7th National Conference on
"Role of Engineers in Nation Building" organized by VIVA Institute of Technology, Mumbai
(1st and 2nd March 2019)
International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

Bankruptcy Prediction: A survey on Financial distress and Solution

Mayur Bodar[1], Mihir Panchal[2], Sunny Maurya[3], Tatwadarshi Nagarhalli[4]
Department of Computer Engineering, Mumbai University, INDIA

Abstract— *Bankruptcy is a legitimate status of an individual or other entity that can't repay debts to loan bosses. The First biggest Financial depression occurred is due to the bankruptcy came in 1930 leading to need of bankruptcy prediction. Bankruptcy is basically a legal status of a person or any other entity that cannot repay debts to creditors. Hence the bankruptcy has a big impact on management, shareholders, employees, creditors, stakeholders, etc. Due to such impacts now the accurate prediction of bankruptcy has become a need in the financial world. This survey provides a synthetic and evaluative survey of issues in predicting bankruptcy. This paper focuses on various techniques with different parameters. In this paper we review the various algorithms like neural networks, Decision Tree, Random Forest, Logistic Regression, Bayesian Network, SVM. This paper compares these algorithms and details the efficient prediction algorithms and its uses. This study focuses on parameters like precision, error rate, performance and accuracy. This paper concludes that how these parameters can be used to predict bankruptcy accurately and efficiently*.

Keywords— Artificial Neural Network, Decision Trees, Economy, Logistic Regression, Support Vector Machine.

I. INTRODUCTION

In the recent years, the problem of corporate bankruptcy has attracted the attention of many stakeholders in the financial sectors such as business investors, market analyst, banking sectors, lawmakers and shareholders. Predicting business failure is not only important for decision making in financial institutions, but also determines the country's financial distress to some extent as wrong decision-making in financial institutions can have a catastrophic effect on national or sometimes a global scale. The current significance of bankruptcy prediction models has grown due to the recent world financial crisis. The importance of the area is due in part to the relevance for creditors and investors in evaluating the likelihood that a firm may go bankrupt. This crisis has seen an increase in the number of bankruptcies in several countries and has served to demonstrate that even the best international companies have to be continuously vigilant concerning their financial situation and the position of the companies they work.

II. LITERATURE REVIEW

In following paper, different techniques used for Bankruptcy Prediction are specified. Prediction of an enterprise bankruptcy is of great importance in economic decision making. A business condition of either small or large firm concerns local community, industry participants and investors, but also influences policy makers and global economy. Therefore, the high social and economic costs because of corporate bankruptcies have attracted attention of researchers for better understanding of bankruptcy causes and eventually prediction of business distress. The quantity of research in this area is also a function of the availability of data: for public firms which went bankrupt or did not, numerous accounting ratios that might indicate danger can be calculated, and numerous other potential explanatory variables are also available.

2.1 Bankruptcy Prediction Using Neural Networks [1]

In this paper G. Pranav and K. Govinda [1] have proposed to predict bankruptcy using Artificial Neural Network (ANN). It is mathematical model that is inspired by the information processing capabilities of our brain. The specialty of this model is that it is able to take into account the past experiences and hence make more accurate decision over a period of time. Neural network is one of the best way used for prediction but even though it's good there are a lot of factors which play vital role in the construction of good predicting model. To get the maximum throughput of the neural. First, it should have a good learning algorithm. Here, they have used Random Forest as a learning algorithm. It uses random forests or random decision forests which are a type of

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

classification and regression methods. They mainly operate by creating decision trees during the training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Second, it should have ideal classification rate so that the model can be used in real world scenarios. G. Pranav and K. Govinda [1] have used Sigmoid function as an Activation function for the neural network. The dataset that G. Pranav and K. Govinda [1] have used is Polish Companies' analyzed data which has 65 attributes but to predict bankruptcy they have considered only 3 such as Solvency, Earnings before Interest and Taxes (EBIT) and Liquidity. The three attributes that are considered for prediction are independent of each other and hence using these parameters to determine bankruptcy is not completely justified. Proposed method yields an error of 4.4349% after 150 epochs. However, the result obtained from random forest classification had a higher degree of error, specifically 5.1954%. This paper mainly focuses on creating Neural Network for predicting bankruptcy and they achieved a yielded 4.43% of error rate with limited data set.

2.2 The impact of payment delays on bankruptcy prediction [2]

In this paper, Z. Fatima and S. Achchab [2] analyzed suppliers and customers' payment delays and based on that they predicted the failure. One of the challenges that can affect the performance of a bankruptcy prediction model is the definition of the appropriate financial ratios to approach the failure distress. In this study, Z. Fatima and S. Achchab [2] have considered 3 subjects. The first one has the objective to study the contribution of variables selection models by comparing Multivariate Discriminant Analysis, Logistic Regression and Decision Trees. The final goal was to study the impact of customers and suppliers' payment delays to the self-financing capacity and risk failure based on neural network models. To identify the significant and the minimum set of variables that can discriminate bankruptcy and healthy firms; they applied two fields of classification techniques, statistic methods and decision trees as artificial intelligence techniques to the same variables. Multivariate Discriminant Analysis (MDA) and Logistic Regression (LR) are the two Statistical models that they used. For Decision tree, they have used CRT (Classification and regression tree) and CHAID (Chi-squared automatic interaction detector). The database used contains annual financial statements data for a sample of Moroccan firms which has lot of missing data leading to misleading or improper prediction. According to the research that they have done the CRT model demonstrated the highest overall accuracy level (85.7%) and LR had the second performance measurement (85.3%). Even though the CHAID model did not perform as well as CRT and LR, it still produces a considerably high accuracy rate (83.5%) and MDA shows the least accuracy of 41.5%. They have created Neural Network with Levenberg-Marquardt algorithm and it gives accuracy of 85.6%. Z. Fatima and S. Achchab [2] used different methods and tried to make work for predicting bankruptcy with limited dataset and even though they that got 86.6% as maximum but when worked with increased data this model tend to fail.

2.3 Banks Bankruptcy Risk Forecasting with Application of FNN [3]

For the solution of bankruptcy, Y. Zaychenko [3] applied fuzzy neural networks' adaptive neuro-fuzzy inference system (ANFIS) and Takagi-Sugeno-Kang (TSK) and fuzzy Group Method of Data Handling (GMDH) is used. The experimental investigations of the suggested fuzzy models application for bankruptcy risk forecasting of European banks are carried out and comparison with classical methods is performed. For training they used 115 banks of Europe and testing sample was of 50 banks. After the analysis they found out that FNN TSK gives better results than ANFIS. When results of FGMDH application compared with the results of FNN TSK it was clear that neural network gives better results on short forecasting. For the comparison they also implemented the regression models, Logit - model, Probit-model. As per their experiments the fuzzy methods FNN TSK, ANFIS and FGMDH give much better forecasting results than statistical methods: regression models; logit - model; probit-model. Y. Zaychenko [3] found that FNN network TSK gives better results than FNN ANFIS. The worst results (maximal error rate) was obtained by conventional linear regression (18%) while the application of probabilistic logit- and probit –models allows to decrease the error rate to 14-16%. He used dataset of European banks according to the International Accountant Standard IFRS. The annual financial indices of about 300 European banks in 2004- 2008 years were collected, preceding the start of crisis of bank system in Europe in 2009 year. Using FNN for predicting bankruptcy was good idea but this gave them much higher error rate and hence failing to be used for prediction.

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

2.4 Anomaly detection methods for bankruptcy prediction [12]

In this paper, S. Fan et. al. [12] have proposed anomaly detection method to detect bankruptcy with Multivariate Gaussian distribution, One-Class SVM and Isolation Forest. Anomaly detection search best combination of financial ratios and corporate governance indicators for bankruptcy prediction. Multivariate Gaussian distribution mainly used support fraction, contamination and store precision attributes. Using a covariance matrix, it is able to fit a robust covariance estimate to the overall pattern. One-Class SVM is a type of unsupervised learning, they fit the model without any class information. To gain the decision boundary, one-class SVM basically separates all training data from the origin input feature space. One-Class SVM is used kernel, gamma, nu, shrinking as hyper parameter. Isolation Forest is able to detect anomalies purely without relying any distance or density measure. The method applies a binary tree called iTree for separating and isolating instances effectively. More specifically, isolation tree randomly selects a feature and then selects a split value at random. Hence, instances with short path lengths from root node to leaf node have a higher probability that it belongs to outliers and also it requires linear time complexity and small memory as compared to conventional method. The experiment dataset are from UCI machine learning repository. It describes bankrupt about financial condition of Polish companies. 1st Year, 2nd Year, 3rd Year, 4th Year and 5th Year cases with respect to the ratio of positive to negative samples is Ratio 1:25, 1:24, 1:20, 1:18 and 1:13 respectively. The proposed system focuses on Multivariate Gaussian distribution, one-class SVM and Isolation Forest to detect bankruptcy Prediction which gives the positive and negative samples Ratio.

2.5 An Attribute Selection Based Classifier to Predict Financial Distress [5]

In this paper, C. Cheng and C. Chan have proposed financial distress prediction (FDP) has become increasingly essential in resolving corporate financial risk. Five classification methods are utilized to identify financial distress in this paper: Decision tree C4.5, IBK, SVM, Random Forest, RBF Network. Financial distress, also called financial crisis which uses feature selection or variable selection to predict bankruptcy, is one of special optimization techniques. The Proposed model to predict Data pre-processing the financial database comes from Taiwan Economic Journal (TEJ) Corporation. Attribute selection six attribute selection methods, Chi-square, Filtered, Gain Ratio, Information Gain, ReliefF, and Symmetrical are used to select relevant and crucial attributes. First, a group of variables generated from six attribute selections are ranked according to their ratios separately, and then system cumulate ratios of variables on the top 95% to be the core attributes. Some classification methods, including Decision tree C4.5, IBK, Support vector machine, Random Forest, and RBF Network, are used to compare the classification models. The proposed system focuses on the feature selection and variable selection to predict bankruptcy with special optimizing techniques.

2.6 A Hybrid Switching PSO Algorithm and Support Vector Machines for Bankruptcy Prediction [6]

In the recent AI approaches, SVM has been used successfully for bankruptcy prediction and became increasingly prevalent. It is based on the statistical learning theory which is different from ANN, it overcame the problem of ANN easy to fall into local minima. Also, SVM has more strong classifying ability and outperformed in generalization performance than ANN. The main idea of Y. Lu, et. al. [6] in this paper is to focus on predicting bankruptcy with the new model of SVM augment by SPSO. The core merits in it is make full of use advantage of Switching PSO Algorithm (SPSO) to search for optimal kernel parameter of SVM. The Data sets that they have used are from UCI Machine Learning Repository donated on 9th Feb, 2014, which consists of 143 Non-Bankruptcy sample and 107 Bankruptcy sample, the total number is 250 sample data sets. Each attribute contains three parameters. They are P(positive), A(Average) & N(negative). Each class contains B(Bankruptcy) & NB(Non-Bankruptcy). First they pre-processed the data by regulation, second they randomly selected relevant data sets as training sample and testing sample, third they randomly initialized parameters and utilized SPSO algorithm to accelerate the learning speed and improve generalization performance of SVM, finally, improve the prediction accuracy of bankruptcy by optimal SVM model. The proposed system mainly focuses on the new model of SVM augment by SPSO to predict bankruptcy which gives the better results as compared to other optimization algorithms.

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

2.7 Bankruptcy Prediction using Data Mining Techniques [7]

In this paper, the majority of the analysis are related to bankruptcy prediction models were performed through WEKA machine learning tool. The dataset that M. Wagle et. al. [7] have used in this paper contains 240 cases, 112 of which are bankrupted cases and 128 are successful cases. They used two popular approaches in WEKA for attribute selection, they were wrapper method and filter method. M. Wagle et. al. [7] have used five classifiers for the attribute selection process which were Bayesian network, decision tree, logistic regression, neural networks and SVM. After comparing the results, the neural networks with the complete set of features produces the highest level of accuracy (70.83%). In order to increase the accuracy of the prediction models, they found the two common techniques in WEKA called the boosting technique and the bagging technique, are applied onto the classifiers to increase their prediction accuracy. After comparing the results, the neural networks using attributes selected by the filter method and applying the bagging technique produces the highest level of accuracy (85.33%). The Proposed system focuses on the neural networks which tends to produce more accurate results and yields the highest level of accuracy over the other classifiers by applying the bagging technique.

2.8 Performance Enhancement of SVM Ensembles Using Genetic Algorithms in Bankruptcy Prediction [10]

D.Kang et. al. [10] paper proposes that the genetic algorithm based coverage optimization techniques of the SVM ensemble to solve the multicollinearity problem. The, D. Kang et.al. [10] paper proposes the hybrid system focused on the resolution of coverage optimization of the SVM ensemble (CO-SVM) in order to improve the performance of the SVM ensemble. The performance of COSVM have shown that the COSVM was effectively applied in the stable performance enhancement of SVM ensembles through the choice of the classifiers by considering the correlations of the ensemble. First of all, they mainly focused on resolving the coverage optimization. But, the most fundamental problems in ensemble is decision optimization process to find optimal combination function. Secondly, the Ensemble learning has a problem with the noise, which in the learning samples distort the classification boundary of learning algorithms like SVM and degrade the learning performance. To deal with those outliers, different SVM ensemble methods like Probabilistic Roulette Selection, Automatic Feature Selection, etc. have been proposed. The proposed system focuses on the genetic algorithm based coverage optimization techniques of the SVM ensemble (CO-SVM) to predict bankruptcy which gives accurate results than the SVM & DT.

2.9 Using Bayesian Networks for Bankruptcy Prediction: Empirical Evidence from Iranian Companies [9]

A. Aghaie et. al. [9] paper proposes that several significant methodological issues were related to the use of naive Bayes Bayesian Network (BN) models to predict the bankruptcy. First, they provide two different methods that guide the selection of predictor variables from a pool of potential variables. The 4 variables that have important correlations with the variable of interest, the status of bankruptcy are selected from a pool of 20 potential predictors. The first naive Bayesian Network consisting of these selected variables have an average prediction accuracy of 90% for the bankruptcy sample and 89% for the non-bankruptcy sample. In second model A. Aghaie et. al. [9] investigated the impact on a naive Bayes model's performance of the number of states into which continuous variables were discretized. They found that the model's performance was the best with the continuous variables being discretized into 4 states. In this paper researcher recommend banks and financial institutes to use these naive Bayes models for loan making decisions to avoid bankruptcy. The proposed system focuses on two method of Bayes Bayesian Network to predict bankruptcy in which first method, only variables that have important correlations with the variable of interest and second method they investigated the impact on a naive Bayes model's performance of the number of states into which continuous variables were discretized.

2.10 Applying Decision Tree to Predict Bankruptcy [11]

In this paper, E. Zibanezhad et. al. [11] have used Clementine software and the method of classification and regression tree are used for mining financial variables. The firms accepted in Tehran Stock Exchange(TSE) are selected as studied society and the used instance consists of two groups of bankrupt and non-bankrupt firms. The variables predicting bankruptcy are divided in to 4 following main categories: Liquidity ratios, which indicate firm's ability for short term commitments. Leverage ratios, which reflect debts and commitment situations of the firm. Activity ratios, which indicate the effective using method of under controlled

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

properties of a firm. Profitability ratios, which reflect the pure yield of selling and properties outcome. Using CLEMENTINE data mining software, a proper model was selected to classify the data and proper patterns were extracted. The biggest goal of classification is to obtain patterns for prediction of bankrupt and non-bankrupt firms in a sample using firms' financial ratios amount. In this study, C&R Tree algorithm was used to make the tree. The functional method of this algorithm is called surrogate splitting which shares the records reached to each tie in to two pats and creates children for that tie which is refined more than their parent. The 94.5% accuracy in training data and 90% in test data indicate the efficiency of this method in predicting the bankruptcy. which shares the records reached to each tie in to two pats and creates children for that tie which is refined more than their parent. The 94.5% accuracy in training data and 90% in test data indicate the efficiency of this method in predicting the bankruptcy.

III. ANALYSIS TABLES

The following Table-1 gives the analysis of literature papers on Bankruptcy Prediction which focuses on Artificial Neural Network, Logistic Regression, Support Vector Machines, Bayesian Networks that uses different datasets and have different accuracy.

TABLE-1
ANALYSIS TABLE

S. No	Title of Paper	Techniques	Dataset	Accuracy
1.	Bankruptcy Prediction Using Neural Networks [1]	Artificial Neural Network (ANN). Random Forest as learning algorithm.	Polish Dataset. In that they have used 10503 financial statements	ANN error rate: 4.4349% RF yields: 5.1954%
2.	The impact of payment delays on bankruptcy prediction [2]	Multivariate Discriminant Analysis (MDA), Logistic Regression (LR) and Decision Trees. For Decision tree, they have used CRT (Classification and regression tree) and CHAID (Chi-squared automatic interaction detector).	The database used contains annual financial statements data for a sample of Moroccan firms.	CRT: 85.7% LR: 85.3% CHAID: 83.5% MDA: 41.5% NN: 85.6%
3.	Banks Bankruptcy Risk Forecasting with Application of FNN [3]	Adaptive Neuro-fuzzy Inference System (ANFIS) and Takagi-Sugeno-Kang (TSK) and fuzzy Group Method of Data Handling (GMDH) is used.	They have used 115 banks of Europe and Testing sample was of 50 European banks	Maximum error rate LR:18% while the application of probabilistic logit- and probit –models allows to decrease the error rate to 14-16%.
4.	An Attribute Selection Based Classifier to Predict Financial Distress [5]	1.Decision tree C4.5(DT) 2. IBK 3. SVM 4. Random Forest (RF) 5.RBF Network	They used financial database which came from Taiwan Economic Journal (TEJ) Corporation.	In this method 34 attribute gives the 98.74% Accuracy among the all methods.
5.	A Hybrid Switching PSO Algorithm and Support Vector Machines for Bankruptcy Prediction [6]	They have used SVM, Genetic Algorithm (GASVM), Particle Swarm Optimization (PSOSVM), and Switching Particle Swarm Optimization (SPSOSVM).	The Data sets that they have used are from UCI Machine Learning Repository donated on 9th Feb, 2014.	SVM: 89.7561% GASVM: 90.8764% PSOSVM: 94.4127 SPSOSVM: 93.2063
6.	Bankruptcy Prediction using Data Mining Techniques [7]	They have used Bayesian Networks (BN), Decision Tree (DT), Logistic Regression(LR), Neural Networks(NN) and SVM.	It analyzes a dataset of 120 companies using different data mining techniques.	NN gives (85.33%), BN gives (73.33%), LR gives (72.50%), & DT (69.16%)

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

7.	Using Bayesian Networks for Bankruptcy Prediction: Empirical Evidence from Iranian Companies [9]	Naïve Bayes Bayesian Network (BN)	They have used Tehran Stocks Exchange (TSE).	Naïve Bayes gives 90% accuracy on predicting and 89% on Non-Predicting Bankruptcy.
8.	Performance Enhancement of SVM Ensembles Using Genetic Algorithms in Bankruptcy Prediction [10]	Decision Trees (DT), Support Vector Machine (SVM), and Coverage Optimization of SVM ensemble (CO-SVM).	Commercial bank in Korea contains 1,200 externally audited manufacturing firms.	DT:75.79% CO-SVM:77.83% SVM:73.07%
9.	Applying Decision Tree to Predict Bankruptcy [11]	They have used Clementine software and the method of classification and regression tree are used for mining financial variables.	Tehran Stock Exchange (TSE) during the 1996 to 2009.	The 94.5% accuracy in training data and 90% in test data.
10.	Anomaly detection methods for bankruptcy prediction [12]	They used Multivariate Gaussian distribution (MG), One-Class SVM and Isolation Forest (IF)	They used Polish Dataset.	MG: 89% SVM: 92% IF: 93%

IV. CONCLUSION

The paper provides a brief survey on various machine learning algorithm which are Random Forest, Linear Regression, and k-means. The survey gives an outcome that larger dataset will provide better result. The survey has shown that numeric prediction, categorical prediction or classification and clustering can be implemented using machine learning algorithm. The study has shown that Random forest gives better results in case of larger dataset whereas K-means is faster clustering algorithm comparing other unsupervised learning algorithm. Simple Linear regression provide better accuracy on different datasets.

REFERENCES

[1] G. Pranav, K. Govinda, "Bankruptcy Prediction Using Neural Network", International Conference on Inventive Systems and Control (ICISC), 2018, pp. 248-251.
[2] Z. Fatima, S. Achchab, "The impact of payment delays on bankruptcy prediction", 3rd International Conference of Cloud Computing Technologies and Applications, 2017.
[3] Y. Zaychenko, "Banks bankruptcy risk forecasting with application of FNN", 11th International Scientific and Technical Conference Computer Sciences and Information Technologies, 2016, pp.196-199.
[4] S. Karlos, S. Kotsiantis, N. Fazakis, "Effectiveness of semi-supervised learning in bankruptcy prediction", 7th International Conference on Information", Intelligence, Systems & Applications (IISA), 2016.
[5] C. Cheng, C. Chan, "An attribute selection based classifier to predict financial distress", 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, 2016, pp. 1119-1122.
[6] Y. Lu, J. Zhu, N. Zhang, "A hybrid switching PSO algorithm and support vector machines for bankruptcy prediction", International Conference on Mechatronics and Control (ICMC), 2014, pp. 1329-1333.
[7] M. Wagle, Z. Yang, Y. Benslimane, "Bankruptcy prediction using data mining techniques", 8th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES), 2017.
[8] G. Kumar, S. Roy, "Development of Hybrid Boosting Technique for Bankruptcy Prediction", International Conference on Information Technology (ICIT), 2016, pp. 248-P253.
[9] A. Aghaie, A. Saeedi, "Using Bayesian Networks for Bankruptcy Prediction: Empirical Evidence from Iranian Companies", International Conference on Information Management and Engineering, 2009, pp. 450-455.
[10] D. Kang, M. Kim, "Performance enhancement of SVM ensembles using genetic algorithm in bankruptcy prediction", 3rd International Conference on Advanced Computer Theory Engineering and (ICACTE), 2010, pp- V2-154 – V2-158.
[11] E. zibanezhad, D. Foroghi, A. Monadjemi, "Applying decision tree to predict bankruptcy", IEEE International Conference on Computer Science and Automation Engineering, 2011, pp. 165-169.
[12] S. Fan, G. Liu, Z. Chen, "Anomaly detection method for bankruptcy prediction", 4th International Conference on Systems and Informatics (ICSAI), 2011, pp. 1456-1460.
[13] https://en.wikipedia.org/wiki/Machine_learning , Last Accessed on 15th Sept. 2018.
[14] https://en.wikipedia.org/wiki/Random_forest , Last Accessed on 16th Sept. 2018.
[15] https://en.wikipedia.org/wiki/Linear_regression , Last Accessed on 16th Sept. 2018.