IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 5, May 2018

Credit Risk Assessment in Debt and Equity Securities

Kunal Dwivedi[1], Manish Mata[2], Samir Mulani[3] ,Samiyan Momin[4]

Department of Computer Engineering' Sinhgad Academy of Engineeing, Kondhwa[1,2,3,4]

Abstract: A precise measure of corporate operating performance play critical role for achieving development during turbulent financial markets. This study proposes a reliable and sophisticated prediction architecture that incorporates risk metrics, dimensionality reduction technique, data envelopment analysis, and artificial intelligence technique for corporate operating performance forecasting. The securities market has deviated from its traditional form due to new technologies and changing investment strategies. The current securities market demands scalable machine learning algorithms supporting identification of market manipulation activities. In this paper we use Support Vector Machine algorithms to identify suspicious transactions in relation to market manipulation in stock market. The usage of ensemble classifiers in machine learning plays a vital role in prediction problems. The aim of this study is to analyze the accuracy of the ensemble methods in classifying the customers as good risk group or bad risk group. The opinions are judged on the basis of unsupervised and supervised learning. Supervised learning has unwavering to be superior to unsupervised mode of view verdict. The proposed paper has given a comparative study of naïve bayes and SVM on the opinions of the reviewers of the stock market. No system has been created for sentiment analysis in the share market. One decision in Stock Market can make huge impact on an investor life. The stock market is a complex system and often covered in mystery, it is therefore, very difficult to analyze all the impacting factors before making a decision. In this research we have tried to design a stock market prediction model which is based on different factors. To find best predicting model we can used the Support Vector Machine algorithm.

Keywords: risk management, classification, data mining, market manipulation, Support Vector Machine(SVM), Stock Market, Machine Learning, Feature Selection

I. INTRODUCTION

The global economic crisis that erupted in 2007 turned the spotlight onto corporate and their risk exposures. A fundamental and timely question is, How the risk of a corporate should be measured? This study proposes an emerging architecture to measure corporate risk utilizing the variance of the performance function. By using the variance function we can measure the risk is low, moderate or high. The theory of accounting and corporate finance, the corporate make risky decisions simultaneously with the perception about the expected profits and of the level of other corporate characteristics, mainly debt and capital structure. Support Vector Machine constructs a hyper plane in high dimensional space that maximizes the margin. This method reduces variance, over fitting problem of individual classifiers. Predicting the future has always been an adventurous and attractive task for the probing individuals. This kind of prediction becomes more fascinating when it involves money and risk like predicting the Stock Market The successful prediction or classification of a stock's trend could yield significant profit. In this paper, we will select some factors that are strongly related to the trend of a stock, and try to analyze the relationship between the selected factors and the trend of a stock by using machine learning technique to train a classifier on some observed data. The major reason for the storage losses is due to the use of traditional methods of storage. The go-down managers dispatch the older stock first and then the new stock. So, in our method, we will use the past data of harvesting time, storage conditions and duration of storage vs. storage losses. The corpus of our machine learning algorithm which will predict which stock will degrade first. The Agronomics will get better as the farmers, the go down owners and the end customers will get more quantity of produce thus decreasing the prices for the end customers and also increasing the income of the farmers and the warehouse owners. To illustrate the use of machine learning in the solution, consider a warehouse which stores various stocks of onions. The storage loss depends upon factors such as number of days the stock has been stored, temperature, humidity in the warehouse, and deflection of harvesting time from the ideal harvesting time (in number of days). The system will predict the loss for each stock when the warehouse owner has to dispatch them. All this will be done on just one click. The machine learning algorithm will output the loss in for each stock and then according to the loss, the system will arrange the stocks so that the losses will be minimum.

II. PROPOSED WORK

The work includes building and developing a system that notifies the share holder about the current status of his share so that the losses in different cultivation stages will be prevented.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 5, May 2018

Market data such as price and volume of securities (i.e. the number of shares or contracts that are traded in a security). These methods are based on expert knowledge but suffer from two issues i) detecting abnormal periods that are not associated with known symptoms (i.e. unknown manipulative schemes), ii) adapting to the changing market conditions whilst the amount of transactional data is exponentially increasing (this is due to the rapid increase in the number of investors and listed securities) which makes designing new rules and monitoring the vast data challenging.

In this paper we focus on adopting supervised learning algorithms for detecting market manipulation in stock market, Commodity market, feature bonds. SVM algorithms to build models for predicting transactions that are potentially associated with market manipulation.

In our proposed system, user can select the company name form the company name list. Once user can select the company they can be enter the number of shares. Then system can fetch trained dataset to particular company. In our shares we can apply the machine learning and SVM algorithm. By using SVM algorithm we can classify the dataset in terms of which share has more risk. Identify risk is divided into three group: low, modulate or high. SVM algorithm suggest highly gain company to avoid a loss. Machine learning algorithm storage losses is due to the use of traditional methods of storage.

In the architecture diagram we can show the working of our system:

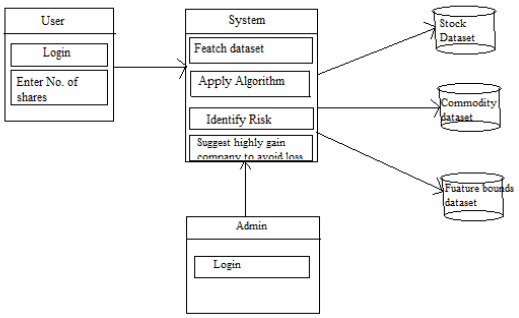

Fig: Architecture diagram

The results compare all four algorithms over training set and test set are show in the Table below. It is evident from the comparison table, that SVM performed best on trained dataset.

Dataset used for verification	**Machine Learning Algorithm**			
	SLP	MLP	RBF	SVM
Trained set	83%	67%	61%	100%
Test set	60%	77%	63%	60%
Average	71.5%	72%	62%	80%

III. METHODOLOGY

A. Support Vector Machine (SVM):

The support vector machine (SVM) is a training algorithm for learning classification and regression rules from data. SVM is used for classification and have recently become an area of intense research owing to developments in the techniques and theory coupled with extensions to regression and density estimation. SVMs are based on the structural risk minimization principle, closely related to regularization theory. This principle incorporates capacity control to prevent over-fitting and thus is a partial solution to the bias-variance trade-off dilemma. Two key elements in the implementation of SVM are the techniques of mathematical programming and kernel functions. The parameters are found by solving a quadratic programming problem with linear equality and inequality constraints; rather than by solving a non-convex, unconstrained optimization problem. SVM can be used for both classification and regression task. In classification case we try to find an optimal hyper plane that separates two classes.

B. Machine Learning:

The major reason for the storage losses is due to the use of traditional methods of storage. the go-down managers dispatch the older stock first and then the new stock. In our method, we will use the past data of harvesting time, storage conditions (temperature, humidity, etc.) and duration of storage vs. storage losses. The input is bounded by number of crops. In this paper, we are focusing on only three crops. We are providing suggestions to the farmers and traders as output. The output is also restricted because of limited number of sensors. The input depends on number and type of sensors installed in the farm. The Agronomics will get better as the farmers, the go down owners and the end

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 5, May 2018

customers will get more quantity of produce thus decreasing the prices for the end customers and also increasing the income of the farmers and the warehouse owners.

IV. LITERATURE SURVEY

In [1] these paper, A precise measure of corporate' operating performance play critical role in it achieving sustainable development during turbulent financial markets, because operating performance is a suitable reflection of corporate management, which has been widely recognized as the main cause of financial troubles. The experimental results show that the proposed architecture can reduce unnecessary information, satisfactorily forecast the corporate operating performance ranking, and yield directions for properly allocating limited financial resource on reliable objects. The introduced architecture is a promising alternative for predicting corporate operating performance ranking, it can assist in both internal and external decision makers.

In [2] these paper, Market manipulation remains the biggest concern of investors in today's securities market, despite fast and strict responses from regulators and exchanges to market participants that pursue such practices. The existing methods in the industry for detecting fraudulent activities in securities market rely heavily on a set of rules based on expert knowledge. The securities market has deviated from its traditional form due to new technologies and changing investment strategies in the past few years. The current securities market demands scalable machine learning algorithms supporting identification of market manipulation activities. In this paper we use supervised learning algorithms to identify suspicious transactions in relation to market manipulation in stock market.

In [3] these paper, Credit scoring prediction is a focus of banking sector to identify trickery customers and to reduce illegal activities. The usage of ensemble classifiers in machine learning plays a vital role in prediction problems. The aim of this study is to analyze the accuracy of the ensemble methods in classifying the customers as good risk group or bad risk group. In this paper experiments are conducted using three ensemble methods namely AdaBoost, Bagging, Random Forest combined with three learning algorithms. Feature selection is applied for selecting important attributes from credit card dataset. This paper provides an assessment on performance of the ensemble classifiers taken for this study.

In [4] these paper, Reduction of post-harvest losses is a critical component of food security. World population is increasing at an alarming rate and thus is the food requirement. Due to limited cultivable land, increasing the food production to meet the needs of people, solely, cannot be the solution. In this paper, we have proposed to build an end-to-end system for farmers and warehouse managers to reduce post-harvest losses. It will consist of a notification-suggestion system which will include data about the current status of farm, suggestions about correct harvesting time and diseases that might affect the crop in its cultivation stages. The system will also include a prediction system for warehouse managers which will suggest the correct dispatch sequence of the stocks and also the optimum temperature and humidity at which one or more crops can be transported so as incur minimum storage and transportation loss. Here, for the prediction-analysis and suggestions, various statistical and probabilistic techniques such as classification and regression are used.

V. CONCLUSION

Corporate operating performance is a good reflection of a corporate management, which has been widely recognized to be the main reason of financial troubles. The paper, explains the gravity of the social issue the world is facing today and also thoroughly explains the practicality of the proposed work. The unsupervised and supervised learning depend methods help to find the results in a better way. The comparative study is done to justify the results.

REFERENCES

[1] D. Diaz, B. Theodoulidis, and P. Sampaio, "Analysis of stock market manipulations using knowledge discovery techniques applied to intraday trade prices," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 12757– 12771, Sep. 2011.
[2] F. Allen and G. Gorton, "Stock Price Manipulation, Market Microstructure and Asymmetric Information," *uropean Econ. Rev.*, vol. 36, pp. 624–630, Oct. 1992.
[3] Rajesh K. Aggarwal and Guojun Wu, "Stock Market Manipulations*," *J. Bus.*, vol. 79, pp. 1915–1953, Dec. 2006.
[4] K. Golmohammadi and O. R. Zaiane, "Data Mining Applications for Fraud Detection in Securities Market," in *2012 European Intelligence and Security Informatics Conference*, 2012, pp. 107–114.
[5] J. Neville, D. Jensen, J. Komoroske, K. Palmer, and H. Goldberg, "Using Relational Knowledge Discovery to Prevent Securities Fraud Categories and Subject Descriptors," in *eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, 2005, pp. 449–458.
[6] D. S. Michael Blume, Christof Weinhardt, "Using network analysis for fraud detection in electronic markets," *Inf. Manag. Mark. Eng.*, vol. 4, no. Studies on EOrganisation and Market Engineering, pp. 101–112, 2006.
[7] M. L. Huang, J. Liang, and Q. V. Nguyen, "A Visualization Approach for Frauds Detection in Financial Market," in *2009 13th International Conference Information Visualisation*, 2009, pp. 197–202.
[8] Z. Ferdousi and A. Maeda, "Unsupervised Outlier Detection in Time Series Data," in *22nd International Conference on Data Engineering Workshops (ICDEW'06)*, 2006, pp. x121–x121.
[9] M. Vlachos, K.-L. Wu, S.-K. Chen, and P. S. Yu, "Correlating burst events on streaming stock market data," *Data Min. Knowl. Discov.*, vol. 16, no. 1, pp. 109–133, Mar. 2007.

[10] H. Öğüt, M. Mete Doğanay, and R. Aktaş, "Detecting stock-price manipulation in an emerging market: The case of Turkey," *Expert Syst. Appl.*, vol. 36, no. 9, pp. 11944–11949, Nov. 2009.
[11] R. Aktaş and Doğanay M., "Stock-price manipulation in the Istanbul stock exchange," *Eurasian Rev. Econ. Financ.*, vol. 2, pp. 21–28, 2006. 11 http://www.r-project.org 12 www.westgrid.ca
[12] L. Breiman, J. Friedman, C. Stone, and R. Olshen, "Classification and regression trees," 1984.
[13] J. Quinlan, *C4. 5: programs for machine learning*. 1993.
[14] W. Loh and Y. Shih, "Split selection methods for classification trees," *Stat. Sin.*, 1997.
[15] C. L. Chih-wei Hsu, Chih-chung Chang, "A practical guide to support vector classification," 2010.
[16] J. D. Kirkland, T. E. Senator, J. J. Hayden, T. Dybala, H. G. Goldberg, and P. Shyr, "The NASD Regulation Advanced-Detection System (ADS)," *AI Magazine*, vol. 20, no. 1. p. 55, 15-Mar-1999.
[17] T. E. Senator, "Ongoing management and application of discovered knowledge in a large regulatory organization," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '00*, 2000, pp. 44–53.
[18] A. Zeileis, T. Hothorn, and K. Hornik, "Model-Based Recursive Partitioning," *J. Comput. Graph. Stat.*, vol. 17, no. 2, pp. 492–514, Jun. 2008.
[19] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5– 32, Oct. 2001.
[20] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee, "Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods," *Ann. Stat.*, vol. 26, no. 5, pp. 1651–1686, Oct. 1998.
[21] S. L. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Mach. Learn.*, vol. 16, no. 3, pp. 235–240, Sep. 1994.