INTERNATIONAL RESEARCH JOURNAL OF ENGINEERING & APPLIED SCIENCES
ISSN: 2322-0821(O)
ISSN: 2394-9910(P)
VOLUME 8 ISSUE 2
Apr 2020 - Jun 2020
www.irjeas.org

BANKRUPTCY PREDICTION MODEL USING MACHINE LEARNING TECHNIQUES – A SURVEY

Nidhi Chandel[1], Uday Ganguly[2]

[1,2]*Department of Electronics & Communication, Amity University, NOIDA*

ABSTRACT

Financial depression and then the resultant failure of a business is usually an extremely costly and disrupting event for any company and organization. Statistical predictions of financial depression models try to predict whether a business will experience financial failure in the future. Discriminate analysis and logistic regression have been the most popular approaches which is used, but there is also a large number of data mining and machine learning techniques that can be used for this purposed. In this paper, studied about the machine learning based bankruptcy prediction model which is used Decision tree, Naïve bayes, Neural network and others classification algorithms. These algorithms have been applied to financial distress prediction and compared with each other as well as the most popular approaches in based on various metrics like accuracy, error rate and model building time.

Keywords: Bankruptcy, Bankruptcy prediction system

Classification, machine learning techniques, Credit Card Fraud Detection

1. INTRODUCTION

Techniques for predicting bankruptcy of companies and financial organization became an important issue in the days. Recently in India bankruptcy become a very hot topic in banking, social and political area. The high individual, economic, and social costs inherent in corporate failures or bankruptcies have prompted efforts to provide better insight into and prediction of bankruptcy events [1]. Given the radical change of globalization, more accurate forecasting of corporate financial distress would provide useful information for decision-makers, such as stockholders, creditors, governmental officials, and even the general public. In fact, corporate bankruptcies can be caused by many factors such as wrong investment decisions, a poor investment environment, low cash flow and so on [1]. Therefore, the many current methods for predicting corporate failure must be continuously improved.

The bankruptcy is a typical binary classification problem: there are only two results of prediction, bankruptcy and non bankruptcy. Up to now, many researchers have proposed some classical bankruptcy prediction models based on statistical methods [2] However, the validity of these traditional statistical methods mainly depends on the

Subjective judgments of the human financial experts when

Applied in the selection of some parameters which, in turn,

Inevitably makes feature selection bias. With the development of data mining techniques, machine learning methods have been exploited by many researchers for the bankruptcy prediction problem since these methods can provide an unbiased feature selection and decision making mechanism.

In this paper, different machine learning techniques are employed to predict bankruptcy. The support system can be utilized by stock holders and investors to predict the performance of a company based on the nature of risk associated.

2. LITERATURE REVIEW

In [1] author compare some traditional statistical methods for predicting financial distress to some more ‘‘unconventional’’ methods, such as decision tree classification, neural networks, and evolutionary computation techniques, using data collected from 200 Taiwan Stock Exchange Corporation (TSEC) listed companies. Empirical experiments were conducted using a total of 42 ratios including 33 financial, 8 non-financial and 1 combined macroeconomic index, using principle component analysis (PCA) to extract suitable variables.

Author [2], proposed a semi-parametric Cox survival analysis model and non-parametric CART decision trees have been applied to financial distress prediction and

International Research Journal of Engineering & Applied Sciences, IRJEAS
www.irjeas.org, ISSN(O): 2322-0821, Volume 8 Issue 2, April-June 2020, Page 16-19

compared with each other as well as the most popular approaches. This analysis is done over a variety of cost ratios (Type I Error cost: Type II Error cost) and prediction

Intervals as these differ depending on the situation. The results show that decision trees and survival analysis models have good prediction accuracy that justifies their use and supports further investigation.

The proposed [3] algorithm is successfully applied in the bankruptcy prediction problem, where experiment data sets are originally from the UCI Machine Learning Repository. The simulation results show the superiority of proposed algorithm over the traditional SVM-based methods combined with genetic algorithm (GA) or the particle swarm optimization (PSO) algorithm alone.

In [4] researchers investigate the effect of sampling methods on the performance of quantitative bankruptcy prediction models on real highly imbalanced dataset. Seven sampling methods and five quantitative models are tested on two real highly imbalanced datasets. A comparison of model performance tested on random paired sample set and real imbalanced sample set is also conducted. The experimental results suggest that the proper sampling method in developing prediction models is mainly dependent on the number of bankruptcies in the training sample set. In this research, authors [5] propose the implementation of Jordan Recurrent Neural Networks (JRNN) to classify and predict corporate bankruptcy based on financial ratios. Feedback interconnection in JRNN enables to make the network keep important information well allowing the network to work more effectively. The result analysis showed that JRNN works very well in bankruptcy prediction with average success rate of 81.3785%. Neural Networks can process a tremendous amount of attribute factors; it results in over fitting frequently when more statistics is taken in. By using K-Nearest Neighbor and Random Forest, authors [6] obtain better results from different perspectives. Research [6] testifies the optimal algorithm for bankruptcy calculation by comparing the results of the two methods.

3. PROPOSED WORK

The framework proposed in this work is depicted in Figure 1. The proposed framework for prediction works for each transaction and separates the transaction with high or low risk using the method proposed. The proposed predictive model can be further used to generate alerts for transaction with high risks. Investigators check these alerts and provide a feedback for each alert, i.e. true positive (fraud) or false positive (genuine). The proposed model uses suitable pre-processing, attributes selection techniques along with proposed classification techniques.

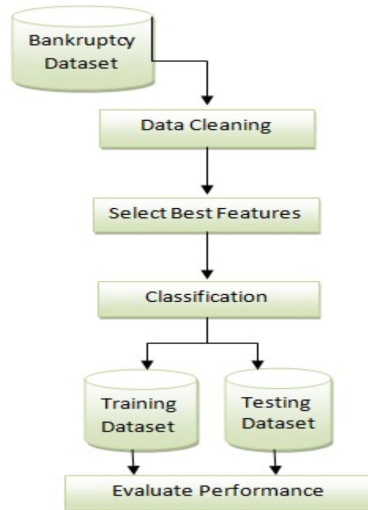

Figure 1: Proposed model

4. EXPERIMENTAL SETUP, METHODOLOGY

3.1 Experimental Setup

Weka 3.6.11 is used as DM tool for simulation purpose. Weka is installed over Windows 7 Operating System. For this research the data used in this study is obtained from sites:https://archive.ics.uci.edu/ml/machine-learningdatabases/00365/ that provided by University of California at Irvine (UCI). The data set consists 1000 Polish companies. 19.4% companies went bankrupt during 2000-2012. Dataset description is presented in Table 1.

Dataset	No. of Features	Total Instances
Bankruptcy (Polish Compines)	65	5910

Table 1: Details of Dataset

3.2 Methodology

The experiment methodology involves following steps:

1. Preprocessing of Dataset
2. Applying Feature selection
3. Applying new ML classifier
4. Evaluate result

4. CONCLUSION & FUTURE WORK

This paper proposed a bankruptcy or financial distress prevention model based on machine learning techniques. Proposed system gives better accuracy and low error rate comparison to existing prediction model. In this survey also studied about the various algorithms and their

www.irjeas.org, ISSN(O): 2322-0821, Volume 8 Issue 2, April-June 2020, Page 16-19

performance metrics. In future this model works on cloud and real time data.

REFERENCES

[1] Mu-Yen Chen "Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches" Elsevier 2011

[2] Adrian Gepp and Kuldeep Kumar∗ "Predicting Financial Distress: A Comparison of Survival Analysis

and Decision Tree Techniques" Elsevier 2015

[3] Yang Lu,1 Nianyin Zeng,2 Xiaohui Liu,3,4 and Shujuan Yi1 "A New Hybrid Algorithm for Bankruptcy Prediction Using Switching Particle Swarm Optimization and Support Vector Machines" Hindawi 2015

[4] Ligang Zhou "Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods" Elsevier 2012

[5] Lingga Hardinata1 , Budi Warsito1 , Suparti1 Bankruptcy prediction based on financial ratios using Jordan Recurrent Neural Networks: a case study in Polish companies IOP Conf. Series: Journal of Physics: Conf. Series 1025 (2018) 012098 doi :10.1088/1742-6596/1025/1/012098

[6] Wenhao Zhang Machine Learning Approaches to Predicting Company Bankruptcy Journal of Financial Risk Management,2017,6,364-http://www.scirp.org/journal/jfrm ISSN Online: 2167-9541 ISSN Print: 2167-9533

[7] Björn mattsson & olof steinert corporate bankruptcy prediction using machine learning techniques department of economics university of gothenburg school of business economics and law,2017

[8] Duaa Alrasheed1 , Dongsheng Che1 Improving Bankruptcy Prediction Using Oversampling and Feature Selection Techniques Int'l Conf. Artificial Intelligence | ICAI'18 |

[9] Jacky C. K. Chow "analysis of financial credit risk using machine learning" Aston University Birmingham, United Kingdom April 2017

[10] 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

[11] M. Krivko, "A hybrid model for plastic card fraud detection systems," Expert Systems with Applications, vol. 37, no. 8, pp. 6070–6076, Aug. 2010.

[12] Benson Edwin Raj, A. Annie Portia, "Analysis on Credit Card Fraud Detection Methods", IEEE International Conference on Computer, Communication and Electrical Technology – ICCCET2011, 978-1-4244-9394-4/11, 2011 IEEE.

[13] David Opitz and Richard Maclin, "Popular Ensemble Methods: An Empirical Study", Journal of artificial intelligence research 169-198, 1999.

[14] L. Breiman, "Bagging predictors," Machine Learning, vol. 24, no. 2, pp. 123–140, 1996.

[15] Freund, Y., & Schapire, R. (1996). Experiments with a new boosting algorithm. In Proceedings of the thirteenth international conference on machine learning, Bari, Italy (pp. 148–156).

[16] Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241–259.

[17] Masoumeh Zareapoor, Pourya Shamsolmolia, "Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier", International Conference on Intelligent Computing, Communication & Convergence, (ICCC 2015), Elsevier, Procedia Computer Science 48 (2015) 679 – 685.

[18]https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4180893/

[19] http://www.cs.waikato.ac.nz/~ml/weka/index.html

[20] Weka, University of Waikato, Hamilton,New Zealand.

[21] V.Mareeswari, Dr G. Gunasekaran, "Prevention of Credit Card Fraud Detection based on HSVM", IEEE, International Conference On Information Communication And Embedded System (ICICES 2016), 978-1-5090-2552-7.

[22] Alejandro Correa Bahnsen, Djamila Aouada, Aleksandar Stojanovic and Bj¨orn Ottersten, "Detecting Credit Card Fraud using Periodic Features", IEEE 14th International Conference on Machine Learning and Applications, 978-1-5090-0287-0/15, 2015 IEEE.

[23] European Central Bank, "Third report on card fraud," European Central Bank, Tech. Rep., 2014.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

International Research Journal of Engineering & Applied Sciences, IRJEAS
www.irjeas.org, ISSN(O): 2322-0821, Volume 8 Issue 2, April-June 2020, Page 16-19

[25] Benson Edwin Raj, A. Annie Portia, "Analysis on Credit Card Fraud Detection Methods", IEEE International Conference on Computer, Communication and Electrical Technology – ICCCET2011, 978-1-4244-9394-4/11, 2011 IEEE.

[26] Alejandro Correa Bahnsen, Aleksandar Stojanovic, Djamila Aouada and Bj¨orn Ottersten, "Cost Sensitive Credit Card Fraud Detection using Bayes Minimum Risk", 12th International Conference on Machine Learning and Applications 2013, 978-0-7695-5144-9/13, 2013 IEEE.

[27] Marwan Fahmi, Abeer Hamdy, Khaled Nagati, "Data Mining Techniques for Credit Card Fraud Detection: Empirical Study", Sustainable Vital Technologies in Engineering & Informatics 2016, Published by Elsevier Ltd.

[28] Wen-Fang YU, Na Wang, "Research on Credit Card Fraud Detection Model Based on Distance Sum", International Joint Conference on Artificial Intelligence 2009, 978-0-7695-3615-6/09, 2009 IEEE.

[29] T. G. Dietterich, "Machine-learning research: four current directions," AI Magazine, vol. 18, no. 4, pp. 97–136, 1997.

[30] R. O. Duda, P. H. Hart, and D. G. Stork, Pattern Classification, Wiley-Interscience, New York, NY, USA, 2000.

[31] R. Bryll, R. Gutierrez-Osuna, and F. Quek, "Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets," Pattern Recognition, vol. 36, no. 6, pp. 1291–1302, 2003.

[32] K. Tumer and N. C. Oza, "Decimated input ensembles for improved generalization," in Proceedings of the International Joint Conference on Neural Networks (IJCNN '99), pp. 3069–3074, Washington, DC, USA, July 1999.

[33] T. Hastie and R. Tibshirani, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2009.

[34] 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