistic Regression, RF, and SVM—across 20 country-specific credit rating datasets, based on five repetitions of 10-fold cross-validation. Higher values indicate better predictive performance.

N°	Country	ANN	GB	Logistic Regression	RF	SVM
1	Australia	0.96	0.96	0.70	0.97	0.76
2	Brazil	0.99	0.99	0.91	0.99	0.88
3	Canada	0.98	0.97	0.70	0.96	0.73
4	Cayman Islands	0.99	0.99	0.92	0.99	0.87
5	China	0.99	0.95	0.87	0.98	0.83
6	France	0.97	0.97	0.78	0.96	0.78
7	Germany	0.97	0.95	0.80	0.96	0.75
8	Hong Kong	0.98	0.97	0.97	0.99	0.90
9	Ireland	0.99	0.98	0.91	0.99	0.89
10	Italy	0.99	0.94	0.93	0.94	0.94
11	Japan	0.97	0.96	0.92	0.98	0.87
12	Republic of Korea	0.99	0.98	0.99	0.97	0.95
13	Luxembourg	0.98	0.95	0.87	0.91	0.87
14	Mexico	0.99	0.97	0.95	0.99	0.96
15	Netherlands	0.99	0.97	0.88	0.98	0.85
16	Spain	0.98	0.96	0.92	0.98	0.94
17	Sweden	0.98	0.97	0.92	0.98	0.94
18	Switzerland	0.99	0.95	0.99	0.97	0.98
19	United Kingdom	0.97	0.96	0.72	0.94	0.74
20	United States	0.86	0.74	0.42	0.86	0.54

Source: Own authorship.

To assess whether the observed performance differences among the algorithms were statistically significant, we employed the methodology outlined by Demšar (2006). First, we applied the Friedman test at a significance level of 0.05 to evaluate the null hypothesis that “there is no statistical difference between the algorithms” across country-specific datasets. In cases where the null hypothesis was rejected, we conducted the Nemenyi post hoc test to identify specific pairwise differences, as shown in Figure 3.

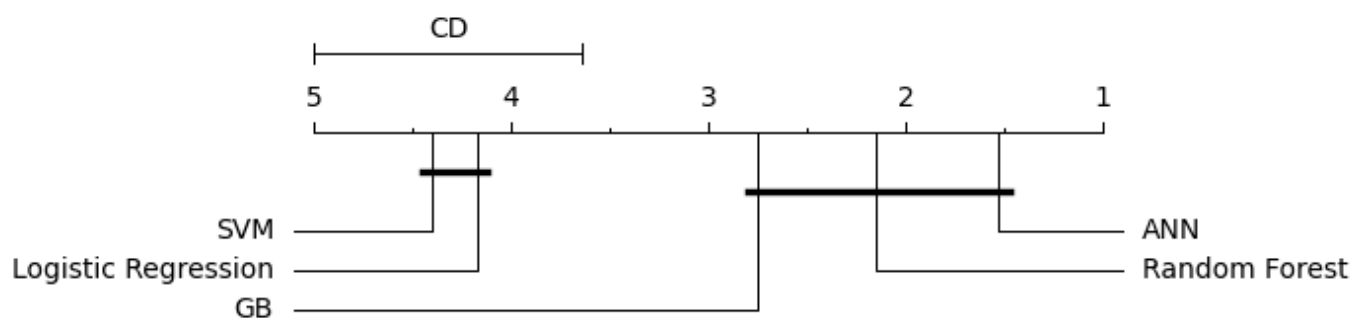


Figure 3. Results of the Friedman and Nemenyi tests comparing the predictive performance of machine-learning algorithms across country-specific datasets. Statistically significant differences indicate superior performance by specific models.

The Friedman test results indicated significant differences in algorithm performance for several countries, particularly those with larger datasets or more complex distributions of credit ratings (e.g., the USA, the UK, and France). For these datasets, the Nemenyi post hoc test revealed that ANN, GB, and RF consistently formed a superior group, significantly outperforming LR and SVMs.

5. Results

The findings reveal distinct trends regarding the predictive performance of ML models in credit rating classification:

- **ANN:** ANN consistently achieved the highest performance across countries, with accuracy scores ranging from 0.86 to 0.99. For instance, Brazil exhibited an accuracy of 0.99, suggesting that ANN models effectively capture financial patterns in complex datasets.
- **GB:** GB demonstrated strong predictive performance, frequently ranking second to ANN. Countries such as Australia (0.96) and Canada (0.97) underscore the robustness of ensemble methods in financial prediction tasks.
- **LR:** While exhibiting slightly lower accuracy compared to ensemble methods, LR performed consistently well, serving as a strong baseline model. For example, Canada recorded an accuracy of 0.70.
- **RF and SVM:** The performance of RF varied, with accuracy ranging from 0.86 to 0.99. Although SVM remained competitive, it frequently recorded lower accuracy compared to ANN, GB, and RF.

These results establish a clear hierarchy in predictive power, with ANN and ensemble methods demonstrating superior performance, while LR provides a reliable benchmark. The findings further suggest that dataset size and complexity influence model effectiveness, as evidenced by the diminished statistical significance in countries with smaller datasets.

An important methodological aspect of this study is that all ML models were applied using their default hyperparameter settings. This choice ensured uniformity across different algorithms and aligns with common practice in applied ML research, where default configurations are often employed as a baseline for initial performance assessment (Biecek and Burzykowski 2021). Nevertheless, recent research emphasizes that systematic hyperparameter optimization—including the adjustment of learning rates, model depth, and regularization parameters—can substantially improve both the accuracy and stability of predictive models (Akiba et al. 2019; Probst et al. 2019; Zhang et al. 2021). Future studies should, therefore, consider incorporating automated hyperparameter tuning techniques, such as Bayesian optimization or evolutionary algorithms, to further enhance the predictive capabilities and generalizability of ML models applied to corporate credit rating prediction.

6. Discussion

The results presented in this study demonstrate the predictive performance of various ML models—ANN, GB, LR, RF, and SVMs—across different countries. A comparative analysis of these findings within the context of existing literature on predictive modeling in financial studies yields several key insights.

ML in Financial Prediction: Prior research (e.g., [Lessmann et al. 2015](#)) has demonstrated that advanced models, particularly ANN and ensemble methods (RF, GB), frequently outperform traditional approaches such as LR in financial prediction tasks. Similarly, [Lokanan and Ramzan \(2024\)](#) highlight the strong predictive capabilities of ANN in financial distress modeling, reinforcing the advantages of advanced ML techniques in financial forecasting. These findings are consistent with the results of this study, where ANN and GB demonstrated superior predictive accuracy, particularly in countries such as Brazil and Australia, where accuracy scores exceeded 0.95.

Performance of ANN: The consistently high performance of ANNs observed across multiple datasets aligns with existing research, highlighting their capability to model complex, non-linear relationships in financial data. [Makridakis et al. \(2018\)](#) demonstrated that ML methods, including NN, significantly enhance forecasting accuracy across various domains, as evidenced by the M4 Competition results. The superior accuracy of ANNs in this study further substantiates their suitability for tasks such as credit rating prediction and financial forecasting.

Traditional vs. Modern Predictive Techniques: While LR remains a widely used benchmark due to its simplicity and interpretability, its performance, as observed in this study, was generally lower than that of ensemble-based models. These findings align with those of [Lessmann et al. \(2015\)](#) and [Ling and Wang \(2024\)](#), who demonstrated that LR, while effective in credit-scoring models, is often surpassed by more sophisticated techniques, particularly ensemble-based models. The improved performance of RF and GB suggests that ensemble learning methods, which integrate multiple weak learners, enhance predictive power by capturing intricate patterns within financial data.

Cross-Country Variability in Model Performance: Variations in model performance across different countries—such as the notably high accuracy of ANN and GB in Brazil—can be attributed to disparities in financial markets, economic structures, and data characteristics. This variability underscores the necessity of context-specific model evaluation, as regional differences in macroeconomic indicators, market volatility, and data availability significantly influence predictive model effectiveness.

For instance, [Pagliaro \(2025\)](#) emphasizes that ML's predictive power depends on varying market efficiencies, with AI able to exploit temporary inefficiencies in certain regions. He also highlights that incorporating macroeconomic and contextual variables improves forecasting, particularly in emerging markets with distinct financial structures.

6.1. Interpretability Considerations

This study has demonstrated the superior predictive accuracy of ML algorithms, particularly ANN and GB, in corporate credit rating prediction. However, an equally critical aspect of practical deployment is model interpretability. In financial applications, where credit ratings influence decisions related to capital allocation, regulatory compliance, and stakeholder confidence, the ability to explain model outputs is indispensable. A well-documented challenge in ML is the inherent trade-off between model complexity and interpretability ([Lipton 2018](#)). While complex models like ANN and GB excel at capturing non-linear relationships within financial datasets, they often function as “black box” models, providing limited transparency about how predictions are derived. In contrast, simpler models such as LR or DT offer more transparent decision processes but may under-

perform in complex contexts like those analyzed in this study. Interpretability is essential for various reasons: regulatory compliance requires transparency and explainability (e.g., [European Banking Authority 2021](#)); model risk management relies on interpretability for effective validation and bias detection (SR 11-7); and stakeholder trust in credit ratings depends on confidence in model outputs. Moreover, interpretability aids in identifying and mitigating bias, thereby promoting fair and equitable outcomes ([Doshi-Velez and Kim 2017](#)). Several approaches can enhance interpretability without sacrificing predictive power, including post hoc explainability tools like SHAP and LIME, model simplification through surrogate models, or adopting inherently interpretable models such as Generalized Additive Models (GAMs) or Explainable Boosting Machines (EBMs) ([Caruana et al. 2015](#)). Both global (overall behavior) and local (individual prediction) explanations are crucial in financial decision-making. Although this study primarily focused on benchmarking predictive accuracy, future research should incorporate explainability techniques to improve model transparency, regulatory acceptability, and practical utility in corporate credit risk assessment.

6.2. Regulatory Considerations

While this study highlights the significant predictive advantages of ML algorithms in corporate credit rating prediction, it is essential to consider the regulatory implications associated with deploying such models in financial risk management. Regulatory authorities, including the [Basel Committee on Banking Supervision \(2022\)](#), the [European Banking Authority \(2021\)](#), and national regulators such as the Federal Reserve (SR 11-7), have articulated clear expectations regarding the transparency, interpretability, and governance of risk models.

One of the primary concerns is the opacity of complex ML models, often referred to as “black box” algorithms. Models such as ANN and GB, while demonstrating superior predictive accuracy in this study, lack the intrinsic interpretability required by regulatory frameworks. Financial institutions are expected to justify and explain credit decisions, especially when these decisions impact capital adequacy and loan origination practices. To align with regulatory expectations for explainability, techniques such as SHAP and LIME can be employed to provide transparent and understandable insights into model predictions ([Rudin 2019](#)).

Additionally, regulators emphasize the need for robust Model Risk Management frameworks. According to the Federal Reserve’s SR 11-7 guidance, models must undergo comprehensive validation, back-testing, and stress testing to ensure their reliability under varying economic conditions. This requirement is particularly pertinent for ML models, which may be prone to overfitting or performance degradation in the presence of data shifts or regime changes in financial markets.

Data governance and bias mitigation represent further regulatory priorities. ML models are highly dependent on the quality and representativeness of input data. Poor data governance can lead to biased outcomes, undermining both model fairness and regulatory compliance. For instance, the proposed EU Artificial Intelligence Act mandates that high-risk AI systems, including those used in financial services, be subject to rigorous oversight to prevent discriminatory impacts. Thus, careful data pre-processing, feature selection, and bias audits are necessary steps when developing ML-based credit rating models.

Moreover, financial regulations such as Basel III link credit risk assessments directly to capital requirements, underscoring the importance of using models that are not only accurate but also compliant with regulatory standards. The EBA Guidelines on Loan Origination and Monitoring ([European Banking Authority 2021](#)) explicitly recommend

that models used for credit decisions be auditable, transparent, and capable of supporting sound credit risk practices.

In light of these considerations, while ML offers substantial potential to improve the efficiency and accuracy of credit rating prediction, its application in regulated financial environments must be accompanied by adherence to regulatory requirements for transparency, governance, and fairness. Future research should focus on the integration of Explainable AI (XAI) methods and the development of hybrid models that balance predictive power with interpretability. Additionally, collaboration between researchers, financial practitioners, and regulators is necessary to establish best practices for the safe and compliant deployment of ML in financial risk management.

7. Conclusions

This study evaluated the predictive performance of five ML models across multiple countries, yielding several key insights relevant to financial forecasting.

Predictive Performance Hierarchy: ANN consistently achieved the highest accuracy, reinforcing their capacity to model complex, non-linear relationships in financial data. Ensemble methods, including GB and RF, demonstrated strong predictive capabilities, further validating their effectiveness in financial applications.

Robustness of LR: Despite the increasing adoption of advanced ML techniques, LR exhibited competitive accuracy, underscoring its continued relevance as a benchmark model in financial prediction.

Cross-Country Variability: The observed variation in model performance across different countries highlights the potential influence of regional economic structures, financial market dynamics, and data characteristics on predictive accuracy.

These findings align with existing literature, supporting the superiority of modern ML approaches over traditional statistical methods for financial predictions. Future research should focus on improving model interpretability and incorporating additional economic indicators to further enhance predictive performance and practical applicability in diverse financial contexts.

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

Funding: This research received no external funding.

Data Availability Statement: The data used in this study were obtained from proprietary databases, specifically S&P Capital IQ Pro and Bloomberg, and are not publicly available due to licensing and confidentiality agreements. Access to these datasets requires a subscription or institutional access. Researchers interested in replicating or extending this study may contact the corresponding author for guidance on accessing similar data sources under appropriate licensing terms.

Acknowledgments: The authors would like to thank the Mackenzie Presbyterian University for providing institutional support throughout the development of this study. We also acknowledge the assistance of the university's data science laboratory for facilitating access to computational resources. During the preparation of this manuscript, the authors used OpenAI's ChatGPT, version gpt-4.0-turbo (OpenAI, 2025) for the purposes of language refinement, consistency checks, and support in drafting summaries and explanations of technical content. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
ANN	Artificial Neural Networks
AUC-ROC	Area Under the ROC Curve
BDTs	Boosted Decision Trees
CAPM	Capital Asset Pricing Model
CNNs	Convolutional Neural Networks
CRAs	Credit rating agencies
DLSS	Dual-Layer Smart Sampling
DT	Decision Trees
EBA	European Banking Authority
EBMs	Explainable Boosting Machines
ESG	Environmental, Social, and Governance
GAMs	Generalized Additive Models
GB	Gradient Boosting
GBM	Gradient Boosting Machines
LIME	Local Interpretable Model-Agnostic Explanations
LR	Logistic Regression
MDA	Multiple discriminant analysis
ML	Machine learning
MPT	Modern Portfolio Theory
NN	Neural Networks
OvO	One-vs-One
OvR	One-vs-Rest
RF	Random Forest
RNNs	Recurrent Neural Networks
S&P	Standard & Poor's
SFBS	Sequential Floating Backward Selection
SFFS	Sequential Floating Forward Selection
SHAP	Shapley Additive exPlanations
SVMs	Support Vector Machines
VIF	Variance inflation factor

References

- Addy, Wilhelmina Afua, Adeola Olusola Ajayi-Nifise, Binaebi Gloria Bello, Sunday Tubokirifuruar Tula, Olubusola Odeyemi, and Titilola Falaiye. 2024. AI in credit scoring: A comprehensive review of models and predictive analytics. *Global Journal of Engineering and Technology Advances* 18: 118–29. [\[CrossRef\]](#)
- Akiba, Takuya, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York: Association for Computing Machinery, pp. 2623–31. [\[CrossRef\]](#)
- Alharthi, Ayesha, Meera Alaryani, and Sanaa Kaddoura. 2025. A comparative study of machine learning and deep learning models in binary and multiclass classification for intrusion detection systems. *Array* 26: 100406. [\[CrossRef\]](#)
- Al-shalif, Sarah Abdulkarem, Norhalina Senan, Faisal Saeed, Wad Ghaban, Noraini Ibrahim, Muhammad Aamir, and Wareesa Sharif. 2024. A systematic literature review on meta-heuristic based feature selection techniques for text classification. *PeerJ Computer Science* 10: e2084. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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., and Anthony Saunders. 1998. Credit risk measurement: Developments over the last 20 years. *Journal of Banking & Finance* 21: 1721–42. [\[CrossRef\]](#)
- Altman, Edward I., Marco Balzano, Alessandro Giannozzi, and Stjepan Srhoj. 2022. Revisiting SME default predictors: The Omega Score. *Journal of Small Business Management* 61: 2383–417. [\[CrossRef\]](#)

- Altman, Edward I., Marco Balzano, Alessandro Giannozzi, and Stjepan Srhoj. 2023. The Omega Score: An improved tool for SME default predictions. *Journal of the International Council for Small Business* 4: 362–73. [CrossRef]
- Altman, Edward I., Marco Balzano, Alessandro Giannozzi, Eric Liguori, and Stjepan Srhoj. 2024. Bouncing back to the surface: Factors determining SME recovery. *Journal of Small Business Management*, 1–28. [CrossRef]
- Barboza, Flavio, Herbert Kimura, and Edward I. Altman. 2017. Machine learning models and bankruptcy prediction. *Expert Systems with Applications* 83: 405–17. [CrossRef]
- Barocas, Solon, Moritz Hardt, and Arvind Narayanan. 2023. *Fairness and Machine Learning: Limitations and Opportunities*. Cambridge: MIT Press.
- Basel Committee on Banking Supervision. 2022. *Principles for the Effective Management and Supervision of Climate-Related Financial Risks*. Basel: Bank for International Settlements. Available online: <https://www.bis.org/bcbs/publ/d532.htm> (accessed on 14 April 2025).
- Basel Committee on Banking Supervision. 2023. *Principles for the Effective Management and Supervision of Climate-Related Financial Risks*. Basel: Bank for International Settlements. Available online: <https://www.bis.org/bcbs/publ/d533.htm> (accessed on 16 March 2025).
- Bello, Oluwabusayo Adijat, Adebola Folorunso, Oluomachi Eunice Ejiofor, Folake Zainab Budale, Kayode Adebayo, and Olayemi Alex Babatunde. 2023. Machine learning approaches for enhancing fraud prevention in financial transactions. *International Journal of Management and Technology* 10: 85–108.
- Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri. 2020. On the rise of fintechs: Credit scoring using digital footprints. *Review of Financial Studies* 33: 2845–97. [CrossRef]
- Biecek, Przemyslaw, and Tomasz Burzykowski. 2021. *Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models*. Boca Raton: CRC Press. [CrossRef]
- Breiman, Leo. 2001. Random forests. *Machine Learning* 45: 5–32. [CrossRef]
- Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. 1984. *Classification and Regression Trees*. Boston: Wadsworth International Group.
- Büyükkeçeci, Mustafa, and Mehmet Cudi Okur. 2023. A Comprehensive review of feature selection and feature selection stability in machine learning. *Gazi University Journal of Science* 36: 1506–20. [CrossRef]
- Cao, Hongliu. 2024. Recent advances in universal text embeddings. *arXiv* arXiv:2406.01607. [CrossRef]
- Caruana, Rich, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. 2015. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York City: Association for Computing Machinery, pp. 1721–30. [CrossRef]
- Chang, Victor, Sharuga Sivakulasingham, Hai Wang, Siu Tung Wong, Meghana Ashok Ganatra, and Jiabin Luo. 2024. Credit risk prediction using machine learning and deep learning: A study on credit card customers. *Risks* 12: 174. [CrossRef]
- Chen, Tianqi, and Carlos Guestrin. 2016. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York City: Association for Computing Machinery, pp. 785–94. [CrossRef]
- Cheng, Xueyi. 2024. A Comprehensive Study of Feature Selection Techniques in Machine Learning Models. *Artificial Intelligence and Digital Technology* 1: 65–78. [CrossRef]
- Cheng, Yu, Qin Yang, Liyang Wang, Ao Xiang, and Jingyu Zhang. 2024. Research on credit risk early warning model of commercial banks based on neural network algorithm. *arXiv* arXiv:2405.10762. [CrossRef]
- Demšar, Janez. 2006. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research* 7: 1–30.
- Doshi-Velez, Finale, and Been Kim. 2017. Towards a rigorous science of interpretable machine learning. *arXiv* arXiv:1702.08608. [CrossRef]
- Duffie, Darrell, and Kenneth J. Singleton. 2003. *Credit Risk: Pricing, Measurement, and Management*. Princeton: Princeton University Press.
- Egan-Jones Ratings. 2024. Methodologies for Determining Credit Ratings. Available online: <https://www.egan-jones.com/> (accessed on 14 April 2025).
- European Banking Authority. 2021. Guidelines on Loan Origination and Monitoring. Available online: <https://www.eba.europa.eu/regulation-and-policy/credit-risk/guidelines-on-loan-origination-and-monitoring> (accessed on 14 April 2025).
- European Commission. 2023. Artificial Intelligence Act. Available online: <https://artificialintelligenceact.eu/> (accessed on 10 February 2025).
- Friedman, Jerome H. 2001. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* 29: 1189–232. [CrossRef]
- Goldmann, Leonie, Jonathan Crook, and Raffaella Calabrese. 2024. A new ordinal mixed-data sampling model with an application to corporate credit rating levels. *European Journal of Operational Research* 314: 1111–26. [CrossRef]
- Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. Empirical asset pricing via machine learning. *The Review of Financial Studies* 33: 2223–73. [CrossRef]

- Guyon, Isabelle, Isabelle Weston, Stephen Barnhill, and Vladimir Vapnik. 2002. Gene selection for cancer classification using support vector machines. *Machine Learning* 46: 389–422. [CrossRef]
- Hassija, Vikas, Vinay Chamola, Atmesh Mahapatra, Abhinandan Singal, Divyansh Goel, Kaizhu Huang, Simone Scardapane, Indro Spinelli, Mufti Mahmud, and Amir Hussain. 2024. Interpreting black-box models: A review on explainable artificial intelligence. *Cognitive Computation* 16: 45–74. [CrossRef]
- Ileberi, Emmanuel, Yanxia Sun, and Zenghui Wang. 2021. Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. *IEEE Access* 9: 165286–94. [CrossRef]
- Investment Grade Capital. 2024. The Role of Credit Rating Agencies in Determining Investment Grade. Available online: <https://investmentgrade.com/investment-grade-credit-rating-agencies/> (accessed on 14 April 2025).
- Jagtiani, Julapa, and Catharine Lemieux. 2019. The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform. *Financial Management* 48: 1009–29. [CrossRef]
- Jarrow, Robert A., and Stuart M. Turnbull. 1995. Pricing derivatives on financial securities subject to credit risk. *The Journal of Finance* 50: 53–85. [CrossRef]
- Kaplan, Robert S., and Gabriel Urwitz. 1979. Statistical models of bond ratings: A methodological inquiry. *Journal of Business* 52: 231–61. [CrossRef]
- Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. LightGBM: A highly efficient gradient boosting decision tree. In *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. New York City: Association for Computing Machinery, pp. 3149–57.
- Kou, Gang, Xiangrui Chao, Yi Peng, Fawaz E. Alsaadi, and Enrique Herrera-Viedma. 2019. Machine learning methods for systemic risk analysis in financial sectors. *Technological and Economic Development of Economy* 25: 716–42. [CrossRef]
- Krauss, Christopher, Xuan Anh Do, and Nicolas Huck. 2017. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research* 259: 689–702. [CrossRef]
- Kuiper, Ouren, Martin van den Berg, Justen van der Burgt, and Stefan Leijnen. 2022. Exploring Explainable AI in the Financial Sector: Perspectives of Banks and Supervisory Authorities. In *Artificial Intelligence and Machine Learning*. Communications in Computer and Information Science. Cham: Springer, vol. 1530. [CrossRef]
- Kurbakov, Mikhail Y., and Valentina V. Sulimova. 2024. Fast SVM-based multiclass classification in large training sets. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* X-2: 17–23. [CrossRef]
- Lessmann, Stefan, Bart Baesens, Hsin-Vonn Seow, and Lyn C. Thomas. 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research* 247: 124–36. [CrossRef]
- Ling, Yi, and Pang Paul Wang. 2024. Ensemble machine learning models in financial distress prediction: Evidence from China. *Journal of Mathematical Finance* 14: 226–42. [CrossRef]
- Lipton, Zachary C. 2018. The mythos of model interpretability. *Communications of the ACM* 61: 36–43. [CrossRef]
- Lokanan, Mark Eshwar, and Sana Ramzan. 2024. Predicting financial distress in TSX-listed firms using machine learning algorithms. *Frontiers in Artificial Intelligence* 4: 654321. [CrossRef]
- Lu, Tian, Yingjie Zhang, and Beibei Li. 2019. The value of alternative data in credit risk prediction: Evidence from a large field experiment. Paper present at 40th International Conference on Information Systems (ICIS 2019), Munich, Germany, December 15–18.
- Lundberg, Scott M., and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 30: 4765–74.
- Majumder, Saibal, Aarti Singh, Anupama Singh, Mykola Karpenko, Haresh Kumar Sharma, and Somnath Mukhopadhyay. 2024. On the analytical study of the service quality of Indian Railways under soft-computing paradigm. *Transport* 39: 54–63. [CrossRef]
- Makridakis, Spyros, Evangelos Spiliotis, and Vassilios Assimakopoulos. 2018. The M4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting* 34: 802–8. [CrossRef]
- Markowitz, Harry. 1952. Portfolio selection. *The Journal of Finance* 7: 77–91. [CrossRef]
- Merton, Robert C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance* 29: 449–70. [CrossRef]
- Mienye, Ibomoiye Domor, and Nobert Jere. 2024. A survey of decision trees: Concepts, algorithms, and applications. *IEEE Access* 12: 86716–27. [CrossRef]
- Mohammed, Ammar, and Rania Kora. 2023. A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University—Computer and Information Sciences* 35: 757–74. [CrossRef]
- Mokheleli, Tsholofelo, and Tinofirei Museba. 2023. Machine learning approach for credit score predictions. *Journal of Information Systems and Informatics* 5: 497–517. [CrossRef]
- Nemati, Khadijeh, Amir Hosein Refahi Sheikhan, Sohrab Kordrostami, and Kamrad Khoshhal Roudposhti. 2024. New Hybrid Feature Selection Approaches Based on ANN and Novel Sparsity Norm. *Journal of Electrical and Computer Engineering* 2024: 7112770. [CrossRef]

- Ohlson, James A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18: 109–31. [CrossRef]
- Owoade, Samuel Jesupelumi, Abel Uzoka, Joshua Idowu Akerele, and Pascal Ugochukwu Ojukwu. 2024. Enhancing financial portfolio management with predictive analytics and scalable data modeling techniques. *International Journal of Scholarly Research and Reviews* 5: 89–102. [CrossRef]
- Pagliaro, Antonio. 2025. Artificial intelligence vs. efficient markets: A critical reassessment of predictive models in the big data era. *Electronics* 14: 1721. [CrossRef]
- Probst, Philipp, Marvin N. Wright, and Anne-Laure Boulesteix. 2019. Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery* 9: e1301. [CrossRef]
- Prokhorenkova, Liudmila, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. 2018. CatBoost: Unbiased boosting with categorical features. In *NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems*. New York City: Association for Computing Machinery, pp. 6639–49.
- Rabinowitz, David, Mahmood I. Surty, Waheeda Mohamed, and Warren Maroun. 2024. Perceptions of the implementation of credit rating agency regulation in South Africa. *South African Journal of Economic and Management Sciences* 27: 5781. Available online: https://hdl.handle.net/10520/ejc-ecoman_v27_n1_a5781 (accessed on 14 April 2025). [CrossRef]
- Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. 2016. “Why should i trust you?” Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York City: Association for Computing Machinery, pp. 1135–44. [CrossRef]
- Rudin, Cynthia. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1: 206–15. [CrossRef]
- Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986. Learning representations by back-propagating errors. *Nature* 323: 533–36. [CrossRef]
- S&P Global. 2024. Credit Ratings Definitions and Criteria. Available online: <https://www.spglobal.com> (accessed on 14 April 2025).
- Sabharwal, Renu, Shah J. Miah, Samuel Fosso Wamba, and Peter Cook. 2024. Extending application of explainable artificial intelligence for managers in financial organizations. *Annals of Operations Research*. advance online publication. [CrossRef]
- Seetharaman, Arumugam, Vikas Kumar Sahu, A. S. Saravanan, John Rudolph Raj, and Indu Niranjana. 2017. The impact of risk management in credit rating agencies. *Risks* 5: 52. [CrossRef]
- Sharpe, William F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19: 425–42. [CrossRef]
- Shiller, Robert J. 2003. From efficient markets theory to behavioral finance. *Journal of Economic Perspectives* 17: 83–104. [CrossRef]
- Srhaj, Stjepan, Vanja Vitezić, Alessandro Giannozzi, and Josip Mikulić. 2024. Tourism SME default: A note on predictors. *Tourism Management* 103: 104910. [CrossRef]
- Talaei Khoei, Tala, and Naima Kaabouch. 2023. Machine learning: Models, challenges, and research directions. *Future Internet* 15: 332. [CrossRef]
- Tran, Dat, and Allan W. Tham. 2025. Accuracy comparison between feedforward neural network, support vector machine and boosting ensembles for financial risk evaluation. *Journal of Risk and Financial Management* 18: 215. [CrossRef]
- Tversky, Amos, and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185: 1124–31. [CrossRef]
- Umeaduma, Chidimma Maria-Gorretti, and Adeniyi Adedapo. 2025. AI-powered credit scoring models: Ethical considerations, bias reduction, and financial inclusion strategies. *International Journal of Research Publication and Reviews* 6: 6647–61. [CrossRef]
- Wang, Yifei. 2024. A comparative analysis of model agnostic techniques for explainable artificial intelligence. *Research Reports on Computer Science* 3: 25–33. [CrossRef]
- Wang, Zhiqin. 2024. Artificial intelligence and machine learning in credit risk assessment: Enhancing accuracy and ensuring fairness. *Open Journal of Social Sciences* 12: 19–34. [CrossRef]
- Weber, Patrick, K. Valerie Carl, and Oliver Hinz. 2024. Applications of explainable artificial intelligence in finance—A systematic review of finance, information systems, and computer science literature. *Management Review Quarterly* 74: 867–907. [CrossRef]
- Zhang, Chiyuan, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. 2021. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM* 64: 107–15. [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.