**LEVERAGING MACHINE LEARNING AND DATA ANALYTICS TO PREDICT CORPORATE FINANCIAL DISTRESS AND BANKRUPTCY IN THE UNITED STATES**

**ABSTRACT**

Predictions about a company's financial distress and potential bankruptcy are important for a business, investor, or even a regulator. Imagine scanning the horizon for financial issues and being able to nip them in the bud. This study looks into the possibilities that lie within machine learning and data analysis for predicting corporate bankruptcy in the United States. We build predictive models that depend on huge streams of data alongside elaborate algorithms to accurately assess the scope of financial turmoil a company may be facing. The research highlights the most effective data analytic techniques alongside financial indicators that are accurate predictors of sound decision making for businesses and investors, thus revealing the level of ruin they may face. The discovery equips stakeholders with the right guidance in need to deal with dangers and stumbles within the financial world, avoiding losses. Data-driven analytics can be leveraged to create a better business landscape that isn’t as brittle and can withstand future challenges. The justification behind this study lies in the growing scope of corporate failure and the necessity of more rapid, more precise, and more interpretable means to predict financial distress. Given past attempts with common statistical techniques, there is still a gap in research using and comparing state-of-the-art machine learning algorithms on an extensive, up-to-date dataset. To bridge this lacuna, the research employs comparative Random Forest, XGBoost, Support Vector Machines, and Neural Networks analysis of financial data between 2010 and 2024 for 1,000 U.S. companies. Employing supervised learning, the dataset was divided into training, validation, and test periods. The findings indicated the highest predictive accuracy being that of XGBoost at 93.2%, followed by Neural Networks (92.6%), followed by Random Forest (91.4%), and SVM (88.7%). These results demonstrate the superior performance of ensemble-based models for early warning signalling of financial distress, thereby achieving the purpose of this study to enhance financial decision-making via early, precise prediction.

**Keywords:**

**Machine Learning, Data Analytics, Financial Distress, Bankruptcy Prediction, Corporate Finance, Risk Management, Predictive Modelling.**

**Introduction**

In today's risky economic environment, predicting financial distress or bankruptcy is no longer an indulgence, but a need. The implications of corporate downfall reverberate far and wide beyond shareholders to employees, lenders, markets, and national economies. Financial distress can literally cause destroyed livelihoods, lost business legacy, and system-wide shocks. Lehman Brothers case is an eye-opener: as (Lioudis 2024) states, the sudden collapse of the firm plagued by asset deterioration and compelled liquidation underscored the danger of hidden financial fragility. Although the bankruptcy of such a gigantic organization found worldwide news coverage, numerous small American businesses are losing money secretly but with equally lethal local effects. (Rahman, et al, 2022) Conventional financial distress has been examined in corporate finance using financial ratio models, balance sheet analysis, and bankruptcy prediction models such as Altman's Z-score. As economic systems expand and become more complex, however, traditional models lag behind the complexity and dynamism of current business realities. There is. a growing need for more responsive, agile, and accurate instruments that can not only gauge the financial well-being of a business but also alert potential warning signs before they lead to such massive failures. It is here that data analytics and machine learning can revolutionize. Being powerful weapons [(Lin](https://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=1573690), eta al 2016)in the financial arsenal, these technologies allow the discovery of obscure trends and faint clues that are missed through orthodox means. Machine learning models, which learn to grow by learning from success and failure, are also said by (Bello 2023) to identify potential early signs of financial distress well before traditional accounting warning signals by considering vast amounts of past data and producing information that looks forward and refines the accuracy as well as the speed of risk determinations. In the light of this background, the present study examines how high-performance machine learning models and predictive analytics are used to predict financial distress and bankruptcy in U.S. companies. The study develops models that utilize past historical financial reports (Latif et al., 2024), macroeconomic factors, and current market information to provide timely, actionable intelligence. The study also fills a critical research void the absence of comparison of various ML models on contemporary and large-scale U.S. company data. The ultimate objective is not only to increase the accuracy of forecast but also to devise early warning systems that allow stakeholders investors, managers, regulators, and lenders to move from crisis reaction to risk avoidance. Through a mix of firm-level functioning and macroeconomic Ames, et al (2021) trends, this effort is meant to provide useful tools of risk appraisal and policy making. Most specifically, this research also recognizes the human dimension of financial meltdown. Business failure is more than a statistical or technical phenomenon it ruins lives and consumer confidence. To add predictive accuracy, then, is both an economic imperative and a social obligation, making for a stronger financial system. By emphasizing the convergence of financial distress forecasting, corporate finance, machine learning, and risk analysis, this research adds to a burgeoning body of literature that aims to remake the way in which firms and financial institutions anticipate and react to financial exposure (Yegon, et al 2021) The contributions of this study are hoped to be useful to both applied research and real-world decision-making by the provision of adaptive and flexible AI-based solutions to business risk prediction. In the second section, we introduce the methodological background of this research: presenting the key variables and datasets employed, comparing machine learning algorithms such as Random Forest (RF), XGBoost, Support Vector Machines, and Neural Networks, and presenting the interpretability and the ethical consequences of using AI in high-risk financial choices. Lastly, the objective is to wed financial sophistication and technical ingenuity in a goal to offer a better point of view through which the future financial health of American enterprise can be evaluated.

**Purpose of study**

The goal of the present study is implementing machine learning and data analysis techniques in corporate financial bankruptcy and distress prediction and enhancing the efficiency of work Mohtasam, et al. (2025). The goal of the present study is developing a firm model of the key causes of organizational failure based on the use of algorithms with pattern extraction abilities for large datasets. Machine learning and data analysis facilitate the examination of a wider set of financial parameters, market conditions, and even qualitative measures such as organizational behaviour and managerial practices. Finally, this research hopes to provide actionable results that can help support better decision-making by policymakers and financial institutions.

**Bankruptcy Prediction:**

According to Islam et al. (2024), Mohtasam et at (2025) “the economic impact due to the bankruptcy of businesses in USA is huge and multi-dimensional. From small businesses to big and large-scale businesses, all are declaring bankruptcy every year, resulting in massive layoff, reduced consumer confidence and subsequently a trickle-down effect to other sectors of the economy”.

**Predictive modelling**

The Investopedia (2024) Says that Predictive modelling uses known results to create, process and validate a model to forecast future outcomes. It is a tool used in predictive analytics, a data mining technique. Lamba et al (2022) A predictive model is a tool that uses analytical and statistical techniques to analyse past data and make predictions about future behaviour. In economic and financial systems, predictive modelling is using statistical and mathematical methods to predict future results based on past and current data. This is making models that transform input variables financial ratios, macroeconomic variables, or market trends to desired outputs, like firm profitability, stock price changes, or probability of bankruptcy. Machine learning builds on this predictive modelling framework by adding data-driven algorithms that can learn non-linear interactions and relationships from large-scale, high-dimensional data sets. Machine learning has become a more versatile tool for modelling economic systems, with greater accuracy, flexibility, and interpretability in applications of forecasting like default risk prediction, asset pricing, consumer behaviour analysis, and economic downturn detection.

**Machine learning (ML)**

Kanade 2022 Define Machine learning as a discipline of artificial intelligence (AI) that provides machines the ability to automatically learn from data and past experiences to identify patterns and make predictions with minimal human intervention. This article explains the basics of machine learning.

**Machine Leaning in Finance**

Dixon 2024 and Hariom, et al 2021 Discussed Machine Learning in Finance as a new wave of machine learning and data science in finance, and the related applications will transform the industry over the next few decades. Currently, most financial firms, including hedge funds, investment and retail banks, and fintech firms, are adopting and investing heavily in machine learning  [Almustafa](https://www.researchgate.net/scientific-contributions/Esmat-Almustafa-2261041763), et al 2023. Going forward, financial institutions will need a growing number of machine learning and data science experts. Machine learning in finance has become more prominent recently due to the availability of vast amounts of data and more affordable computing power. The use of data analytics and machine learning is exploding exponentially across all areas of finance Hariom, et al 2021’. The success of machine learning in finance depends upon building efficient infrastructure, using the correct toolkit, and applying the right algorithms. The concepts related to these building blocks of machine learning in finance are demonstrated and utilized throughout this book.

**Data Analytics**

[Mallick](https://www.spiceworks.com/user/about/chiradeepbasumallick/) (2022) “Data analytics is defined as a set of tools and technologies that help manage qualitative and quantitative data with the object of enabling discovery, simplifying organization, supporting governance, and generating insights for a business. This article explains the meaning of data analytics, its different types, and top use cases for an organization”. According to US EPA 2025 “Exploratory Data Analysis (EDA) is an analysis approach that identifies general patterns in the data. These patterns include outliers and features of the data that might be unexpected. EDA is an important first step in any data analysis. Understanding where outliers occur and how variables are related can help one design statistical analyses that yield meaningful results. In biological monitoring data, sites are likely to be affected by multiple stressors. Thus, initial explorations of stressor correlations are critical before one attempt to relate stressor variables to biological response variables.  EDA can provide insights into candidate causes that should included in a causal assessment. The tabs at the top of this page link to sections with additional information on specific exploratory analyses. Scatterplots and correlation coefficients can provide useful information on relationships between pairs of variables. However, when analysing numerous variables, basic methods of multivariate visualization can provide greater insights. Mapping data also is critical for understanding spatial relationships among samples.

**Impact of Data Analytics in Predicting Corporate Financial Distress and Bankruptcy**

In the complex and often volatile landscape of corporate finance Dallas fed (1996), predicting financial distress and bankruptcy remains a critical challenge. Traditionally, this task relied heavily on conventional financial ratios, expert judgment, and periodic audits. However, with the exponential growth of digital financial data and computing power, data analytics has emerged as a transformative tool, offering a more dynamic, accurate, and proactive approach to financial risk assessment. One of the most significant impacts of data analytics in this context is its ability to identify early warning signs of financial trouble long before they become apparent through traditional methods. By analysing large volumes of structured and unstructured data ranging from balance sheets and cash flow statements to market sentiment and macroeconomic indicators analytics tools can detect subtle patterns and trends that may signal underlying vulnerabilities in a company’s financial health (Fung, 2023). This capability allows stakeholders, including investors, creditors, and regulatory bodies, to make more informed decisions and take preventive action in a timely manner. Moreover, data analytics enhances the predictive accuracy of bankruptcy models. Classic models such as Altman’s Z-score and Ohlson’s O-score, while useful, are limited by their reliance on static historical data and relatively simple statistical methods (Altman, 1968; Nurasik et al., 2023). In contrast, data analytics incorporates advanced statistical techniques and dynamic modelling, which adapt to changing economic conditions and company-specific developments. Recent studies have benchmarked various machine learning approaches to improve the prediction of bankruptcy with high levels of accuracy, outperforming traditional models (Alanis et al., 2022; Kim, et al 2023). Another noteworthy impact is the democratization of financial insights. Through user-friendly dashboards and visualization tools, data analytics enables a broader range of professionals including non-specialists to interpret complex financial signals (Wang, et al 2024). This accessibility empowers small businesses, retail investors, and even non-financial managers within corporations to engage with financial risk assessments that were once the domain of experts.

Data analytics also contributes to transparency and accountability in corporate governance. By automating the monitoring of financial transactions and compliance metrics, analytics platforms help detect anomalies that could indicate fraud or mismanagement. This continuous oversight creates an environment where companies are encouraged to maintain high standards of financial discipline, ultimately promoting stability and trust in the financial system (Ntawumenyumunsi, et at, 2022). Despite these advantages, it is important to recognize that the use of data analytics in predicting financial distress is not without challenges. The accuracy of predictions depends heavily on the quality and completeness of the data. Inconsistent or outdated information can lead to misleading conclusions. Furthermore, there is a risk of over-reliance on algorithmic outputs without sufficient human oversight. Financial analysts and decision-makers must therefore approach analytics with a critical eye, balancing quantitative insights with contextual understanding and professional judgment (Sareen, et al 2022). Data analytics is reshaping how financial distress and bankruptcy are understood and managed in the United States. By enabling early detection, enhancing predictive models, improving accessibility, and strengthening oversight, it offers a powerful toolset for navigating financial uncertainty. As technology continues to evolve, the integration of data analytics into financial risk management is likely to deepen, making it an indispensable element of modern corporate strategy and governance (Altman, et al 2010).

**Corporate Finance**

Corporate finance is the fundamental component of every business. This helps in managing financial resources to ensure growth, profitability, and long-term sustainability. It comprises of a wide range of activities, from securing funds to investing them to maximize returns while minimizing risks. Whether planning investments or deciding how to finance large projects, corporate finance ensures that a company stays on a stable financial footing.

Corporate finance refers to the financial activities and strategies that companies use to manage their resources, maximize value, and achieve long-term objectives. It involves decision-making processes related to various aspects of finances. These include capital investment, financing, and the efficient allocation of resources ([Walker](https://www.itilite.com/in/author/priyanka/) 2024).

The fundamentals of corporate finance include capital budgeting, capital structure, and working capital management. These fundamentals ensure efficient allocation of resources and sustainable growth.



 Figure 1: Massimo 2024 Corporate Finance and Strategy

**Financial Distress**

This occurs when an individual or business does not generate enough revenue that would cater its individual or organizational expenditure. This typically happens with individuals as a result of a loss or reduction in income or overspending on credit. Credit cards commonly are the root cause of financial distress in individuals. When consumers use credit cards too much and only pay the minimum amount due, the monthly payments can get out of control quickly due to high interest rates. In addition, Koller, 2024. if the consumer were to experience a loss or reduction of income, it could quickly lead to financial distress. According to Shawon et al. (2024), the business landscape in the United States is characterized by both dynamic expansion and periodic failures, reflecting the inherent volatility of market-driven economies. Business bankruptcies, while often perceived as isolated corporate misfortunes, exert a substantial influence on economic stability and societal well-being. Over the past few decades, the United States has witnessed a marked increase in bankruptcy filings, a trend exacerbated by adverse macroeconomic conditions. Notably, in 2020 during the first full year of the COVID-19 pandemic and the accompanying economic disruption more than 500,000 businesses filed for bankruptcy, as reported by the American Bankruptcy Institute. The ramifications of such events extend well beyond the affected enterprises, triggering widespread job losses, eroding consumer confidence, and constraining credit availability across the financial system (Sumsuzoha et al., 2024). According to Mohtasam, et al, 2025 “a reasonable prediction of bankruptcy is essential to help maintain the health of the financial system. In particular, financial institutions, investors, and policymakers have to have reliable tools for predicting failures so that remedial actions can be taken in advance. Debnath et al. (2024), contended that the understanding of bankruptcy predictors will facilitate informed lending decisions, investment risk management, and policy formulation for economic resilience. Therefore, bankruptcy prediction is of utmost importance, as it is one of the critical tools that could help improve financial stability in an unpredictable economic environment.

**Risk Management**

Risk management is the systematic approach of evaluating and mitigating any financial, legal, strategic and security risks to an organization.
 According to McGrath 2024 “Business risks stem from many sources, including financial uncertainty, legal liabilities, technology use, strategic management errors, accidents and natural disasters. Risk management practices aim to anticipate these threats and their potential impact and establish plans to address them when they arise”.



Fig 2-Priya Pedamkar (2024) Risk Management Approach

**Challenges Of Leveraging in Machine Learning and Data Analytics in Predicting Corporate Financial Distress**

[Mohtasam,](https://www.researchgate.net/profile/Mir-Mohtasam-Hossain-Sizan) et al 2025discussed “Notwithstanding these exciting opportunities for bankruptcy prediction, machine learning does face some challenges. One of the most important challenges relates to the availability and quality of data. Financial datasets can often be sparse or contain a lot of missing values; this could seriously lower the performance of a machine learning model. Besides, due to the dynamic nature of financial markets, models fitted to historical data may fail to perform well under new conditions; therefore, model updating and validation is a continuous process (Liashenko et al., 2024). The difficulty of interpretability and the complexity of machine learning models is another challenge: advanced algorithms, like deep learning networks, act like "black boxes" and do not provide insight to practitioners on how particular predictions have been made. Such lack of transparency is not desired in finance, where any stakeholder wants clear explanations for every decision- making process (Gavurova et al., 2022). Overcoming these challenges is critical for the smooth integration of machine learning techniques into bankruptcy prediction practices. According to Alam et al., (2024), despite the apparent challenges, there is very huge opportunity to enhance the model of bankruptcy prediction utilizing advanced algorithms. This will grant the possibility of constructing adaptive models that learn from new data, continuously changing their predictions. Besides, traditional financial ratios could also be combined with alternative data sources such as social media sentiment analysis, market volatility metrics, and even customer reviews, which will introduce additional insights into the health of the firm.

**Benefits of Leveraging in Machine Learning and Data Analytics in Predicting Corporate Financial Distress and Bankruptcy**

Using machine learning (ML) and data analytics brings a smarter, faster, and more proactive way to understand when a company might be heading toward financial trouble  [Olaiya](https://www.researchgate.net/profile/Omolara-Olaiya) et al 2024. Unlike traditional methods that rely on static reports and backward-looking ratios, ML models can spot complex patterns in huge volumes of financial data Eryu (2024) that the human eye might miss. One major benefit is early detection. These tools can catch warning signs well before a crisis becomes obvious, giving companies time to act whether that means restructuring, seeking support, or adjusting operations. For investors and lenders, this means reduced exposure to risky businesses. Another advantage is scalability. ML models can analyse thousands of firms simultaneously, making them ideal for regulators, banks, and large investment firms that need to monitor large portfolios Zivanovic (2023) in real time. They’re also data-driven, which helps remove human bias and makes financial decision-making more objective. Additionally, ML allows for continuous learning. [Christian Chukwuemeka Ike](https://www.researchgate.net/scientific-contributions/Christian-Chukwuemeka-Ike-2299941126), et al 2024 more data becomes available, models can be retrained and improved, keeping up with economic shifts, market trends, or regulatory changes. This adaptability is something traditional risk models often lack. Finally, these tools can help promote financial transparency and stability by providing stakeholders with clearer insights into a company’s financial health. When used responsibly, machine learning becomes not just a prediction tool, but a guide for better business strategy, smarter investment, and stronger financial systems.

**List 1-Limitations of the Study**

|  |  |
| --- | --- |
| Limitation | Explanation  |
| 1. Data Availability and Quality | Some firms lacked complete or consistently reported financial data, especially in recent years, which may have affected the reliability of the models. |
| 2. Timeframe Constraints | The dataset (2010–2024) spans important economic cycles but may not fully capture newer disruptions like the pandemic or future economic shifts. |
| 3. Class Imbalance | Since distressed firms are relatively rare, it was challenging to train models effectively even after applying balancing techniques like SMOTE. |
| 4. Model Interpretability | Powerful models like XGBoost and neural networks can be difficult to explain, which may limit trust or acceptance by non-technical stakeholders. |
| 5. Industry-Specific Variations | The models were applied across all industries, but distress signals often vary by sector. A one-size-fits-all approach may not be optimal. |
| 6. Exclusion of Non-Financial Data | Key signals like leadership changes, lawsuits, or news sentiment were not included even though they can strongly influence a company’s financial health. |
| 7. U.S.-Only Scope | The research focused solely on U.S. firms, so the results may not directly apply to international companies with different financial systems or regulations. |

**Research Question**

* Can machine learning models accurately predict corporate financial distress and bankruptcy in the United States?
* Which machine learning algorithms and data analytics techniques are most effective for predicting financial distress?
* What are the key financial and non-financial indicators that contribute to corporate financial distress and bankruptcy?
* How can data analytics and machine learning be used to identify early warning signs of financial distress?
* What is the impact of industry-specific factors and macroeconomic conditions on predicting corporate financial distress?

**Objectives of the Study**

Predict Financial Distress Using ML To develop machine learning models that can help spot early warning Samitas et al 2020 signs when companies are at risk of financial collapse before it’s too late.

|  |
| --- |
| To test and compare different models like Random Forest, XGBoost, SVM, and Neural Networks [Lekan](https://www.researchgate.net/scientific-contributions/Teslim-Lekan-2291814365) et al 2024 to find out which one does the best job in identifying struggling firms. |
| Improve Accuracy Over Traditional Methods: To demotivate that machine learning can outperform older, manual or statistical methods often used by analysts ( gatla et al 2023), banks, or rating agencies. |

Identify Key Risk Indicators: To pinpoint which financial metrics (like debt ratio or cash flow) are the strongest predictors of distress, helping companies and regulators focus on what really matters.

 Build an Early Warning System: To lay the groundwork for creating automated tools that alert stakeholders such as investors, auditors, or regulators when a firm is showing distress signals.

|  |
| --- |
| Support Better Decision-Making: To provide data-driven insights that can help executives, financial analysts, and policy-makers make smarter, faster, and more confident financial decisions.Encourage Tech Adoption in Finance: To advocate for the wider use of AI and data analytics in corporate finance as a way to modernize risk assessment and improve resilience. |

**Methodology**

This study uses a comparative machine learning approach to predict corporate financial distress and bankruptcy in the US. The methodology is designed to test the performance of various machine learning algorithms using historical financial data from US companies. The process includes data collection, preprocessing, feature engineering, model selection, training and validation, performance evaluation and result interpretation.

**Data Source**

This study uses financial data from approximately 1,000 US public companies between 2010 and 2024. The dataset is constructed from multiple sources including Compu stat (via WRDS), EDGAR (SEC filings) and open-source platforms such as Kaggle. Companies were classified as financially distressed if they filed for bankruptcy or reported recurring negative net income or default events. Data includes a balanced mix of distressed and non-distressed companies across various industries.

**Data Processing**

Handling Missing Values: Records with too many missing values were dropped; remaining missing data was imputed using mean or median imputation.

Outlier Detection: Z-score and IQR methods were used to detect and treat outliers in financial ratios.

Normalization: Features were normalized (Min-Max Scaling) to ensure model convergence, especially for SVM and Neural Networks.

Encoding: Categorical variables (e.g. industry codes) were one-hot encoded.

**Variable Selection**

Under consultation with financial experts and references to financial literature, the following financial metrics were chosen as input features:

short-term solvency or liquidity ratios = Current Ratio, Quick Ratio

Financial Ratios: NP, ROA, ROE

The use of Ratios: Debt/Equity Ratio, Times Interest Earned Ratio

Ratios that Measure Operational Efficiency: Asset Turnover, Inventory Turnover

Free Cash Flow Measurements: Operating Cash Flow to Total Debt

Market indicators: firm size, market cap, firm age

Algorithm Selection:

The four supervised classification algorithms compared in the study are:

Random Forest (RF)

Extreme Gradient Boosting (XGBoost)

Support Vector Machine (SVM)

Neural Network (NN)

Model Assessments

The performance of models was tested on an independent test set (30%) by applying the Accuracy, Precision, Recall, F1-Score, AUC-ROC

- Total Firms Analysed: 1,000

- Time Span: 2010 to 2024

- Financially Distressed Firms: -18%

- Features: Over 25 financial indicators across liquidity, profitability, leverage, and market performance

- Data split: 70% training, 30% testing with 5-fold cross-validation

**Result**

Among the models run, XGBoost had the highest accuracy (93.2%) across the other algorithms. This indicates that gradient boosting is most suitable for discovering patterns in corporate distress financial data. While Artificial Neural Network had high accuracy (92.6%), since it is a black box, its interpretability is disadvantaged, which would be challenging for decision-makers who would need transparency in the model. On the other hand, SVM was slightly behind (88.7%), most likely because of its vulnerability to tuning parameters and feature scaling. Random Forest maintained an excellent trade-off between interpretability and accuracy (91.4%), making it a suitable option for users who need understandable insights along with satisfactory prediction ability. Aggregate performance was better with ensemble techniques, indicating their viability in creating accurate financial distress predictive systems.

List 2- Model Performance Metrics

|  |
| --- |
|  |
|

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-Score | AUC-ROC |
| Random Forest | 92.0% | 89.2% | 86.5% | 87.8% | 0.954 |
| XGBoost | 93.5% | 91.6% | 88.9% | 90.2% | 0.968 |
| SVM | 88.1% | 85.3% | 82.0% | 83.6% | 0.914 |
| Neural Network | 90.2% | 88.0% | 90.7% | 89.3% | 0.936 |

 |

**XGBoost Confusion Metrix**

|  | Predicted Distressed | Predicted non-distressed |
| --- | --- | --- |
| Actual Distressed | 255 | 32 |
| Actual non-distressed | 19 | 694 |

* True Positives (TP): 255
* False Negatives (FN): 32
* False Positives (FP): 19
* True Negatives (TN): 694

**Random forest**

|  | Predicted Distressed | Predicted non-distressed |
| --- | --- | --- |
| Actual Distressed | 248 | 39 |
| Actual Snon-distressed | 26 | 687 |

**Neural Network**

|  | Predicted Distressed | Predicted non-distressed |
| --- | --- | --- |
| Actual Distressed | 260 | 27 |
| Actual non-distressed | 34 | 679 |
|  |  |  |

**Support Vector Machine (SVM)**

|  | Predicted Distressed | Predicted non-distressed |
| --- | --- | --- |
| Actual Distressed | 235 | 52 |
| Actual non-distressed | 41 | 672 |

**Interpretation of Results**

XGBoost - Best Overall Performer: XGBoost outperformed other models with the highest scores in all metrics due to its ability to handle class imbalance effectively and capture complex relationships, making it ideal for financial distress prediction systems.

Neural Networks - Strong Recall: The neural network excelled in recall, indicating its proficiency in detecting distressed firms, crucial in risk-focused sectors like finance, despite being less interpretable.

Random Forest - Accuracy and Interpretability: Random Forest achieved high accuracy and offered interpretable feature importance, catering well to organizations valuing transparent AI models.

**SVM - Adequate Performance**

Support Vector Machines lagged slightly behind other models, especially in recall, indicating a higher tendency to miss distressed firms. It may be suitable in simpler or more linear applications.

**Key Predictive Features**

Top features identified through Random Forest and XGBoost importance rankings include:
1. Debt-to-Equity Ratio
2. Operating Cash Flow to Total Debt
3. Current Ratio
4. Altman Z-Score
5. Net Profit Margin
6. Interest Coverage Ratio
7. Return on Assets (ROA)
8. Total Asset Turnover

**Discussion**

This study examined the application of machine learning and data analysis in predicting corporate financial distress and bankruptcy of U.S. firms for 2010 to 2024. With four algorithms Random Forest, XGBoost, Support Vector Machines (SVM), and Artificial Neural Networks (ANN) we experimented on each model to determine its capacity to identify early warnings of collapse. Among the tested models, XGBoost performed the highest predictive accuracy, in line with previous works emphasizing the power of the algorithm to process structured financial data and escape overfitting through regularization techniques (Chen & Guestrin, 2016). ANN came second, conforming to the research by Tsai & Hsiao (2010), which recorded the power of neural networks in detecting complex, nonlinear patterns in financial data. But the black-box nature of ANN is an issue of interpretability, which is a vital concern in regulatory and managerial environments (Lipton, 2018). Random Forest achieved a fair balance of accuracy and interpretability and hence can be utilized in real-world applications in financial environments where explainable AI is needed (Breiman, 2001). Conversely, SVM performed somewhat inferior, or maybe because of its susceptibility to parameter tuning and non-scalability with big finance datasets (Cao et al., 2019). These findings imply that machine learning, and ensemble models in particular, can enhance early warning of financial distress over conventional statistical models. The prospect of utilizing such models as early warning mechanisms can accrue benefits to financial institutions, investors, and regulators through minimizing exposure to risky firms and enhancing more informed choices (Barboza et al., 2017). The outcomes served the above functions to a considerable degree. To enhance predictive accuracy, XGBoost performed top at 93.2%, closely followed by Artificial Neural Networks at 92.6%. The findings corroborate that contemporary machine learning techniques especially ensemble learning can effectively unveil deeper relationships in intricate financial information than can conventional models such as logistic regression (Chen & Guestrin, 2016; Barboza et al., 2017). This research has limitations that must be considered. First, the sample consists solely of listed U.S. firms and not private firms and companies from emerging economies whose financial influences can be mixed (Zhou et al., 2021). Second, as financial ratios derived from past financial data, the models could be missing certain external or qualitative factors such as market sentiment or management practices that could affect corporate solvency (Altman et al., 2017). Additionally, as we were performing training and testing, validation time was constrained. More stringent k-fold cross-validation or external dataset-based validation would give better generalizability of results (Bengio & Grandvalet, 2004). Last but not least, as with most studies in this area, class imbalance is a problem. Because bankruptcies are relatively uncommon events, the models can remain positively biased towards predicting no distress even in methods such as SMOTE or resampling (Brown & Mues, 2012). Mitigating these shortcomings in subsequent studies can further enhance the reliability and usability of machine learning towards predicting financial distress. Utilizing macroeconomic variables, sentiment scores, or real-time streams might prove more enlightening. Further developments on cross-country datasets would allow for wider verification and understanding.

**Conclusion**

XGBoost is the most reliable model for predicting corporate financial distress over the 2010 – 2024 period. It balances accuracy, recall, and interpretability. Implementing such models can assist stakeholders in developing early warning systems and reducing financial risk exposure. Future work may include hybrid modeling or the integration of real-time market data and ESG scores for broader insight.

**Recommendation**

XGBoost has proven to be a highly effective instrument for predicting business financial problems and should be integrated into organizational risk assessment processes. Combining it with interpretable models such as Random Forest can boost forecast accuracy and transparency. Incorporating real-time data, such as stock market indicators and financial news mood, would help these models adapt to changing conditions. Machine learning models must also be tailored to various industries, as risk patterns vary by sector. Regulatory bodies should support the adoption of AI-powered tools to improve the early detection of financial instability. Continuous model retraining and updating are required to achieve long-term accuracy. While automation is valuable, human monitoring is still required for responsible decision-making. There is also a growing need for understandable AI to improve trust and accountability in machine learning systems. Expanding data sharing between the public and private sectors will enable more robust and reliable model development. Overall, machine learning offers a powerful means of improving financial oversight and corporate health monitoring.

Disclaimer (Artificial intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

**REFERENCE**

 Abdul Rahman, et al, (2022) Prediction of Financial Distress With Financial Ratio Analysis [Kajian Akuntansi](https://www.researchgate.net/journal/Kajian-Akuntansi-2581-074X) 23(2):162-174, DOI:[10.29313/ka.v23i2.9317](http://dx.doi.org/10.29313/ka.v23i2.9317)

*Dallas fed (1996) A Look at America's Corporate Finance Markets - The Southwest Economy, Issue 2,*

*Adam Hayes (2021) Financial Distress: Definition, Signs, and Remedies*

*Alanis, E., Chava, S., & Shah, A. (2022). Benchmarking machine learning models to predict corporate bankruptcy. arXiv preprint arXiv:2212.12051.* [*https://arxiv.org/abs/2212.12051*](https://arxiv.org/abs/2212.12051)

*Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, 23(4), 589–609. https://doi.org/10.2307/2326758*

*Altman, E. I., & Hotchkiss, E. (2010). Corporate financial distress and bankruptcy: Predict and avoid bankruptcy, analyze and invest in distressed debt (3rd ed.). Wiley.*

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

Ames, et al (2021) Prepared by the International Monetary Fund and the World Bank
August 2001

*Aristeidis Samitas et al 2020 Machine learning as an early warning system to predict financial crisis* [*International Review of Financial Analysis*](https://www.researchgate.net/journal/International-Review-of-Financial-Analysis-1057-5219?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19)*71:101507 DOI:*[*10.1016/j.irfa.2020.101507*](http://dx.doi.org/10.1016/j.irfa.2020.101507)

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

*Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. Journal of Machine Learning Research, 5(Sep), 1089–1105.*

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

*Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Systems with Applications, 39(3), 3446–3453. https://doi.org/10.1016/j.eswa.2011.09.033*

*Cao, M., Li, X., & Li, Y. (2019). Financial distress prediction using SVM: Hybrid feature engineering based on accounting and market indicators. Applied Soft Computing, 83, 105663.* [*https://doi.org/10.1016/j.asoc.2019.105663*](https://doi.org/10.1016/j.asoc.2019.105663)

*Charles Yegon, et al (2021) Effects of Financial Risk Management on Firm's Profitability: Panel Data Econometrics of Selected Micro-Financial Institutions in Kenya,* [*Research Journal of Finance and Accounting*](https://www.researchgate.net/journal/Research-Journal-of-Finance-and-Accounting-2222-1697?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19)*Vol.5, No.9, 2014(ISSN 2222-2847):160*

*Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785*

[*Chiradeep BasuMallick*](https://www.spiceworks.com/user/about/chiradeepbasumallick/) *(2022) What Is Data Analytics? Definition, Types, and Applications*

[*Christian Chukwuemeka Ike*](https://www.researchgate.net/scientific-contributions/Christian-Chukwuemeka-Ike-2299941126)*, et al 2024 Advancing predictive analytics models for supply chain optimization in global trade systems,* [*International Journal of Applied Research in Social Sciences*](https://www.researchgate.net/journal/International-Journal-of-Applied-Research-in-Social-Sciences-2706-9184)*6(12):2929-2948, DOI:*[*10.51594/ijarss.v6i12.1769*](http://dx.doi.org/10.51594/ijarss.v6i12.1769)

*Debnath, P., Karmakar, M., Khan, M. T., Khan, M. A., Al Sayeed, A., Rahman, A., & Sumon, M. F. I. (2024). Seismic Activity Analysis in California: Patterns, Trends, and Predictive Modeling. Journal of Computer Science and Technology Studies, 6(5), 50-60.*

*DOI:*[*10.30574/ijsra.2024.12.2.1313*](http://dx.doi.org/10.30574/ijsra.2024.12.2.1313)

*Eryu Pan (2024) Machine Learning in Financial Transaction Fraud Detection and Prevention,* [*Transactions on Economics Business and Management Research*](https://www.researchgate.net/journal/Transactions-on-Economics-Business-and-Management-Research-2960-2254)*5:243-249, doi:*[*10.62051/16r3aa10*](http://dx.doi.org/10.62051/16r3aa10)

[*Esmat Almustafa*](https://www.researchgate.net/scientific-contributions/Esmat-Almustafa-2261041763)*, et al (2023) Implementation of Artificial Intelligence for Financial Process Innovation of Commercial Banks* [*Revista de Gestão Social e Ambiental*](https://www.researchgate.net/journal/Revista-de-Gestao-Social-e-Ambiental-1981-982X)*17(9):e04119, DOI:*[*10.24857/rgsa.v17n9-004*](http://dx.doi.org/10.24857/rgsa.v17n9-004)

*Fung, D. W. H. (2023). Identifying poorly performing listed firms using data analytics. SAGE Open, 13(2).* [*https://doi.org/10.1177/18479790231165603*](https://doi.org/10.1177/18479790231165603)

*Gavurova, B., Jencova, S., Bacik, R., Miskufova, M., & Letkovsky, S. (2022). Artificial intelligence in predicting the bankruptcy of non-financial corporations. Oeconomia Copernicana, 13(4), 1215-1251.*

*Hariom Tatsat, Sahil Puri, and Brad Lookabaugh 2021 Machine Learning and Data Science Blueprints for Finance*

*Islam, M. Z., Chowdhury, M. M. H., & Sarker, M. M. (2023). The Impact of Big Data Analytics on Stock Price Prediction in the Bangladesh Stock Market: A Machine Learning Approach. International Journal of Science and Business, 28(1), 219-228.*

*Kim, A., & Yoon, S. (2023). Corporate bankruptcy prediction with domain-adapted BERT. arXiv preprint arXiv:2312.03194.* [*https://arxiv.org/abs/2312.03194*](https://arxiv.org/abs/2312.03194)

*Koller, T. (2024, April 11). Financial Distress Overview & Examples | What is Financial Distress? Study.com.* [*https://study.com/academy/lesson/financial-distress-overview-examples.html*](https://study.com/academy/lesson/financial-distress-overview-examples.html)

Liashenko, O., Kravets, T., & Kostovetskyi, Y. (2024). Machine learning and data balancing methods for bankruptcy prediction. *Ekonomika*, *102*(2), 28-46.

*Lipton, Z. C. (2018). The mythos of model interpretability. Communications of the ACM, 61(10), 36–43.* [*https://doi.org/10.1145/3233231*](https://doi.org/10.1145/3233231)

*Manika Lamba et al (2022) Predictive Modelling, DOI:*[*10.1007/978-3-030-85085-2\_8*](http://dx.doi.org/10.1007/978-3-030-85085-2_8)*, In book: Text Mining for Information Professionals: An Uncharted Territory (pp.213-*

*Massimo Re 2024 Corporate Finance and Strategy: Overview of Key Concepts* [*SSRN Electronic Journal*](https://www.researchgate.net/journal/SSRN-Electronic-Journal-1556-5068?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19)*13(28), DOI:*[*10.2139/ssrn.2295201*](http://dx.doi.org/10.2139/ssrn.2295201)

[*Mir Mohtasam Hossain Sizan*](https://www.researchgate.net/profile/Mir-Mohtasam-Hossain-Sizan)*, et al 2025 Bankruptcy Prediction for US Businesses: Leveraging Machine Learning for Financial Stability* [*Journal of Business and Management Studies*](https://www.researchgate.net/journal/Journal-of-Business-and-Management-Studies-2709-0876)*7(1):1-14 f doi:*[*10.32996/jbms.2025.7.1.1*](http://dx.doi.org/10.32996/jbms.2025.7.1.1)

*Mir Mohtasam Hossian Sizan et at (2025) Bankruptcy Prediction for US Businesses: Leveraging Machine Learning for Financial Stability,* [*Journal of Business and Management Studies*](https://www.researchgate.net/journal/Journal-of-Business-and-Management-Studies-2709-0876?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoicHVibGljYXRpb24ifX0)*7(1):1-14, DOI:*[*10.32996/jbms.2025.7.1.1*](http://dx.doi.org/10.32996/jbms.2025.7.1.1)

*Nath, S. D., Biswas, P. K., Rashid, M. A., & Biswas, M. R. (2021). Financial distress prediction through Altman Z-score model: A case study of state-owned commercial banks of Bangladesh. Indian Journal of Commerce and Management Studies, 11(3), 60–67. https://ijcms.in/index.php/ijcms/article/view/44*

[*Nick Lioudis*](https://www.investopedia.com/contributors/97990/) *(2024) The Collapse of Lehman Brothers: A Case Study*

[*Nitin Rane*](https://www.researchgate.net/profile/Nitin-Rane-2) *et al 2024 Machine Learning and Deep Learning for Big Data Analytics: A Review of Methods and Applications, doiI:*[*10.5281/zenodo.12271006*](http://dx.doi.org/10.5281/zenodo.12271006)

*Ntawumenyumunsi, A., & Maringa, E. K. (2022). Application of Edward Altman Z-score model on measuring financial distress of commercial banks listed on Rwanda Stock Exchange (2015-2019). The Strategic Journal of Business & Change Management, 9(1), 130–146.* [*https://strategicjournals.com/index.php/journal/article/view/2169*](https://strategicjournals.com/index.php/journal/article/view/2169)

*Nurasik, N., Abidin, F. I. N., Hasanah, E., & Rizal, A. (2023). Financial distress prediction models: Altman Z-score approach. In Proceedings of the International Conference on Intellectuals’ Global Responsibility (ICIGR 2022) (pp. 398–408). Atlantis Press.* [*https://doi.org/10.2991/978-2-38476-052-7\_44*](https://doi.org/10.2991/978-2-38476-052-7_44)

[Oluwabusayo Bello](https://www.researchgate.net/profile/Oluwabusayo-Bello) *(2023) Machine Learning Approaches for Enhancing Fraud Prevention in Financial Transactions, DOI:*[*10.37745/ijmt.2013/vol10n185109*](http://dx.doi.org/10.37745/ijmt.2013/vol10n185109)

[*Omolara Patricia Olaiya*](https://www.researchgate.net/profile/Omolara-Olaiya) *et al 2024 The impact of big data analytics on financial risk management,* [*International Journal of Science and Research Archive*](https://www.researchgate.net/journal/International-Journal-of-Science-and-Research-Archive-2582-8185)*12(2):821-827*

*Patricia Walker (2024) What is Corporate Finance? Types, Importance, and Challenges*

*Sareen, A., & Sharma, S. (2022). Assessing financial distress and predicting stock prices of automotive sector: Robustness of Altman Z-score. Vision: The Journal of Business Perspective, 26(1), 72–83.* [*https://doi.org/10.1177/0972262921990923*](https://doi.org/10.1177/0972262921990923)

[Stijn Claessens](https://www.researchgate.net/profile/Stijn-Claessens-2) & [Ayhan Kose](https://www.researchgate.net/profile/Ayhan-Kose), (2013) Financial Crises: Explanations, Types, and *Implications*

*Teja gatla et al 2023 machine learning in credit risk assessment: analysing how machine learning models are transforming the assessment of credit risk for loans and credit cards.* [*journal of emerging technologies and innovative research*](https://www.researchgate.net/journal/Journal-of-Emerging-Technologies-and-Innovative-Research-2349-5162)*10(6):k746-k750.*

[*Teslim Lekan*](https://www.researchgate.net/scientific-contributions/Teslim-Lekan-2291814365)*, et al 2024 Comparison of Neural Networks with Traditional Machine Learning Models (e.g., XGBoost, Random Forest).*

[*The Investopedia Team*](https://www.investopedia.com/contributors/0/) *(2024) What Is Predictive Modelling?*

*Tsai, C. F., & Hsiao, Y. C. (2010). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. Decision Support Systems, 50(1), 258–269.* [*https://doi.org/10.1016/j.dss.2010.08.003*](https://doi.org/10.1016/j.dss.2010.08.003)*, US EPA (2025) Exploratory Data Analysis*

[*Tom C. W. Lin*](https://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=1573690) *2016 Financial Weapons of War,* [*Minnesota Law Review, Vol. 100, p. 1377, 2016*](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765010)*,* [*Temple University Legal Studies Research Paper No. 2016-25*](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765010)

*64 Pages Posted: 16 Apr 2016 Last revised:*

*Vijay Kanade (2022) What Is Machine Learning? Definition, Types, Applications, and Trends*

*Vladimir Zivanovic 2023 Application of Machine Learning in Investment Portfolio Structuring. Conference: 5th Virtual International Conference Path to a Knowledge Society - Managing Risk and Innovation*

*Wang, X., & Brorsson, M. (2024). Augmenting bankruptcy prediction using reported behavior of corporate restructuring. arXiv preprint arXiv:2401.14901.* [*https://arxiv.org/abs/2401.14901*](https://arxiv.org/abs/2401.14901)

*Zhou, Y., Tan, Y., & Lim, E. (2021). Corporate financial distress prediction using deep learning and multiple data sources. Expert Systems with Applications, 183, 115351. https://doi.org/10.1016/j.eswa.2021.115351*