International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

Bankruptcy Score Indexing

Bhavik H. Prajapati[1], Akhand C. Dubey[2], Meghank S. Upadhyay[3], Prof. Neelam Phadnis[4]

[1,2,3]*Department of Computer Engineering, Shree L.R. Tiwari College of Engineering, Maharashtra, India*
[4]*Assistant Professor .Department of Computer Engineering, Shree L.R. Tiwari College of Engineering, Maharashtra, India*

---***---

Abstract - *Bankruptcy in the corporate world is one of the main drivers of the credit risk and gains primary attention from creditors and investors. The financial assets of companies are sold out to clear the debt which results in huge financial losses to the investors; here is a need to design effective strategies for prediction of bankruptcy to avoid financial crisis at an earlier stage. Bankruptcy can also be predicted using mathematical techniques, hypothetical models as well as soft computing techniques. Studies in bankruptcy prediction routinely adopt measures including firms' stock market trading information and accounting data from company's financial statements to forecast bankruptcy. This system focuses on predicting corporate bankruptcy by making use of textual disclosures from the SEC EDGAR section of 10-K & 10-Q annual filings of the firm & It was found that learning algorithms could enable users to predict bankruptcies with satisfying accuracy long before the final bankruptcy.*

***Key Words*: Natural language processing; Bankruptcy Prediction; Data Extraction; Co-variance vectorization model.**

1. INTRODUCTION

According to the U.S. Securities and Exchange Commission the U.S. Bankruptcy Code is stated as "the company stops all operations and goes completely out of business. A trustee is appointed to liquidate (sell) the company's assets, and therefore the money is employed to pay off debt." Corporate bankruptcy is one among the most drivers of the credit risk and gains primary attention from creditors and investors. The financial damage inflicted by corporate bankruptcy cannot be overstated.

The 2008-2010 financial crisis has also shown that in aggregate the corporate bankruptcy events have a profound influence on the economy. Corporate bankruptcy may incur a robust negative social cost and further propagate recession and thus jeopardize the economy at large. An accurate bankruptcy forecasting model, therefore, is effective to practitioners, regulators, and academic researchers alike. Regulators can use the model to watch the financial health of individual institutions and curb systemic risks.

Studies in bankruptcy prediction routinely adopt measures including firms' stock market trading information and accounting data from company's financial statements to forecast bankruptcy. Research has shown that accounting-based ratios and stock exchange data offer signals on whether a firm is financially healthy or may step into severe trouble like bankruptcy. Given the high impact of corporate bankruptcy events, researchers in operational research (OR) and AI (AI) further propose intelligent models to forecast bankruptcy. A common element of these models is that the application of market-based and accounting-based variables, which are usually constructed using numeric data during a well-structured format. Yet, there's growing recognition that text disclosure — a sort of unstructured, qualitative data — plays an equally important role in how information is conveyed to the general public. For example, a vast proportion of public firm's annual filings to regulatory agencies are textual disclosures.

This introduces machine learning models for forecasting corporate bankruptcy using textual filings which are filed quarterly. Although textual data is common, it's rarely considered within the financial decision network. We've proposed novel inputs extracted from the equity markets. As are often seen from the results, the new indicators improve the bankruptcy score considerably. This will be explained by the tendency of the equity markets to be highly predictive, not only of the health of a firm, but also of the health of the economy, which successively affects the creditworthiness of the firm. In this system we reviewed the matter of bankruptcy prediction using covariance matrix. From the various studies existing within the literature, it are often seen that neural network & deep learning are generally more superior to other techniques. Once this is often established, the logical next step for the research community is to enhance further the performance of covariance vectors for this application, perhaps through better training methods, better architecture selection, or better inputs. Learning model uses Glove word vector which returns the vector of word to urge Covariance vectors and extract features from textual data for prediction. We've constructed a comprehensive bankruptcy database of 1000 U.S. public companies.

The goal is to estimate how financial ratios can have different bankruptcy-indicating abilities across industries and time .The results from this study will increase the understanding among business researchers and academics on how financial ratios vary as bankruptcy indicators across industries and time, and encourage researchers for further research.

The final purpose is to match the prediction accuracy of the estimated models to ascertain what effect the industry-adaptation can wear the results.

2. LITERATURE REVIEW

In Literature review, we discuss about the various aspects of the project by taking reference of the existing projects that are similar to the makers of this current project.

Kalyan Nagaraj and Amulyashree Sridhar[1] have used different machine learning techniques and are employed to predict bankruptcy. Soft computing techniques helps in developing predictive models in finance. Variety of the favored soft computing techniques include Bayesian networks, logistic regression, decision tress, support vector machines, and neural networks.

Yi Qua, Pei Quan, Minglong Leid and Yong Shia[2] have reviewed the machine learning or deep learning models utilized in bankruptcy prediction, including the classical machine learning models like Multivariate Discriminant Analysis (MDA), Logistic Regression (LR), Ensemble method, Neural Networks (NN) and Support Vector Machines (SVM), and major deep learning methods like Deep Belief Network (DBN).In each model, the precise process of experiment and characteristics are going to be summarized through analyzing some typical articles.

Hironori Takeuchi, Shiho Ogino, Hideo Watanabe[3] have analyzed the sentences in financial reports in Japan and extracted key phrases/descriptions to predict bankruptcy. Our research revealed that if some particular expressions appear alongside the word ‘‘dividend’’ or ‘‘retained earnings’’ within the same section of an annual report, they were effective in distinguishing between bankrupt companies and non-bankrupt companies.

Shantanu Deshpande[4] have focused on new deep learning methods for bankruptcy forecast by assessing the predictive power of textual disclosures. The aim of the research was to use deep learning approach to forecast bankruptcy using textual disclosures.

Mai, Feng & Tian, Shaonan & Lee, Chihoon & Ma, Ling [5] introduced deep learning into the prediction of bankruptcy, using layers of neural networks to extract features from textual data from over 10000 U.S. public companies. It's been found that if some textual data (e.g. news, public report of companies) are conjunction with classical numerical data (e.g. financial ratio data), then the deep learning will yield superior performance in forecasting bankruptcy using textual disclosures, which will improve the accuracy of the prediction.

S. Kotsiantis, D. Tzelepis , E. Koumanakos and V. Tampakas[6] had reported here to research the efficiency of machine learning techniques in such an environment. to the present end, sort of experiments are conducted using representative learning algorithms, which were trained employing a knowledge set of 150 failed and solvent Greek firms within the recent period 2003-2004.It had been found that learning algorithms could enable users to predict bankruptcies with satisfying accuracy long before the last word bankruptcy.

3. PROBLEM DEFINATION

To build such a predictive model that will be robust and efficient in forecasting corporate bankruptcy based on not only the financial indicators but also the textual disclosure.

3.1 Corporate Finance

Corporate bankruptcy is a major challenge for the economic stakeholders including the investors, venture capitalist, etc. The major concern in predicting the reason of a business failure comprises various number of factors responsible for a financial crisis.

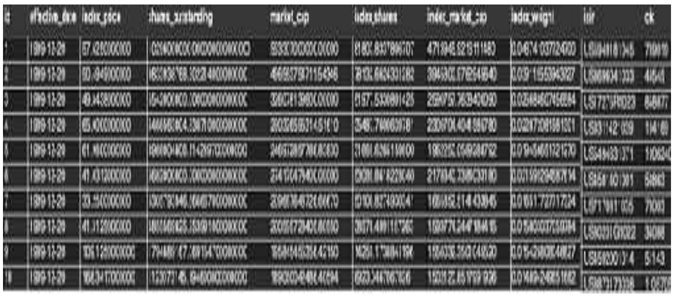
The system focuses on predicting bankruptcy of the firms by considering the following financial terms listed in the United States of America's stock market.

Central Index Key which is a unique identification number for company registered in United States of America.

There are two kind of filings which will be the first level in generation of the bankruptcy score.10-Q filings are generated annually; 10-K filings are registered quarterly, summing of these two filings will make four filings in the year which are extracted through CIK (central index key) which in turn are found through ISIN (International securities identification number) and company names. Each of these links consists of financial statements; commentaries are drawn out with the help of these 10k, 10Q filings links. These statements explains the company scenario for the past three months, on how the company managed through its financials comparing its various portfolios present in their organization.

3.2 Data Exploration

In this interface we will be considering the financial ratios (includes accounting data and equity trading data) and textual disclosures data from the 10-K annual filings for extraction of information pertaining to the cause of bankruptcy of an organization. The data will be collected from two sources: S Network for the equity trading data and Securities Exchange Commission (SEC) for textual disclosure data from 10-K & 10-Q filings. It will include yearly data of publicly traded organization from 1999 to 2019.



id	effective_date	index_price	shares_outstanding	market_cap	index_shares	index_market_cap	index_weight	isin	cik
1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Fig 1-Working Database

The primary sample of the financial ratios includes more than 450 firms with close to 45,000 to 50,000 firm-year observations. And for the numerical predictors, 13 predictor variables have been identified based on above literature works and by considering only those values that are responsible for companies' profitability, liquidity, and liability status. Some of the prominent parameters are company name, effective date, CIK (Central index key), ISIN (International security identification number), Index price, market capital, index weight etc.

4. SYSTEM ARCHITECTURE

Bankruptcy Score Indexing System forecasts corporate bankruptcy using textual filings which are filed quarterly. Learning model uses Glove word vector which returns the vector of word to urge Covariance vectors and extract features from textual data for prediction. We've have constructed a comprehensive bankruptcy database of 1000 U.S. public companies.

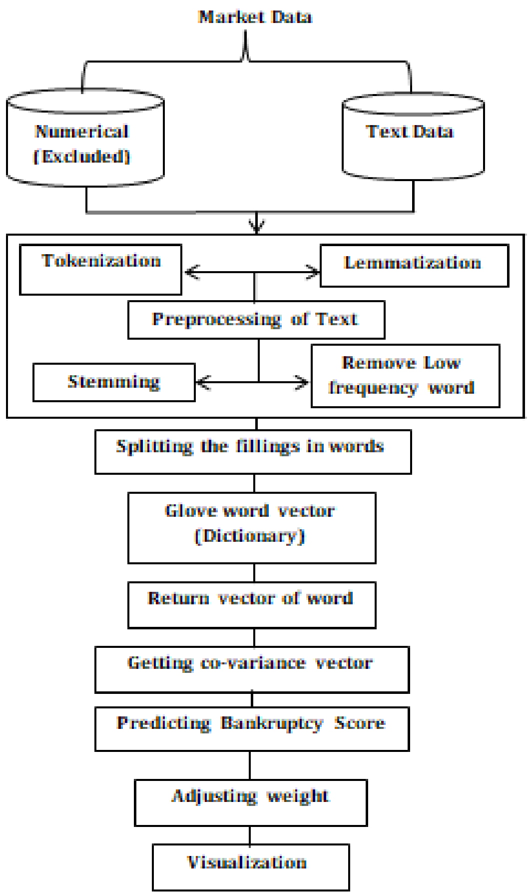


Fig -2: System Architecture

5. IMPLEMENTATION OF THE SYSTEM

Here we will discuss about how we implemented our system and is partially represented in a flowchart manner in Figure 1.

5.1 Text Pre-processing

For each linked 10-K filing, we remove the HTML tags, tables, and exhibits. Extraction of MD&A section is done using Python scripts. In the text pre-processing stage, we transform the original MD&A section from 10-K annual filings to plain-text documents in three phases: (1) We tokenize each filing into individual words using the Natural Language Toolkit (NLTK) ; (2) We also use NLTK stem to lemmatize each word and remove the inflectional forms of words and return them to understandable forms. For instance, connection and connecting become connect; (3) Removal of low-frequency words & stop words with help of NLTK corpus and only include 15,000 most frequent words. Such filtering procedure is a common practice in NLP as it can help reduce the complexity of statistical models.

5.2 Algorithms

In this part we have provided two simple steps to find bankruptcy score.

5.2.1 Mean Embedding Vectorizer

We first describe how we use word2vec a word embedding model, a learning layer for textual filings, to extract meaning from the texts and turn words into real-valued vectors i.e. a class which simply returns the vector of word; over here glove word vector is operated where each word is read and vector is given to each word. Dictionary is returned where all the words in corpus as key with their vector as value. Next, our model uses the out-turn from the word embedding layer as inputs to generate bankruptcy score. We compare two learning model architectures: Mean embedding vectoriser and covariance vector.
.

5.2.1 Co-variance Vector Model

We show how traditional models handle text features and describe these benchmark models. Over here effective transformation of each document to get a matrix of document, where n is the document length. In practice, each text filings varies by length and normalizing each document to length which varies according to the length of documents if a text is empty we should return a vector of zeros with the same dimensionality as all the other vectors. One of the core processing layer Covariance vectors which are calculated by transposing word vectors and calculating by taking upper triangular matrix as vector because of them being symmetric which includes objects from mean embedding vectoriser which suggests that this model architecture gives the average of every word vector dimension for all the words present in the document.

5.3 Bankruptcy Score

The bankruptcy score determines the financial health of the company and a curving point is identified to distinguish between failing and non-failing firms which are derived from co-variance vector model. If the bankruptcy score is below the cut-off point i.e <0.5 then the firm is classified as bankrupt and non-bankrupt if it is above the cut-off point i.e .0.5.

5.4 Weightage Adjustment

Fundamentally, due to the nature of the indexing, the risk was controlled internally through adjusting bankruptcy score weightage by deploying techniques to evaluate models performance and estimating the models accuracy by initially introducing cumulative distributive function and then probability distributive function which would then adjust the bankruptcy score according to firm's presence in the market shares that are outstanding.

6. RESULTS

After the procedure of building portfolio and structuring with appropriate functions, we equip decision rules that would be applied to risk management of each trade. Fundamentally, due to the nature of long-only strategy, the risk was controlled internally through adjusting bankruptcy score weightage. The following visualization displays the comparison between Benchmark Indexes which are designed to provide accurate coverage of publicly listed US stocks.

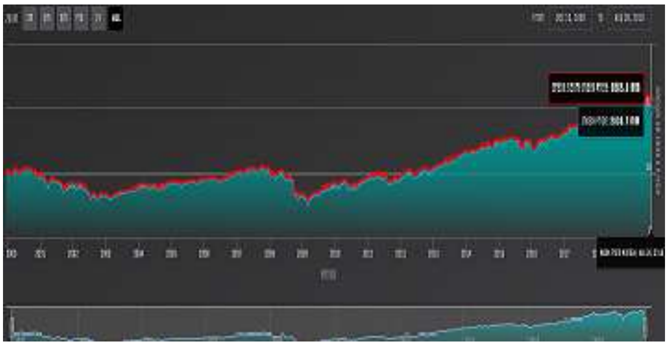


Fig -3 Comparison with Benchmark Index

7. CONCLUSION

In the domain of corporate bankruptcy, it can be noted that however large scale the firm is, it can still come down to a situation of bankruptcy and this can adversely affect a good segment of individuals including investors, employees, management, etc. The final purpose is to match the prediction accuracy of the estimated models to ascertain what effect the industry-adaptation can wear the results. The results from this model increases the understanding among business researchers and academics on how financial ratios vary as bankruptcy indicators across industries and time, and inspire researchers for further research.

8. REFERENCES

[1] Kalyan Nagaraj and Amulyashree Sridhar," A PREDICTIVE SYSTEM FOR DETECTION OF BANKRUPTCY USING MACHINE LEARNING TECHNIQUES", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.1, January 2015.

[2] Qu, Y., Quan, P., Lei, M., & Shi, Y. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. Procedia Computer Science,162,895899. doi:10.1016/j.procs.2019.12.65.

[3] Shirata, C. Y., Takeuchi, H., Ogino, S., & Watanabe, H. (2011). “Extracting Key Phrases as Predictors of Corporate Bankruptcy: Empirical Analysis of Annual Reports by Text Mining”. Journal of Emerging Technologies in Accounting, 8(1), 31–44. doi:10.2308/jeta-10182.

[4] Shantanu Deshpande.“Corporate Bankruptcy Prediction using Deep Learning Techniques” National College of Ireland.

[5] Mai, Feng & Tian, Shaonan & Lee, Chihoon & Ma, Ling, 2019. "Deep learning models for bankruptcy prediction using textual disclosures” ,European Journal of Operational Research, Elsevier, vol. 274(2), pages 743-758.

[6] S. Kotsiantis, D. Tzelepis , E. Koumanakos and V. Tampakas,” Efficiency of Machine Learning Techniques in Bankruptcy Prediction”, 2nd International Conference on Enterprise Systems and Accounting (ICESAcc’05) 11-12 July 2005, Thessaloniki, Greece.