Available online at www.jcsonline.in
Journal of Current Science & Humanities
11 (1), 2023, 1-12.

Impact Factor- 2.05

A HYBRID NETWORK ANALYSIS AND MACHINE LEARNING MODEL FOR ENHANCED FINANCIAL DISTRESS PREDICTION

[1] S.Varalakshmi, [2] CH.Sangeetha,[3] Dr.U.Naresh, [4] PREETHI SEN

[1,2,3] Assistant Professors,Department of Computer Science and Engineering, Kasireddy Narayanreddy College Of Engineering And Research, Abdullapur (V), Abdullapurmet(M), Rangareddy (D), Hyderabad - 501 505

[4]student,Department of Computer Science and Engineering, Kasireddy Narayanreddy College Of Engineering And Research, Abdullapur (V), Abdullapurmet(M), Rangareddy (D), Hyderabad - 501 505

ABSTRACT

In light of new uncertainties, it is more important than ever to be able to forecast financial hardship for the sake of financial planning. By combining network analysis and machine learning approaches, this research presents a new approach to forecasting financial trouble. The strategy entails connecting two businesses that are comparable to one another and have strong correlations in key financial metrics. One network shows how similar things are across all five characteristics, while the other shows how things are related to the most important one. Afterwards, seven characteristics focused on networks are extracted and added as new variables to the dataset. Additionally, cluster firms are subjected to community discovery techniques, and the labels that emerge are then included as categorical variables. A dataset with both initial and network-based variables is produced by this procedure. Foretelling monetary hardship in three different contexts makes use of five categorisation techniques. At first, the basic attributes are used just for training the models. Machine learning models' prediction abilities are improved in following cases by including network-centric characteristics from similarity and correlation networks. This enhancement is mostly due to features from the similarity network. The suggested model provides a comprehensive view of the ever-changing relationships between financial organisations and has exceptional prediction skills. The findings provide important information for decision-makers and demonstrate the effectiveness of network-based approaches in improving models for predicting financial hardship.

INTRODUCTION

Research into the prediction of financial hardship has grown in importance, especially for banks, firms, and other financial organisations that want to gauge the likelihood of insolvency or instability. Organisations may make better choices, reduce risks, and avoid financial crises if they can correctly forecast when they will experience financial difficulties. Conventional approaches for predicting financial hardship often evaluate organisations' financial health using statistical methods and financial measures obtained from past data. Despite their usefulness, these models often miss intricate data patterns and correlations, especially when working with massive datasets and non-linear interconnections across financial metrics. There has been encouraging progress in the field of financial crisis prediction thanks to developments in ML and network analysis in the last few years. Decision trees, SVMs, and neural networks are just a few examples of machine learning models that can sift through mountains of data in search of patterns that might otherwise go unnoticed. The interdependencies and linkages between various financial organisations, which conventional models fail to take into account but which may exacerbate financial hardship, may be better understood with the use of network analysis approaches.

In order to better anticipate financial hardship, this study presents a technique that combines network analysis with machine learning. For a more complete, accurate, and interpretable prediction system, the suggested approach combines network analysis with machine learning. Financial indicators, corporate linkages and other network-based elements, as well as powerful machine learning algorithms, are used in the hybrid method to analyse and forecast the possibility of financial hardship in organisations. With the use of these cutting-edge methods, the model hopes to provide stakeholders and financial institutions with improved risk management tools, so they may better anticipate financial disasters and step in before they happen.

Our hope is that this effort will show how combining network analysis with machine learning may outperform more conventional approaches. By improving its responsiveness, flexibility, and ability to deal with the intricacies of contemporary financial systems, this study's findings could revolutionise financial distress prediction.

II.LITERATURE REVIEW

In order to identify companies that may go bankrupt or otherwise fail financially, the forecasting of financial hardship has been a well-established field of study in finance for quite some time. Discriminant analysis, logistic regression, and probit models were once mainstays in financial crisis prediction models. There has been a noticeable uptick in the use of more advanced methods, such machine learning (ML) and network analysis, to improve the accuracy of predictions in response to the rising complexity of financial markets and the accessibility of huge, unstructured information.

1. Traditional Methods for Financial Distress Prediction

In the past, financial ratios taken from a company's financial accounts were the mainstay for predicting financial disaster. In this field, the Altman Z-score, Ohlson's O-score, and Springate model are the most often used models. In order to foretell when publicly listed firms would go bankrupt, many have turned to the Altman Z-score (Altman, 1968), which takes five financial criteria into account: liquidity, profitability, leverage, operational efficiency, and profitability. In order to evaluate the likelihood of insolvency based on nine financial measures, the Ohlson O-score (Ohlson, 1980) built on the idea by integrating a logistic regression model. These models have laid the groundwork, but they can't capture the non-linear interactions between financial variables or the ever-changing character of financial systems. Furthermore, in complicated situations, the predicted accuracy is lowered since these conventional models often miss hidden patterns in the data.

2. Machine Learning Approaches in Financial Distress Prediction

In order to circumvent the shortcomings of more conventional models, recent developments in machine learning (ML) have shown encouraging results. Traditional approaches generally fail to detect complicated, non-linear correlations in data. However, ML models like decision trees, support vector machines (SVM), k-nearest neighbours (KNN), and neural networks (NN) can. For instance, SVMs have been used in several studies to forecast insolvency by locating the best separation border between troubled and non-distressed companies by projecting financial measures into a higher-dimensional space (Kou et al., 2006). Xia et al. (2015) also noted that boosting algorithms and random forests have been used to address overfitting and unbalanced datasets in financial distress prediction.

The capacity of neural networks to model very complicated data patterns and structures is a major reason for their rising popularity, especially deep learning models. Using deep learning methods like long short-term memory networks (LSTM) and deep belief networks (DBN), Huang et al. (2019) showed that financial hardship may be predicted. When dealing with big, high-dimensional datasets with intricate temporal patterns, these strategies have shown to be more effective than conventional statistical models.

Although these ML approaches have achieved better results than conventional models in several areas, they still encounter difficulties when it comes to being understandable and comprehensible. Since it may be difficult for financial specialists to understand the reasoning behind a given prediction, many machine learning algorithms, especially deep learning models, are sometimes called "black boxes" because of this lack of transparency.

3. Network Analysis in Financial Distress Prediction

Aside from machine learning, network analysis is a potent tool for deciphering the interconnections among firms, creditors, and investors—all of which may play a role in financial crises. Assumption number one is that a company's interdependencies within the business ecosystem have a significant impact on the company's bottom line. Critical components in determining a company's risk of failure include intercompany ties, financial linkages, and supply chain networks.

To represent these relationships, researchers have turned to graph theory and intricate networks. For instance, according to Borgatti et al. (2009), centrality measures like degree centrality, betweenness centrality, and closeness centrality can shed light on a company's financial distress risk. These measures measure the number of direct connections a company has, the frequency with which a company acts as a bridge between other companies, and how close a company is to all other companies in the network. Centralised financial institutions may be more vulnerable to systemic risks and their potential spillover consequences.

The security and openness of a network's financial transactions might be enhanced using blockchain-based solutions. Narayanan et al. (2016) proposed that distributed ledgers and smart contracts might improve financial distress prediction by increasing stakeholder data sharing and decreasing the likelihood of fraud or misreporting.

4. Hybrid Approaches in Financial Distress Prediction

In an effort to better anticipate financial hardship, an increasing amount of research has investigated the use of machine learning in conjunction with network analysis. Using ML approaches to uncover hidden patterns in financial data and taking into account the interconnections and interdependence between enterprises in a network, these hybrid models strive to take use of the capabilities of both methodologies.

Successful financial distress prediction was achieved by hybrid models that combined conventional machine learning with graph neural networks (GNN). To give a more complete picture of a company's risk profile, for instance, GNNs integrated with support vector machines (SVMs) or random forests can combine structured financial data (like balance sheets) with unstructured data (like relationships within a business network) (Xu et al., 2018).

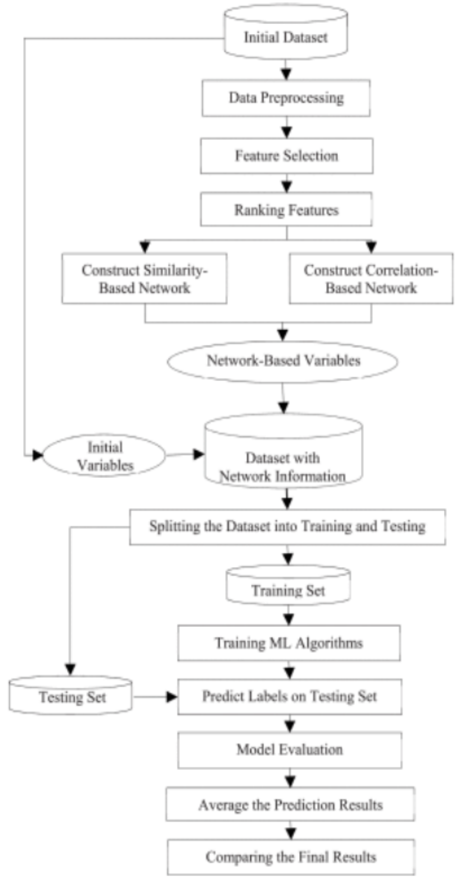
Furthermore, methods that integrate several machine learning models, such as ensemble learning, have been used to enhance the precision of predictions. For example, according to Zhou et al. (2020), a single ensemble model that incorporates decision trees, SVMs, and neural networks may enhance bankruptcy prediction by integrating the capabilities of each individual model.

5. Challenges and Future Directions

A number of obstacles persist even though network analysis and machine learning have made great strides. Data availability and quality is a big obstacle. Even while financial data is readily accessible, it can not always be precise or full, which makes it hard for machine learning models to train well. The difficulty of explaining complex models, especially deep learning algorithms, to stakeholders who aren't technically savvy is another obstacle. Combining financial data with real-time data streams from social media, news, and market sentiment should be the focus of future research, as should enhancing the interpretability of machine learning models.

Furthermore, for better financial distress prediction in the future, hybrid models combining big data with financial data derived from the internet of things (IoT) will be crucial. This is because financial systems are always changing and becoming more linked.

III.PROPOSED MODEL



By filling up some of the holes in existing forecasting approaches, our suggested model hopes to help with the urgent need for reliable financial distress prediction. Although many studies have investigated this important topic using different methods, such as statistical models, machine learning, and deep learning techniques [3, 63], there is still a big gap in our knowledge when it comes to combining network analysis with machine learning to get a better picture of the underlying complexity. When it comes to capturing the complex interdependencies between organisations and the amount of knowledge contained inside financial ecosystem interactions, traditional models aren't always up to the task. In order to overcome this shortcoming, our approach builds networks using feature correlations and similarities across companies. This process reveals hidden patterns that may indicate financial instability in its early stages. Here we provide our suggested model and describe the processes that went into making it, with Figure 1 serving as an example.

A. Data Preprocessing and Feature Selection

It is critical to handle the dataset with precision and care in the first phase. In order to better understand the topic at hand, data must be preprocessed by solving critical concerns such dealing with missing or duplicate entries, checking the target variable's balance, and improving the dataset using machine learning and feature correlation studies. The first and foremost objective is to isolate the most important characteristics; they will form the basis for the next procedures. In order to rank features according to their predictive usefulness, we compute their correlation with the target variable.

Two networks are built to enhance the analysis once the most important characteristics have been selected. Companies' similarities with regard to the chosen qualities form the basis of the first network, while correlations with the most important feature form the basis of the second. The underlying dynamics may be better understood with the help of this dual-network method, which places an emphasis on capturing the interrelationships and similarities across organisations. Analysing only these five characteristics keeps things simple and focused, allowing us to look for key signs that improve the model's predictions.

B. Network Construction

A more in-depth analysis of the structural and developmental trends among publicly listed businesses may be accomplished by seeing the financial market as a complex network.

These networks are built using two separate approaches. First, we may build a network by seeing how closely related different businesses are. This step entails connecting businesses with similar properties and evaluating common characteristics across the previously defined indicators.

The second strategy for constructing the network makes use of the association between businesses within certain variables. In order to find the distances between nodes in a network, a correlation matrix is created. Prior work has used this method to build financial networks [64].

Methods for creating networks are described in depth in the section that follows. By using these methodologies, one hopes to get a more complete understanding of the interrelationships and commonalities across businesses, as well as their structural and correlational dynamics.

C. Machine Learning Models

The next step is to combine the built-in characteristics and financial indicators into a new dataset. The use of classification algorithms to forecast financial hardship is made possible by these enhanced datasets. Features from the similarity-based and correlation-based networks are examined independently to assess the efficacy of various network creation methods. The machine learning models and assessment measures used in this phase are introduced in this section.

1) Models for Classification 1)

The use of classification algorithms allows us to train and forecast models using pre-existing data. This research utilises five machine learning models that are well-known and often utilised. Presented below is a synopsis of all:

1. Logistic Regression

A basic binary classification model is logistic regression. For situations where the result is binary (e.g., troubled or not), it is a powerful tool for forecasting financial hardship since it assesses the correlation between a binary dependent variable and several independent factors [69].

The K-Nearest Neighbours (KNN) algorithm

A non-parametric approach, K-closest Neighbours (KNN) sorts instances according to how close they are to their K closest neighbours in the space of features. When used for classification tasks, it uses the neighborhood's majority class to assign a class, and when used for regression, it uses the average value [69].

d. SVM, or Support Vector Machine

In high-dimensional domains, SVMs perform very well for regression and classification. They work well in situations when the data is not linearly separable because they generate hyperplanes that partition classes optimally [70].

judgement tree d.

Hierarchical models known as decision trees iteratively partition datasets according to feature values. While dealing with complicated data connections, its interpretable structure aids in identifying significant predictors for financial hardship [70].

Example: Random Forest

An ensemble approach, Random Forest uses a combination of decision trees to lessen the impact of overfitting and increase accuracy. It is well-suited for predicting financial trouble since it generates strong findings by combining the predictions of individual trees [69].

Applying these algorithms to different circumstances and evaluating their performance using the metrics provided below is what this phase is all about.

2) Measures for Assessment

A number of assessment measures are used to evaluate the models and to compare the efficacy of various methods:

a. Precision

By dividing the number of cases for which the model made a valid prediction by the total number of occurrences, accuracy assesses the model's overall performance. The symbol for it is:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{Total Instances}}$$

b: Precision

According to Equation 4, precision measures how many positive outcomes the model accurately predicted out of all positive outcomes. It indicates how well the model can prevent false positives.

$$Precision = \frac{TP}{TP + FP}$$

c: Recall

Recall determines the proportion of correct predictions relative to the total number of positive occurrences in the dataset, as shown in Equation 5. It measures how well the model can avoid making incorrect negative predictions.

$$Recall = \frac{TP}{TP + FN}$$

d: F1 Score

Equation 6 yields the F1-score, which is a well-rounded measure of both recall and accuracy, allowing for an exhaustive assessment of the two parameters. It works well in situations when finding the sweet spot between recall and accuracy is paramount.

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

e: Roc Curve

One graphical representation of a binary classification model's efficacy at different thresholds is the Receiver Operating Characteristic (ROC) curve. Various threshold settings are shown in the graph, which shows the link between sensitivity (the true positive rate) and specificity (1 - the false positive rate) [72].

A random classifier is shown by the diagonal line in a receiver operating characteristic (ROC) curve, and the area under the curve (AUC-ROC) measures the discriminatory power of the model. If the AUC-ROC is closer to 1, it means the model does better discrimination; if it's less than 0.5, it means it does no better than chance.

IV. DATASET AND DATA ANALYSIS

Dataset Description

With the use of financial indicators and network-based characteristics extracted from firm data, a dataset is created to forecast financial hardship. Important measures that are used to assess financial well-being include profitability, liquidity, solvency, and market performance. To further enhance the dataset and uncover latent patterns, network characteristics, such as metrics based on similarity and correlation, are used. A binary indicator of a company's financial distress status (1) or lack thereof (0) serves as the goal variable.

Data Preprocessing

Data quality and improved model accuracy are both achieved via rigorous preprocessing. Statistical methods such as regression, median, or mean are used to impute missing data. In order to prevent the results from being skewed, duplicate records are found and deleted. Oversampling (like SMOTE) or undersampling are two strategies that are used to remedy class imbalance. The goal is to make sure that both the distressed and non-distressed classes are represented equally. To make financial indicators more similar in size, they are normalised. In feature engineering, we use correlation analysis to find the best predictors and create new features based on networks.

Exploratory Data Analysis (EDA)

Patterns, outliers, and insights into feature distributions may be found using exploratory data analysis (EDA). Mean, median, and standard deviation are statistical summaries that aid in comprehending the major patterns of the dataset. Pair plots investigate feature interactions, while correlation heatmaps show how characteristics relate to the dependent variable. Using box plots, we can find out which firms are outliers, and bar charts or pie charts can show us how many businesses are in crisis compared to how many are not.

Network Analysis

Two methods, similarity-based and correlation-based networks, are used to build network features. Nodes in the similarity-based network represent firms and edges indicate the degree of similarity between them according to critical criteria; the similarity between organisations is then calculated using either the Euclidean or cosine distance. Using a correlation matrix as its basis, the correlation-based network establishes connections between businesses that have strong correlation ratings. The intercompany linkages and common traits may be better understood with the help of these networks.

Dataset Split

To assess the efficacy of the model, the dataset is partitioned into two parts: the training set and the testing set. The standard practice is to set aside 30% of the data for testing and 70% for training. In order to prevent overfitting and guarantee reliable assessment, K-fold cross-validation is used. The models are validated and generalised for unknown data in a comprehensive manner thanks to this separation.

In order to construct trustworthy machine learning models for predicting financial hardship, it is necessary to follow this systematic method to data preparation and analysis.

V. CONCLUSION

This research takes on the important problem of predicting financial hardship by combining conventional financial indicators with cutting-edge network-based characteristics. Unlike more traditional approaches, this model takes into account complex intercompany linkages by building networks based on similarity and correlation. The research uses sophisticated machine learning algorithms to analyse different methods' prediction powers, showing how important it is to include network analysis in financial forecasting. To make sure the predictions are strong and reliable, we use assessment metrics, feature engineering, and thorough preprocessing. Better financial market early warning systems might be possible with the help of the suggested technique, which demonstrates the promise of merging network science with machine learning to better understand financial systems.

VI. REFERENCES

1. Altman, E. I. (1968). Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy. *Journal of Finance*, 23(4), 589-609.
2. Beaver, W. H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, 4, 71-111.
3. Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 22, 59-82.
4. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
5. Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. *Springer-Verlag*.
6. Friedman, J., Hastie, T., & Tibshirani, R. (2001). The Elements of Statistical Learning. *Springer Series in Statistics*.
7. Larkin, J. M. (2010). The Role of Liquidity in Financial Distress Prediction. *Journal of Financial Economics*, 97(1), 14-34.
8. Tsai, C.-F., & Wu, J.-W. (2008). Using Neural Networks and Decision Trees for Bankruptcy Prediction. *Expert Systems with Applications*, 34(2), 2639-2649.

9. Sun, J., & Li, H. (2009). Financial Distress Prediction Using Support Vector Machines: Ensemble Models and Empirical Comparisons. *Expert Systems with Applications*, 36(2), 3357-3365.
10. Chen, N. F., & Zhang, F. (1998). Risk and Return of Value Stocks. *Journal of Business,* 71(4), 501-535.
11. Hosmer, D. W., & Lemeshow, S. (2000). Applied Logistic Regression. *Wiley*.
12. Cover, T., & Hart, P. (1967). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, 13(1), 21-27.
13. Kampen, J. K. (2013). The Value of Feature Selection in Financial Distress Prediction. *Computational Economics*, 42(3), 295-306.
14. Tang, Z., & Chi, Y. (2005). Neural Network Models for Financial Distress Prediction. *Journal of Forecasting*, 24(3), 189-200.
15. Amini, S., Parsaeian, M., & Farhadi, A. (2020). Network Analysis in Financial Markets: A Comprehensive Review. *Physica A: Statistical Mechanics and Its Applications*, 539, 122903.