A Methodological Review of Financial Distress Prediction Techniques

1Dr. Ajay Kr. Kansal,
2Mr. Shashank Sharma,
1,2School of Management, Gautam Buddha University, Gr. Noida

Abstract

In the recent times, cases of companies going into financial crisis have been increased to a great extent. The legal procedure to deal with these distress situations was limited. In 2016, Indian government introduced “Indian Bankruptcy Code, 2016” to deal with the financial distress of the individuals, partnerships and corporates. But, If there will be an EWS (Early Warning Systems) in place that can predict the possibility of going into financial distress well in advance. This will help financial decision makers in timely resolution of the concerned areas to prevent the distress in first place. This paper aims for a methodological review of the literature of the prediction techniques that have been used widely for Financial Distress Prediction.

Keywords: Financial Distress, Prediction techniques, AI (Artificial Intelligence), EWS (Early Warning Systems), Financial health, Crisis

Introduction

Financial distress refers to a situation where a company cannot pay or faces difficulties to pay off its obligations to creditors. If a company suffers from such situation for a long time it has every possibility to become financially distressed in near future. A company may be financially distressed if it is unable to pay its financial obligations as they mature (Lin, 2009).

Studying financial health in terms of financial distress prediction is always an exploratory topic of finance. The work was started by Beaver (1966), using univariate analysis, refined by Altman (1968) who introduced multivariate discriminant methodology using the financial ratios. Prediction of financial distress is important, as it gives a signal to the stakeholders and the investors of the company.

Financial reporting provides the primary source of independently verified information
to its stakeholders. Researchers generally benefit from a variety of different statistical methods to analyze data from historical financial reports to establish models and make decisions based on their estimations to conduct the research.

Although financial distress does not necessarily imply that the distressed firms will ultimately fail, a significant and persistent decline in a company’s financial performance may eventually result in bankruptcy, making investors and creditors suffer considerable financial loss.

Methodology

Firstly, we define a recent review period of last three years; we also ensured the inclusion of initial publications in this field of research. Then we conducted a keyword search that includes financial distress prediction, financial distress, crisis and bankruptcy.

We used these keywords to retrieve Scopus indexed articles from ScienceDirect, JSTOR, Emerald Insight, and Elsevier. We included papers being published in reputed academic journals in the field of accounting and finance with large numbers of citations and high impact factor.

The main challenge of Distress prediction starts with the selection of the prediction technique. Many advanced techniques for predicting bankruptcy have been developed. Statistical and deep learning techniques are the two broad categories used to predict financial distress Data Analysis Techniques: -

Following Are The Various Techniques Are Available For Prediction Analysis

Beaver (1966) studied the corporate distress initially and provided the prediction models
based on ratios. Beaver proposed two types of single variable analysis viz., profile analysis and univariate discriminant model. Through profile analysis for five years before failure, he found that the means of financial ratios in two groups were significantly different, and the gap was noticeable nearer to the failure.

**Traditional Statistical Techniques**

1. **Multivariate Discriminant Analysis (MDA)**


   In this paper author calculated bankruptcy probabilities using ratio analysis for the data on US manufacturing companies. The idea behind use of ratio analysis was that the failing company will exhibit very different ratio measurements than their counterparts. The ratios measuring profitability, liquidity and solvency were found to be most significant indicators. Using Multiple Discriminant Analysis (MDA), he converted the linear combinations of independent variables in scores (called Z scores) so that they could be classified into one of the two groups of bankrupt and non-bankrupt company.

   The final discriminant function estimated by Altman (1968) is as follows:

   \[ Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5 \]

   Where, \( X_1 = \) Working Capital/Total Assets;
   \( X_2 = \) Retained Earnings/Total Assets;
   \( X_3 = \) Earnings before Interest and Taxes/Total Assets;
   \( X_4 = \) Market Value of Equity/Book Value of Total Liabilities;
   \( X_5 = \) Sales/Total Assets;
   \( Z = \) Overall Index.

2. **Logistic Regression (Lr)** Logistic Regression Is A Type Of Generalized Linear Model

   Using simple linear regression is inappropriate when the variable to be predicted is binary; due to normality assumptions.


   An empirical study was conducted between 2007 and 2012 using a matched pairs research design with 308 observations, with half of them classified as distressed and non-distressed. Results confirm that in difficult situations prior to bankruptcy, the impact of board ownership and proportion of independent directors on business failure likelihood are
similar to those exerted in more extreme situations.

3. Decision Trees (Dt) It Is A Tree Branch Structure, Used In Predictive Modelling To Move From Observations To Conclusions.


This paper showed how a decision support system provides EWS of financial distress to a company one year ahead. This paper offered a practical solution for predicting financial distress in Romania by developing an integrated decision support system and analysing the effectiveness of various prediction models based on decision trees, logit and hazard models, as well as neural networks. From the four prediction models tested, best results were obtained by the CHAID decision tree model. The prediction accuracy of the classification tree was quite high, reaching over 90% in the testing phase, as compared to the neural networks (83.9%), the Hazard model with time invariant function (82.8%) or the single-period logit model (77.3%).

4. Naive Bayes (NB) a classification tool simply uses Bayes conditional probability rule. Each attribute and class label are considered random variable, and assuming that the attributes are independent, the naive Bayes finds a class to the new observation that maximizes its probability given the values of the attributes. Bayesian belief network (BBN) allow for the representation of dependencies among subsets of attributes. A BBN is a directed acyclic graph, where each node represents an attribute and each arrow represents a probabilistic dependence.

5. Bivariate Probit Model (BP) is typically used where a dichotomous indicator is the outcome of interest and the determinants of the probable outcome includes qualitative information in the form of a dummy variable where, even after controlling for a set of covariates, the possibility that the dummy explanatory variable is endogenous cannot be ruled out a priority.

Intelligent AI Techniques

1. Genetic Algorithm (GA) It is inspired from natural evolution, randomly generated rules are considered as an initial population. It is mostly used to find optimal or near-optimal solutions to difficult problems which otherwise would take a lifetime to solve.

The aim of this paper is to investigate whether network based variables can improve the predictive power of financial distress prediction. This study introduced a genetic algorithm (GA) approach to parameter selection in gradient boosting decision tree and integrated network-based variables for financial distress prediction.

The experiment results indicate that the introduction of network-based variables and GA based gradient boosting method for financial distress prediction can further enhance predictive performance in terms of accuracy, recall, precision, F-score, type I error, and type II error.


It is a search methodology belonging to the family of evolutionary computation. GP randomly generates an initial population of solutions. After that, the initial population is manipulated through various genetic operators to produce new populations.

3. Sentiment Analysis Using Deep Learning


Aspect based sentiment analysis aims to detect an aspect or features in a given text and then perform sentiment analysis of the text with respect to that aspect. This paper performed aspect-based sentiment analysis on the micro blogs and headlines of financial domain to predict the financial health.

4. Support Vector Machines (SVM) Use A Linear Model To Implement Nonlinear Class Boundaries By Mapping Input Vectors Nonlinearly Into A High-Dimensional Feature Space.


Empirical results indicate that for one year prior to financial distress, Support Vector Machine is the best classifier with an overall accuracy of 88.57%. Meanwhile, in the case of two years prior to financial distress, the hybrid model outperforms Support Vector Machine, Logit model, Partial Least Squares, and Artificial Neural Networks with an overall accuracy of 94.28%.

5. Text Mining and Big Data Analytics.

Text content in blogs, e-mails, tweets, community forums and other forms of textual communication constitutes information, analysing that information is what we call text analytics. Currently Text Analytics is often considered as the next step in Big Data analysis. Text analytics is applicable to most industries: it can help analyse millions of emails; you can analyse customers’ comments and questions in forums; you can perform sentiment analysis using text analytics by measuring the perceptions about a company, brand, or product. Text Analytics has also been called text mining, and is a subcategory of the Natural Language Processing (NLP) field.


In this review paper, it is concluded that the research focus is on stocks price prediction, financial fraud detection and market forecast utilizing online text mining and the current research trends of text mining are related to the need to analyse large amounts of data on websites and pages on social media, and to identify and test various text-mining techniques.

6. Artificial neural network (ANN) artificial neural networks were first created with the purpose to imitate the behaviour of the human brain. A neural network is the connection of elementary objects called the simple neuron.


The aim of this paper is to compare Logit and ANN. The monthly dataset covered 18 countries and a time span of January 1970 to June 2003. The results suggest that an ANN-based model is a highly promising tool for banking crisis prediction and it can clearly outperform the usual models such as the logit regression.

7. Multi-layer feed forward neural network (MLFF-NN) is one of the most common NN structures, as they are simple and effective, and have found home in a wide assortment of machine learning applications.


The aim of this paper was to develop an early warning system to predict financial crisis using Monthly data of 7 key macroeconomic and financial indicators of Turkish economy. Its found that predictive power of MLFNN is quite striking. Out-of-sample forecasts indicate that the Turkish economy remains at high risk due to major negative developments and potential political instability between 2014 and 2016.
Findings and Conclusion

We have provided a methodological literature review for the financial distress, and covered the techniques that have been known to deal with the problem of distress prediction, which are broadly divided to two sections: statistical and intelligent techniques. We have discussed their sub types and need of being implemented in the recent research. Also we have presented a quick reference of the recent literature.

An comparative analysis of the various techniques w.r.t. accuracy and precision is summarised below; (12)*

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Accuracy</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Discriminant analysis (LDA)</td>
<td>75.47</td>
<td>68.75</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>79.25</td>
<td>71.87</td>
</tr>
<tr>
<td>Decision Trees</td>
<td>86.79</td>
<td>78.12</td>
</tr>
<tr>
<td>Artificial Neural Networks (ANN)</td>
<td>90</td>
<td>94.11</td>
</tr>
<tr>
<td>Support Vector Machines (SVM)</td>
<td>90</td>
<td>94.11</td>
</tr>
<tr>
<td>Support Vector Machines with Genetic Algorithm (SVM-GA)</td>
<td>92.5</td>
<td>94.44</td>
</tr>
<tr>
<td>Support vector machines with particle swarm optimization algorithm (SVM-PSO)</td>
<td>95</td>
<td>94.73</td>
</tr>
</tbody>
</table>

References/Bibliography


