An Integrated Network Analysis and Machine Learning Approach for Improved Financial Distress Prediction

Nataraj Swetha[1], Bitra Gowtham[2], Shaik Lal Ahamad[3], Shaik Mahaboob Basha[4], Damaraju Vara Prasada Padmaja[5]

[1, 2, 3, 4, 5] *Department of CSE (AI&ML), SRK Institute of Technology, Vijayawada, A.P., India*

Abstract- ***This examination presents a cross breed model that consolidates network investigation and AI to work on the expectation of monetary pain. By using gathering learning techniques, for example, casting a ballot classifiers and Irregular Woods calculations, the model tends to the challenges of determining monetary misery with a dataset involving 86 elements and 3,672 examples. The joining of cutting edge AI methods with hearty troupe methodologies is intended to upgrade both the exactness and dependability of distinguishing associations in danger of monetary issues. The review surveys the presentation of these techniques in separating among upset and non-bothered elements, showing significant enhancements in expectation exactness. This headway gives important experiences to monetary investigators and leaders, offering a more complex way to deal with assessing monetary wellbeing pointers.***

Keywords: Ensemble learning, voting classifiers, Random Forest algorithms.

INTRODUCTION

Anticipating financial distress is a crucial area of focus for investors, financial analysts, and managers, as it helps identify companies at risk of financial instability or bankruptcy. Early intervention can minimize potential losses, but traditional methods of predicting financial trouble rely heavily on financial indicators such as profitability, liquidity, leverage, and market signals. These methods may fall short in capturing the complex relationships among various financial factors that affect a company's overall financial health. As financial markets become more complex and data availability increases, advanced techniques are being explored to model the intricate relationships and underlying patterns in the data.

Hybrid models combining network analysis with machine learning (ML) algorithms provide a promising solution to this challenge. Network analysis, which has been applied in fields like social networks, ecological systems, and financial markets, helps uncover hidden patterns and relationships between entities such as companies, industries, and financial markets. When combined with machine learning, particularly ensemble learning methods, the predictive power of such models is significantly enhanced. Ensemble learning, such as voting classifiers, aggregates predictions from multiple models, like decision trees, support vector machines, or neural networks, to produce a final prediction based on the majority vote, improving accuracy and reducing overfitting. Random Forest, another popular ensemble algorithm, builds multiple decision trees and combines their outcomes to improve prediction accuracy. By using bagging methods, Random Forest reduces overfitting and enhances the model's ability to handle noisy and complex datasets. These ensemble methods, when applied to financial distress prediction, significantly improve accuracy in identifying at-risk companies, enabling early detection and more informed decision-making. The study uses a dataset of 86 financial entities and 3,672 instances representing companies across various industries. The features include traditional financial ratios like profitability, liquidity, and solvency, as well as advanced financial metrics derived from network analysis. Network analysis captures relationships between company conditions, supply chain linkages, and broader economic impacts—factors often overlooked by traditional financial indicators. By integrating both traditional financial metrics and network-derived features, the hybrid model is better equipped to capture the full range of factors influencing a company's financial health, leading to more reliable predictions of financial distress. Ensemble learning further strengthens the hybrid model, improving its prediction accuracy and reducing the influence of noisy or

inconsistent data. Combining multiple algorithms leverages their individual strengths, addressing shortcomings such as overfitting in decision trees, which is mitigated by averaging results in Random Forest. By merging network analysis with these robust machine learning methods, the hybrid model provides a promising solution for more accurate financial distress prediction. The research demonstrates the effectiveness of this hybrid model in distinguishing between distressed and non-distressed companies, showing significant improvements in prediction accuracy compared to traditional methods. The use of advanced machine learning techniques and network analysis offers a more sophisticated approach to identifying financial vulnerabilities and understanding the underlying causes of financial distress. Furthermore, ensemble classifiers and Random Forest algorithms ensure the model's robustness, making it a valuable tool for financial analysts and managers assessing a company's financial stability.

LITERATURE SURVEY

[1] D. Kuiziniene, T. Krilavičius, R. Damaševičius, and R. Maskeličnas, ˙ ''Systematic review of financial distress identification using artificial intelligence methods,'' Appl. Artif. Intell., vol. 36, no. 1, Dec. 2022, Art. no. 2138124.

The paper "Systematic Review of Financial Distress Identification Using Artificial Intelligence Methods" by D. Kuiziniene, T. Krilavičius, R. Damaševičius, and R. Maskeličnas (2022) provides an in-depth review of 232 studies focused on AI applications in predicting financial distress. Utilizing PRISMA methodology, the review examines various challenges and techniques such as handling data imbalance, feature dimensionality reduction, and the role of non-financial indicators in financial distress prediction. Key machine learning methods and performance evaluation metrics are discussed. Findings emphasize the importance of data balancing and dimensionality reduction in preprocessing, revealing gaps for future research. This comprehensive review offers valuable insights for improving AI-based financial distress prediction models.

[2] F. Mai, S. Tian, C. Lee, and L. Ma, ''Deep learning models for bankruptcy prediction using textual disclosures,'' Eur. J. Oper. Res., vol. 274, no. 2, pp. 743–758, Apr. 2019. This study explores the use of deep learning models for predicting corporate bankruptcy through textual disclosures, an area often overlooked in financial decision-making. By utilizing layers of neural networks, these models can extract features from textual data to improve prediction accuracy. The research involves a comprehensive bankruptcy database of 11,827 U.S. public companies, demonstrating that deep learning models outperform traditional methods in forecasting bankruptcy. The integration of textual data with accounting ratios and market variables enhances prediction performance. Additionally, the study compares different deep learning architectures, finding that simpler models, like averaging embedding, are more effective than convolutional neural networks, providing significant evidence of the predictive power of textual disclosures.

[3] M. Farooq, A. Noor, and S. F. Qureshi, ''The impact of corporate social responsibility on financial distress: Empirical evidence,'' Social Responsibility J., vol. 18, no. 5, pp. 1050–1067, Jun. 2022. The study explores the influence of corporate social responsibility (CSR) on financial distress in 139 Pakistan Stock Exchange (PSX) firms from 2008-2019. Using panel logistic regression and the dynamic generalized method of moments estimator, it assesses CSR through charitable donations, employee welfare, and research and development contributions, alongside financial distress indicators like the Altman Z-score and ZM-Score. Findings reveal a significant negative impact of CSR on financial distress, supporting the stakeholder view of CSR as beneficial for aligning shareholder and stakeholder interests and reducing financial distress risk. The study highlights implications for corporate managers, investors, and policymakers in developing countries, emphasizing the strategic importance of high-quality CSR investments.

[4] S. Sehgal, R. K. Mishra, F. Deisting, and R. Vashisht, ''On the determinants and prediction of corporate financial distress in India,'' Managerial Finance, vol. 47, no. 10, pp. 1428–1447, Oct. 2021. The study by Sehgal, Mishra, Desisting, and Vashisht aims to identify key microeconomic determinants of financial distress and develop a prediction model for India. Using two accounting-based definitions of financial distress, the study employs a binomial logit model and machine learning models—support vector

machine (SVM) and artificial neural network (ANN)—to compare prediction accuracy. Findings indicate that five financial ratios and firm size significantly predict distress. SVM and ANN outperform the logit model, with one-year-ahead forecasts being more accurate. Practical implications include monitoring six critical financial ratios for early warning systems, demonstrating the superiority of machine learning models. This research provides valuable insights for creditors, policymakers, and regulators in emerging markets.

[5] S. Ashraf, E. G. S. Félix, and Z. Serrasqueiro, ‘‘Do traditional financial distress prediction models predict the early warning signs of financial distress?’’ J. Risk Financial Manage., vol. 12, no. 2, p. 55, Apr. 2019. This study examines the prediction accuracy of traditional financial distress models for firms in Pakistan from 2001 to 2015. The methodology compares model scores of distressed and stable firms to their actual financial status, also evaluating model performance before, during, and after the financial crisis. The findings reveal that the three-variable probit model has the highest overall accuracy, while the Z-score model better predicts insolvency at both early and advanced distress stages. The study highlights that the effectiveness of these models diminishes during financial crises. An important contribution is expanding the definition of financial distress to include early warning signs, emphasizing the need for models tailored to both developed and developing markets.

[6] R. Lumbantobing, ‘‘The effect of financial ratios on the possibility of financial distress in selected manufacturing companies which listed in Indonesia stock exchange,’’ in Proc. 6th Annu. Int. Conf. Manage. Res. (AICMaR). Atlantis Press, 2020, pp. 60–63. The study by R. Lumbantobing examines the impact of financial ratios on the likelihood of financial distress in manufacturing companies listed on the Indonesia Stock Exchange. Analyzing data from 30 companies over 2015-2017, the research found that activity ratios do not significantly influence financial distress. However, liquidity ratios have a significant negative impact, while debt ratios and earnings ratios positively affect the likelihood of financial distress. The study concludes that liquidity and debt ratios are reliable predictors of financial distress and suggests including earnings ratios related to liquidity in financial statements to help users make informed decisions during financial distress conditions. Keywords include debt ratio, liquidity, earnings, activity, and financial distress possibility.

EXISTING SYSTEM

Existing financial distress prediction systems primarily rely on traditional financial indicators and machine learning algorithms to forecast potential risks. These methods often utilize standard features such as profitability ratios, liquidity measures, and leverage indicators to train models like Support Vector Machines (SVM) and other classification algorithms. However, these approaches may lack depth in capturing the intricate relationships and interactions between companies. Network analysis, while promising, is seldom integrated with machine learning in financial distress prediction, leaving a gap in leveraging network-centric insights for enhanced predictive accuracy.

Disadvantages :

Limited Feature Scope: Traditional systems rely solely on standard financial indicators, potentially missing nuanced risk factors that can affect financial stability.

Surface-Level Analysis: Methods focusing only on profitability ratios and leverage may not capture deeper, complex relationships between financial metrics and distress signals.

Neglected Interactions: Existing models often overlook the interactions between different financial indicators, which can provide critical insights into financial health.

Lack of Network Insights: The absence of network analysis in current models misses out on leveraging interconnected financial relationships for improved predictions.

Inflexibility to Dynamic Factors: Traditional approaches might not adapt well to rapidly changing financial environments, reducing their accuracy in real-time distress prediction.

PROPOSED SYSTEM

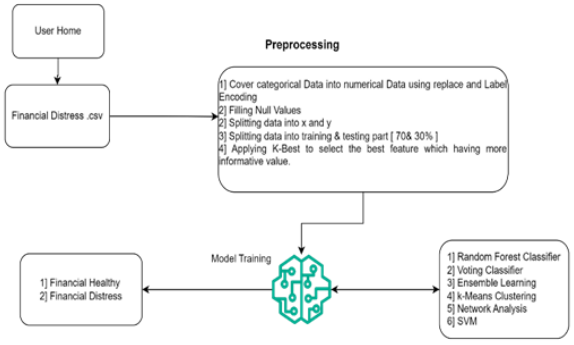

Fig:[1] Architecture

Proposed system algorithms:

SVM: Support Vector Machines (SVM) is a powerful and versatile supervised machine learning algorithm commonly used for classification and regression tasks. In the context of enhanced financial distress prediction, SVM plays a crucial role due to its ability to handle high-dimensional data and its robustness against overfitting, especially in scenarios with limited data points. By constructing optimal hyperplanes, SVM separates the data into distinct classes, effectively identifying patterns indicative of financial distress. The algorithm's kernel trick allows it to operate in a higher-dimensional space, making it capable of capturing complex relationships within financial data.

Random Forest: The Random Forest algorithm, integral to the title "A Hybrid Network Analysis and Machine Learning Model for Enhanced Financial Distress Prediction," offers robust predictive capabilities for financial distress scenarios. This ensemble learning method constructs multiple decision trees during training and merges their results to enhance accuracy and mitigate overfitting. Random Forest excels in handling high-dimensional data and complex interactions, making it ideal for financial distress prediction where numerous variables influence outcomes.

K-means clustering: K-means clustering is a widely used unsupervised machine learning technique that partitions a dataset into distinct clusters based on the similarity of data points. In the context of "A Hybrid Network Analysis and Machine Learning Model for Enhanced Financial Distress Prediction," K-means clustering can be effectively applied to identify patterns and group companies or financial entities exhibiting similar financial behaviors and risk profiles. The insights gained from K-means clustering can be integrated with other machine learning models and network analysis techniques.

Network analysis: Network analysis plays a pivotal role in the proposed model for predicting financial distress by leveraging the intricate connections and interactions within financial systems. By examining the relationships between various financial entities, such as companies, banks, and markets, network analysis can uncover hidden patterns and insights that traditional models might overlook. In this model, financial data is represented as a network, where nodes signify entities and edges represent the financial transactions or dependencies between them. Centrality measures, such as degree, betweenness, and closeness, are employed to identify key players and influential connections that could indicate systemic risks.

Voting Classifier: In financial distress prediction, leveraging a hybrid approach that combines network analysis and machine learning techniques can significantly enhance predictive accuracy. The proposed model integrates a voting classifier to amalgamate the strengths of multiple algorithms, thereby optimizing prediction outcomes. This ensemble model combines Decision Trees, Random Forests, and Logistic Regression to create a robust classifier that considers various perspectives of financial health. Each algorithm contributes its unique analytical strengths: Decision Trees offer interpretability, Random Forests provide resilience against overfitting, and Logistic Regression delivers probabilistic predictions.

Methodology:

Data Collection: The dataset includes financial and non-financial data from Kaggle, sourced from audited statements, with 83 features and 3672 samples.
Data Preprocessing: This involved handling missing values, removing duplicates, and selecting key features for financial distress prediction. Correlation coefficients for all 83 variables were calculated, identifying the top five.
Network Construction and Analysis: What is Network Analysis?

It helps identify relationships between companies based on financial data.

Instead of looking at companies separately, we analyze how they are connected financially.

Type of Network Used in Prediction:

Similarity-Based Network

Finds companies with closely related financial patterns based on key indicators.

Builds a weighted graph where stronger links mean higher similarity.

Model Training: Five different machine learning algorithms were used for model training: Random Forest, Gradient Boosting, AdaBoost, Voting Classifier, Slacking Classifier

Evaluation and Performance Analysis: The models were trained and evaluated using metrics such as accuracy, precision, recall, F-Score, and ROC curves.

Advantages:

Enhanced Accuracy
Robust Analysis
Improved Reliability
Advanced Insights
Enhanced Predictive Performance

RESULT

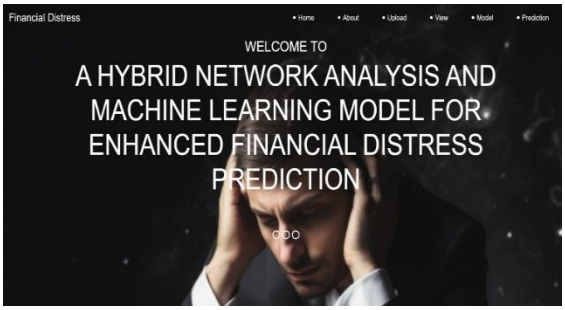

Fig:[2] Home page

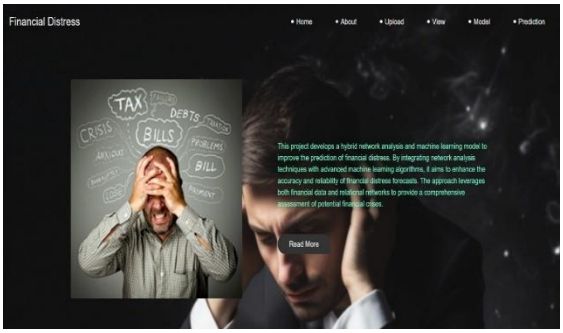

Fig:[3] information Related Financial Distress

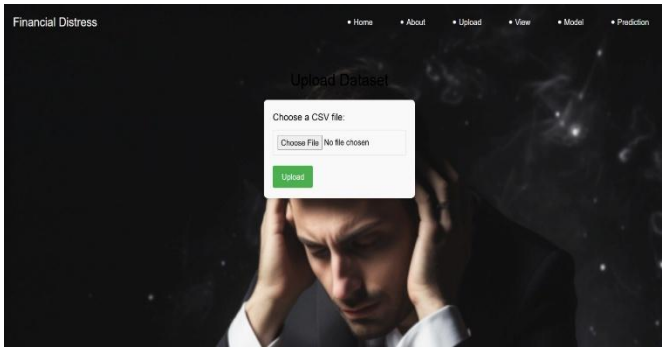

Fig:[4] User input page

Fig:[5] Dataset

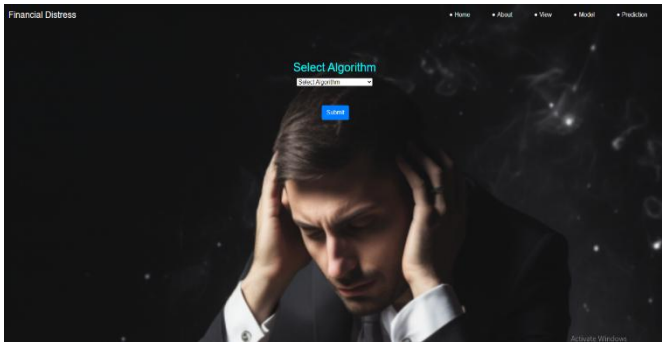

Fig:[6] training the algorithm to see which one has the best score

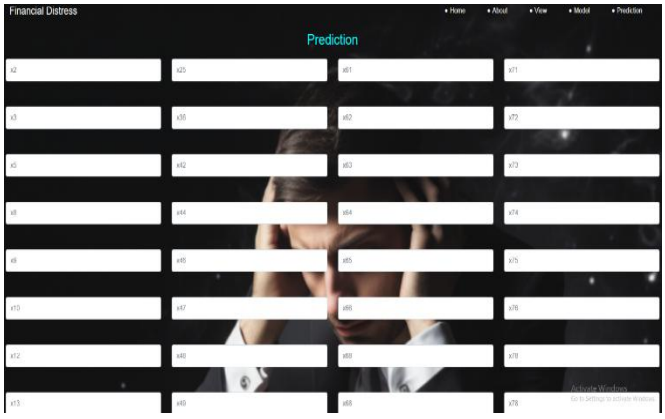

Fig:[7] Final Output

CONCLUSION

In conclusion, the hybrid network analysis and machine learning model demonstrated significant improvements in predicting financial distress by leveraging ensemble learning techniques, including voting classifiers and Random Forest algorithms. With a comprehensive dataset of 86 features and 3,672 samples, the model effectively captures the complexities inherent in financial distress prediction. The enhanced accuracy and reliability of this approach provide a valuable tool for financial analysts and decision-makers, offering deeper insights into financial health indicators. This research not only advances the predictive capabilities in financial distress assessment but also underscores the potential of integrating advanced machine learning strategies with network analysis for more robust financial forecasting.

FUTURE SCOPE

Future enhancements to this hybrid network analysis and machine learning model could involve several key developments. First, incorporating additional data sources such as real-time market indicators and macroeconomic factors may further refine prediction accuracy. Integrating advanced deep learning techniques, such as neural networks, could enhance the model's ability to capture complex patterns in financial distress. Additionally, expanding the dataset to include more diverse financial entities and geographic regions would improve generalizability. Employing explainable AI methods could also provide greater transparency in the decision-making process, helping stakeholders better understand the model's predictions. Finally, continuous model updating and validation with emerging financial trends will ensure sustained relevance and reliability in dynamic financial environments.

REFERENCE

[1] Al Ali, A. I., S, S. R., & Khedr, A. M. (2024). Enhancing financial distress prediction through integrated Chinese Whisper clustering and federated learning. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(3), 100344

[2] Che, W., Wang, Z., Jiang, C., & Abedin, M. Z. (2024). Predicting financial distress using multimodal data: An attentive and regularized deep learning method. *Information Processing & Management*, *61*(4), 103703.

[3] Elhoseny, M., Metawa, N., Sztano, G., & El-hasnony, I. M. (2022). Deep Learning-Based Model for Financial Distress Prediction. *Annals of Operations Research*, 1–23.

[4] El Madou, K., Marso, S., El Kharrim, M., & El Merouani, M. (2024). Evolutions in machine learning technology for financial distress prediction: A comprehensive review and comparative analysis. *Expert Systems*, *41*(2), e13485.

[5] Huang, Y. P., & Yen, M. F. (2019). A new perspective of performance comparison among machine learning algorithms for financial distress prediction. *Applied Soft Computing*, *83*, 105663. https://doi.org/10.1016/J.ASOC.2019.105663

[6] Hua, Z., Wang, Y., Xu, X., Zhang, B., & Liang, L. (2007). Predicting corporate financial distress based on integration of support vector machine and logistic regression. *Expert Systems with Applications*, *33*(2), 434–440.

[7] Kadkhoda, S. T., & Amiri, B. (2024). A Hybrid Network Analysis and Machine Learning Model for Enhanced Financial Distress Prediction. *IEEE Access*, *12*, 52759–52777.

[8] Liu, J., Wu, C., & Li, Y. (2019). Improving Financial Distress Prediction Using Financial Network-Based Information and GA-Based Gradient Boosting Method. *Computational Economics*, *53*(2), 851–872.

[9] Sun, J., Li, J., Fujita, H., & Ai, W. (2023). Multiclass financial distress prediction based on one-versus-one decomposition integrated with improved decision-directed acyclic graph. *Journal of Forecasting*, *42*(5), 1167–1186.

[10] Wu, D., Ma, X., & Olson, D. L. (2022). Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decision Support Systems*, *159*