Volume 1, Issue 2 (2019)
Article No. 8
PP 1-6

VIVA-TECH IJRI
ANNUAL RESEARCH JOURNAL
VIVA-TECH INTERNATIONAL JOURNAL FOR RESEARCH AND INNOVATION
ANNUAL RESEARCH JOURNAL
ISSN(ONLINE): 2581-7280

An Innovative Approach to Predict Bankruptcy

Mihir H. Panchal[1], Mayur B. Bodar[1], Sunny R. Maurya[1], Tatwadarshi P. Nagarhalli[1]

[1](*Computer Engineering Department, VIVA Institute of Technology, India*)

Abstract: *Bankruptcy is a legal status of a person or other organization that cannot repay their debts to creditors. Bankruptcy prediction is the task of predicting bankruptcy and by doing various surveys we can avoid financial distress of firms. It is a huge area of accounting and finance research. The significance of this area is an important part of financial specialists and creditors in assessing the probability that a firm may go bankrupt or not. Estimating the risk of corporate bankruptcies is very important as the effect of bankruptcy is on a global level. The aim of predicting financial distress is to develop a predictive model that combines various economic factors which allow foreseeing the financial status of a firm. In this domain, various methods were proposed that were based on neural networks, Support Vector Machines, Decision Trees, Random Forests, Naïve Bayes, Balanced Bagging and Logistic Regression. In this paper, we document our observations as we explore and build a Restricted Boltzmann Machine to Bankruptcy Prediction. We started by carrying out data pre-processing where we impute the missing data values using Mean Imputation. To solve the data imbalance issue, we apply the Synthetic Minority Oversampling Technique (SMOTE) to oversample the minority class labels. Finally, we analyze and evaluate the performance of the model.*

Keywords – *Artificial Neural Network, Decision Trees, Logistic Regression, Naïve Bayes, Random Forests, Restricted Boltzmann machine, Support Vector Machine.*

1. INTRODUCTION

In 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 bankruptcy 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 [1].

Predicting bankruptcy is of great importance in financial decision making. No matter how big or small a company is when it goes bankrupt it affects everyone on a global level and hence prediction of enterprise bankruptcies is very necessary. Researchers have been working on this domain for a good amount of time. They tried to understand the reasons for bankruptcy and eventually trying to avoid bankruptcy. The research on this domain is also functional to the availability of data.

The aim of the bankruptcy prediction is to predict the financial condition of a company and its future perspectives within the context of long-term operation on the market. It is a vast area of finance and econometrics that combines knowledge about the historical data of prosperous and unsuccessful companies. Typically, enterprises are quantified by numerous indicators that describe their business condition that is further used to induce a mathematical model using past observations.

There are many related works which focus on predicting the best model for the given financial decision-making problem. The proposed system not simply examine the probability of liquidation but also to analyze the best training algorithm. The best training model will be a retreat based on the highest classification rate. In the proposed method system, we use "Restricted Boltzmann machine".

The dataset is based on the Polish companies' bankruptcy data analyzed over 2000-2012. The informational collection is assembled by Emerging Markets Information Services (EMIS).

Volume 1, Issue 2 (2019)
Article No. 8
PP 1-6

2. RELATED WORKS

G. Pranav and K. Govinda [1] have proposed to predict bankruptcy using Artificial Neural Network (ANN). The specialty of this model is that it is able to take into account past experiences and hence make a more accurate decision over a period of time. Here, they have used Random Forest as a learning algorithm. The 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. 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.

Z. Fatima and S. Achchab [2] analyzed suppliers and customers' payment delays and based on that they predicted the failure. In this study, Z. Fatima and S. Achchab [2] have used Multivariate Discriminant Analysis, Logistic Regression, and Decision Trees. The database used contains annual financial statements data for a sample of Moroccan firms which has a lot of missing data leading to misleading or improper prediction. Z. Fatima and S. Achchab [2] used different methods and tried to make work for predicting bankruptcy with the limited dataset.

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 probably of work under fuzzy and incomplete data to use expert knowledge in a form of fuzzy methods. For training, they used 115 banks of Europe and the testing sample was of 50 banks

S. Fan et. al. [12] have proposed anomaly detection method to detect bankruptcy with Multivariate Gaussian distribution, One-Class SVM and Isolation Forest. The experiment dataset is from the UCI machine learning repository. It describes bankrupt about the 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 are Ratio 1:25, 1:24, 1:20, 1:18 and 1:13 respectively.

C. Cheng and C. Chan [5] have proposed financial distress prediction (FDP) has become increasingly essential in resolving corporate financial risk. Five classification methods utilized to identify financial distress are Decision tree C4.5, IBK, SVM, Random Forest, and RBF Network. Data pre-processing the financial database comes from Taiwan Economic Journal (TEJ) Corporation. The proposed system focuses on the feature selection and variable selection to predict bankruptcy with special optimizing techniques.

Y. Lu, et. al. [6] is to focus on predicting bankruptcy with the new model of SVM augment by SPSO. The core merits in it 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 the UCI Machine Learning Repository donated on 9th Feb 2014, which consists of 143 Non-Bankruptcy samples and 107 Bankruptcy sample, the total number is 250 sample data sets. Each attribute contains three parameters.

The dataset that M. Wagle et. al. [7] have contains 240 cases, 112 of which are bankrupted cases and 128 are successful cases. Observations come from two to five years before bankruptcy took place. Among these companies, 56 went bankrupt two to five years later. The data from the first year of all the 120 companies will be used as training set to train the classifier and the second-year data of all the 120 companies will be used as the testing set to test the classifier. 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. 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.

G. Kumar et. al. [8] objective of bankruptcy prediction was to determine whether an organization or financial firm will go bankrupt or not. The objective of boosting algorithm is to assign more weights to the misclassified instances so that the learner can focus more on them for the succeeding round to classify it accurately. The dataset they had obtained from 500 French industrial firms during the year 2002 and 2003. The proposed system focuses on various boosting algorithms like Logit-Boost, NFS-Boost to predict bankruptcy which gives better results after comparing with the other boosting algorithms.

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. The proposed system focuses on two methods 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.

D. Kang et. al. [10] paper proposes that the genetic algorithm-based coverage optimization techniques of the SVM ensemble to solve the multicollinearity problem. Considering this background, D. Kang et.al. [10] 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.

E. Zibanezhad et. al. [11] have used Clementine software and the method of classification and regression tree are used for mining financial variables. The data collected from financial statements of firms accepted in Tehran Stock Exchange (TSE) from 1996 to 2009 which contained a total of 25 required financial ratios. The

Volume 1, Issue 2 (2019)
Article No. 8
PP 1-6

structure of the decision tree is a tree structure similar to a flowchart the biggest tie in a tree is the root tie and the leaf ties show groups or group distribution.

3. PROPOSED SYSTEM

The paper proposes Bankruptcy prediction using Restricted Boltzmann Machine (RBM). Restricted Boltzmann Machines (RBMs) are neural networks that belong to so called Energy Based Models. This type of neural networks may be not that familiar, yet this kind of neural networks gained big popularity in recent years in the context of the Netflix Prize where RBMs achieved state of the art performance in collaborative filtering and have beaten most of the competition. A restricted Boltzmann machine (RBM) is generative stochastic artificial neural network that can learn a probability distribution over its set of inputs.

Polish dataset has been used for proposed system. It is hosted by the University of California Irvine (UCI) Machin Learning Repository which is a huge repository of freely accessible datasets for research and learning purposes for the Machine Learning/Data Science community. This information was collected from the Emerging Markets Information Service [6] (EMIS), which is a database containing data on developing markets far and wide. The bankrupt organizations were examined in the period 2000-2012, while the as yet working organizations were assessed from 2007 to 2013.

Table 1: Summary of the Polish bankruptcy dataset

Dataset characteristic	Multivariate			
Number of Features	64			
Number of Instances	**Data**	**Total Instances**	**Bankrupt instances**	**Number of Instances**
	1st year	7027	271	6756
	2nd year	10173	400	9773
	3rd year	10503	495	10008
	4th year	9792	515	9227
	5th year	5910	410	5500
Feature characteristics	Real values			
Has missing data?	Yes			
Associated tasks	Classification			
Date donated	04-11-2016			

In table 2, 2nd column display, the total number of instances in all dataset and 3rd column display the number of instances or rows with missing values for at least one of the features 4th Column display the number of instances that would remain in each dataset if all rows with missing values were dropped. 5th Column displays the percent of data loss if all the rows with missing data values were dropped. As the data loss rate in most of the datasets is more than 50%, it is now clear that we cannot simply drop the rows with missing values, as it leads to a loss in the representativeness of data.

Table 2: Missing Data

Data Set	**#Total Instances**	**# Instances with missing values**	**# Instances that would remain if all rows with missing values were dropped**	**% Data loss if rows with missing values were dropped**
Year 1	7027	3833	3194	54.54 %
Year 2	10173	6085	4088	59.81 %
Year 3	10503	5618	4885	53.48 %
Year 4	9792	5023	4769	51.29 %
Year 5	5910	2879	3031	48.71 %

There are mainly three issues comes in missing data
1. Missing data information can introduce a major amount of bias.
2. Handling and analysis are of the data more difficult.
3. Create more reductions in efficiency.

Volume 1, Issue 2 (2019)
Article No. 8
PP 1-6

Table 3 display below summarizes the populations of class labels. 2nd Column displays the total instances, while the 3rd Column and 4th Column display the total number of instances with the class as Bankrupt and Non-Bankrupt respectively. While 5th Column shows the population percentage of the minority class, i.e., the Bankruptcy class label, among the total population of the dataset. These numbers in column 5 display that there is a huge data imbalance.

Table 3: Data Imbalance

Data Set	# Total Instances	# Bankrupt instances in this forecasting period	# Non-Bankrupt instances in this forecasting period	Percentage of minority class samples
Year 1	7027	271	6756	3.85 %
Year 2	10173	400	9773	3.93 %
Year 3	10503	495	10008	4.71 %
Year 4	9792	515	9277	5.25 %
Year 5	5910	410	5500	6.93 %

To deal with missing data in the proposed system mean imputation is used. It is the process in which all the missing value replace with their mean of that context variable. Once completed dataset will not contain any missing values. Data imbalance will be dealt by oversampling the dataset using Synthetic Minority Oversampling Technique (SMOTE). It is a majorly used oversampling technique.

Figure 1 shows the System flow of proposed system. In this above-mentioned dataset will be taken and pre-processing will be done on that to create balanced data. This balanced data will not have any missing data and will be balanced for both the instances. This balanced data now will be used to train Restricted Boltzmann Machine.

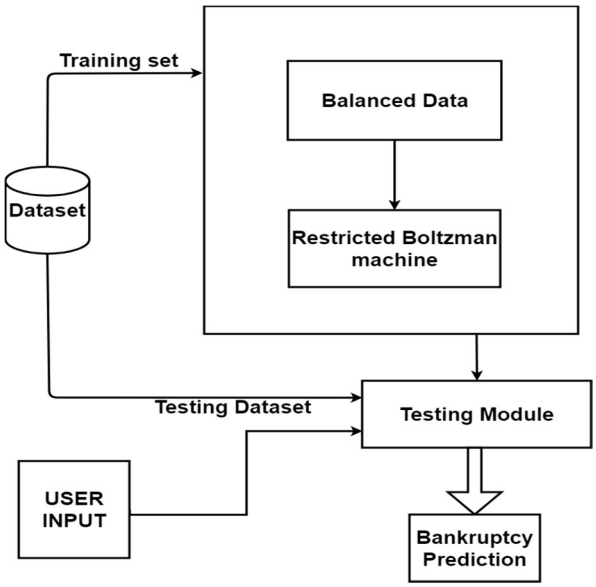

Figure 1: Proposed System

As the dataset used contains 64 attributes. Users will have to upload provided .csv file for prediction which contains the 64 attributes with their descriptions.

4. RESULTS AND ANALYSIS

Proposed system outperformed all the previously implemented models with giving much higher accuracy of 70.86%. While previously implemented best model i.e. Support Vector Machine (SVM) struggled to give accuracy higher than 60%. Hence our proposed model is more efficient to predict bankruptcy and gives much

better performance and result. Table 4 shows the comparison of accuracy of different models and proposed system. In this we can clearly see that proposed system has higher accuracy than other models. Based on this predictions user can take required actions and by using proposed system for prediction user can foresee the financial status of company.

Table 4: Comparative Analysis

Model	Dataset	Accuracy
Multivariate Gaussian distribution (MG)	Polish Dataset	0.89
Support Vector Machine (SVM)	Polish Dataset	0.8975
RBM (Proposed)	Polish Dataset	0.9614

4. CONCLUSION

In global level bankruptcy is the major issue. Bankruptcy prediction is very important for corporate world and it must be done to reduce financial distress. Models that have been implemented before were only able to give results around 60% which is less for such a big problem. The dataset that has been used in proposed system had a lot of missing data and data that was in that was also unbalanced that is the companies that went bankrupt were very less as compared to compares to the companies which didn't go bankrupt. To deal with missing values Mean imputation technique was used and to deal with unbalanced data issue Synthetic Minority Oversampling Technique (SMOTE) was used. After pre-processing was done model was trained and it gave 70.86% which is higher than all the other models that were created before as they were only able to give results around 60%. Most was given by Support Vector Machine (SVM) at 63.8%. Having a better model and being able to foresee the financial status of a company or organization can help in avoiding global financial problem.

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.

Volume 1, Issue 2 (2019)
Article No. 8
PP 1-6

[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/Linear_regression , Last Accessed on 16th Sept. 2018.

[15] https://en.wikipedia.org/wiki/Random_forest , Last Accessed on 16th Sept. 2018.