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], Kiran Avhad[2], Manish Mata[3], Samir Mulani[4], Samiyan Momin[5]

Department of Computer Engineering, Sinhgad Academy of Engineering, Kondhwa[1,2,3,4,5]

Abstract: Credit Risk Assessment is a crucial part of decision making in any financial institutes. The aim of this paper is to highlight and illustrate the use of some quantitative techniques for risk estimation in finance and insurance. We will study the theoretical properties, the accuracy of modelling the economic phenomena and the computational performances of the risk measures Value-at-Risk, Conditional Tail Expectation, Conditional Value-at-Risk and Limited Value-at-Risk in the case of logistic distribution. We also investigate the most important statistical estimation methods for risk measure evaluation and we will compare their theoretical and empirical behavior.

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

I. INTRODUCTION

Credit risk refers to the risk that a borrower may not repay a loan and that the lender may lose the principal of the loan or the interest associated with it. Credit risk arises because borrowers expect to use future cash flows to pay current debts; it's almost never possible to ensure that borrowers will definitely have the funds to repay their debts. Interest payments from the borrower or issuer of a debt obligation are a lender's or investor's reward for assuming credit risk. [3]. Risk prediction depends on assessing the current holding in various financial instruments. These financial instruments include stocks, commercial bonds, future options, commodities etc.[1]. Although the economic definitions of these financial instruments are different, the way these instruments are traded remains the same. So it is believed if an adequate prediction model is created for any one of the financial instrument then the same model can be used for the future value of other instruments. [4] The basic idea here is that once the future value of the holding is known then the potential risk (exposure) can be calculated. The most amount of research has been done on predicting the stock prices, thus for research purpose the "Stock Market prediction Models were searched". Over the past three decades many important changes have taken place in the environment of financial markets. The development of powerful communication and trading facilities has enlarged the scope of selection.[2] The objective of the paper is to describe the previous research that has been done on the topic. The Risk assessment can be divided into two parts: Risk Production / Calculating Exposure and Suggesting Possible Actions to Minimize the Exposure.

II. PROPOSED WORK

The work includes building and developing a system that predicts the risks of the users current holding in the portfolio and gives the user an estimated value of his forthcoming loss so that he can strategies accordingly. 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, future bonds. SVM algorithms to build models for predicting transactions that are potentially associated with market manipulation.

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

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

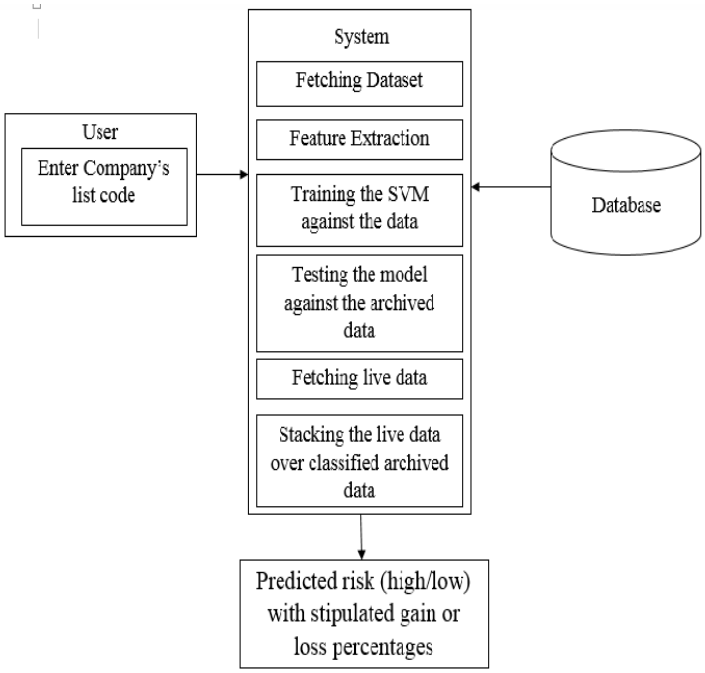

Fig [1]. Architecture of the System.

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%

Fig [2]. Comparison of all algorithms

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.

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

B. Machine Learning:

Machine leaning is an important tool for building predictive models, especially the Supervised Learning algorithms. Here a training data set is provided containing historical trade patterns for a specified period of time (6-12 months). Now the thing that makes the machine learning techniques special is the fact that the algorithm (SVM or decision trees) alter their approach to best fit the desired result. They do this based on the experience they gain from going through the training data set, i.e. the machine "learn" how to predict the future stock prices and compare them with the current values to get an estimate of the financial risk posed by the stock in the coming future.

IV. LITERATURE SURVEY

[1] paper describes 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.

[2] Paper describes 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.

[3] Paper describes credit scoring prediction is a focus of banking sector to identify trickery customers and to reduce illegal activities. The usage of ensemble classifiers in machinelearning 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.

[4] Paper describes 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. ALGORITHM

Following are the steps:

i. Enter Company's List code.
ii. Fetch Dataset.
Iii. Extract features from the dataset.
iv. Train the SVM model against the fetched data.
v. Test the model against the data that is archived.
vi. Fetch live data according to user's choice.
vii. Stack the live data over the archived data.
viii. Predict the risk with stipulated loss or gain percentages.

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

VI. RESULT

1. Training the data.

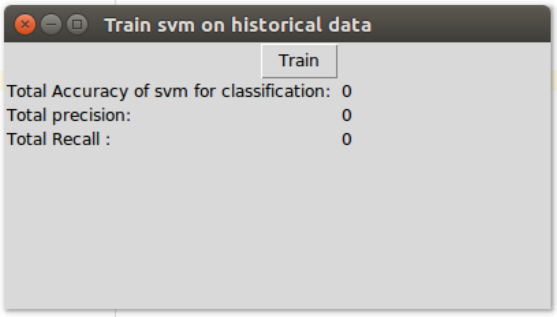

Fig [4] Training Dataset First Snippet

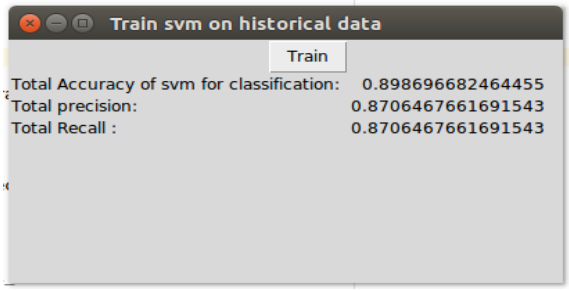

Fig [5] Training Dataset Second Snippet

2. Risk Prediction

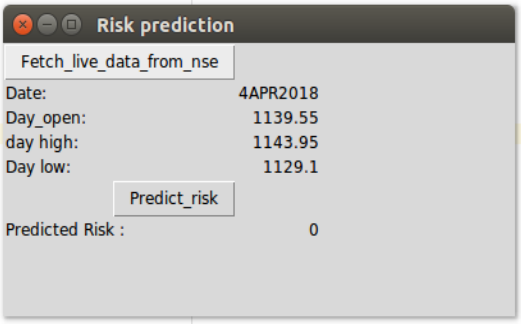

Fig [6] Risk Prediction First Snippet

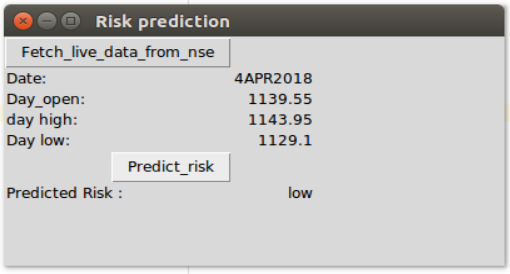

Fig [7] Risk Prediction Second Snippet

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

VII. CONCLUSION

Risk Assessment is a very important part of our financial system and thus there has been a lot of research on this topic. The main objective of this research is to find a better way of getting the job done as with each passing day, the volume of trade is increasing and with this the amount of data is also increasing. And with this plethora of data one is prone to make mistakes. Thus predictive model and decision trees and learning algorithms play an important role in obtaining the solutions.

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", 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.
[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, andP. 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," inProceedings 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.

BIOGRAPHIES

Kunal Dwivedi, B.E (Computer Engineering), Sinhgad Academy of Engineering, Kondhwa

Prof. K. A. Avhad, Project Guide, Sinhgad Academy of Engineering, Kondhwa, Kondhwa

Samir Mulani, B.E (Computer Engineering), Sinhgad Academy of Engineering, Kondhwa

Manish Mata, B.E (Computer Engineering), Sinhgad Academy of Engineering, Kondhwa

Samiyan Momin, B.E (Computer Engineering), Sinhgad Academy of Engineering, Kondhwa