ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

A Survey on Data Mining in the Financial Sector and Stock Market

Vibhor Chandra Gupta[1], Varun Chandra Gupta[2]

Bachelors of Technology Student, Department of Computer Engineering, Mukesh Patel School of Technology Management and Engineering, NMIMS University, Mumbai, Maharashtra, India[1]

Bachelors of Technology (Integrated) Student, Department of Computer Engineering, Mukesh Patel School of Technology Management and Engineering, NMIMS University, Mumbai, Maharashtra, India[2]

ABSTRACT: Big Data is constantly being generated by various sources such as the internet, mobile devices, personal computers and various other electronic devices. This data can be very useful in the prediction of various fields of the financial sector such as the quality of a loan, what are the best interest rates etc, and the stock market where various transaction data and external information can be analysed to predict future values of the stock. This paper endeavours to determine the feasibility and practicality of applying various data mining techniques and analysis as a forecasting tool for the individual investor and the company at large.

KEYWORDS: Co-relation Analysis, Data Clustering, Data Mining, Neural Networks, Stock Market, Stock selection

I.INTRODUCTION

The stock market refers to the collection of markets and exchanges where the issuing and trading of equities or stocks of publicly held companies take place. There is a need for a system that enables individuals to make a propitious investment decision. This can be done by studying and analysing, various metrics to gauge the performance and characteristics of the stock behaviour. A model can then be designed to increase return on investments and reduce investment risks. There are various approaches which will be analysed and compared in the following sections, such as using clustering methods to classify and predict the trend of stocks, using lagged correlation method of predicting stocks with strong correlation which are suited for short term capital gains, using technical and fundamental analysis, which refers to the analysis of companies by their financial value, potential growth which is found by business trends, general economic conditions, etc. Finally, we will be looking into Stock data analysis based on BP neural network.
The focus of this paper is to determine an efficient technique to select stocks in an inferential manner to maximize profit returns with minimum risks.

II. RELATED WORK

The papers on which the survey was done primarily focused on 3 different types of approaches. The first one is, stock selection based on correlation analysis. Correlation analysis measures the inter-relationship between the stock prices of two companies. In the second approach, stock selection was done based on data clustering method. The K-means algorithm was used to create a set of k clusters of the stocks. Five financial indicators were chosen for clustering purposes. The third paper discusses about how stock selection can be done using Stock data analysis based on BP neural network(Back Propagation). Here, neural networks were used, and various technical indicators in economics were used asinput for the BP neural network.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

III.DATA MINING

Big Data

Big data is an evolving term that describes any voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information. The use of big data management and processing in the financial sector were discussed since the 1980s, but at the time due to lack of attention given to it, it was temporarily put on hold.

In forthcoming years with the development of the internet, mobile computers and other technological advances, big data became readily available in massive quantities, the great potential use of big data was finally recognized and strides were made by various companies and governments to build a system to deal with the storage and processing of big data.

Data Analysis

Data analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data. According to Shamoo and Resnik (2003) various analytic procedures "provide a way of drawing inductive inferences from data and distinguishing the signal (the phenomenon of interest) from the noise (Statistical fluctuations) present in data.

IV. FINANCIAL SECTOR

Nowadays, there are a plethora of financial products offered by various financial institutions such as banks, insurance companies, brokerage firms etc, along with this, the presence of network terminals in large amounts allow for the creation of huge complex financial data each and every day. Adding to this, there are integrated data platforms tailored around an individual person or group of people but large companies like social networks. This data too can be used to generate financial data.

Use of this data is the key to success in the highly competitive financial sector where the decision makers can make use of data mining and extra useful information from the highly complex data and then make scientific, reasonable, logical and feasible plans and strategies for the future.

Financial Decisions Based in the Data Mining Model Design

The data mining model design selects samples or a subset from the data warehouse, analyses them using a particular algorithm and generates a mathematical model. In a system, all models will be stored in the model base and will be managed by the model base management system.

In Financial decision system model database, the models are generally divided into two categories: (1) unit model and (2) combination model. The unit model is further divided into the financial cost decision-making model, financial scale forecast model, capital structure analysis model financial risk analysis model, etc. Combination mode is a combination of multiple unit models, used in process of policy making for complex questions.

A complete model includes: model head, model name, input parameters, output parameters, modelling time, model function, model connection situation, model main body, model structure and so on.

The process of model designing is an iterative one. A good model takes feedback based on the result and if the results are unsatisfactory, the model is modified till we get satisfactory results.

Applications

- To construct data mining design of multidimensional data analysis and to construct a data warehouse.
- To analyse client status.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

- Enterprise risk management. Financial risk prevention and financial fraud identification have been the important aspects that the financial sector needs to focus on and deal with appropriately. Data mining techniques can be used to analyse and find the risk information concerning the industry development and the user's personal credit information, forecast risk level, and provide reference data for decision makers to avoid risks, from a large number of financial business data and customer transaction information.
- To forecast market trends.
- To provide support for the detection of economic crime.
- Credit Risk Assessment - Credit assessment process can use data such as Gender, Age, Credit amount, Monthly income, expenses each month, Current payment per month, Saving (income expenses payment), Collateral types, Collateral values, Loan period, Type of business activities, Source of funding, Previous credit status/rating.

V.STOCK MARKET

Prediction system of stock market is very crucial and essentially important because it deals with the huge amount of money and in today's growing and forward time money is first priority. The predicted value directly affects the stock price and no one take risk to drop down in stock market index. So,the due to money involvement and the reputation of the shares stock market needs to be a perfect or more accurate prediction about their upcoming market trends.

Methodology of Prediction

When performing an experiment, it is necessary to have a dataset and a proper methodology as to how to work on that dataset so that a proper prediction could be made in lieu of future decisions to be made. Before using any model with the data set it is imperative that we ensure that it's pre-processed: that is the data should be in the .csv format without any noise or null values.

The next step is to divide the dataset into the training data and the testing data. Classification models are then executed on the data subsets after which the results are generated in the form of evaluation parameters such as Gini, F-measure, AUC, H, TPR, Sensitivity, FPR, Specificity, Recall, Error Rate, Accuracy, Precision and Time. Any of these parameters can be used to compare the results.

Final results help the experimenter to predict stock markets such that correct decisions can be made. The figure below depicts the major processes that are to be applied on the data, and the respective flow of operations.

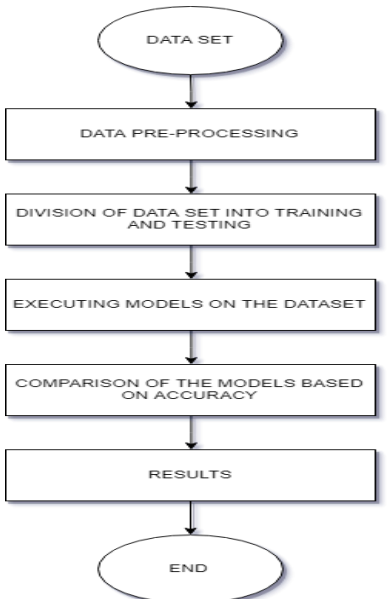

Fig. 1. Steps to be executed on Data

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

Association Rules for Stock Trading

Association rule mining is a useful technique for discovering correlations among items. In various experiments, three transactional datasets were generated using different values of change percentage: 1%, 2% and 3%. After three transactional datasets were constructed, an association rule mining technique was applied to find the association rules separately for each dataset.

Since there are three different sets of association rules according to a certain percentage of change to a stock price, in reality, traders shall select an interesting rule from a group of acceptable price change. As an example, if an investor aims to buy or sell a stock when the price change is more than equal to 2%, the rule should be from the 2% group.

VI. ANALYSIS

Stock selection based on Correlation Analysis

Correlation analysis measures the inter-relationship between the stock prices of two companies. The software tool - Stock Strategy Analyser (SSA) is used for analysing and gauging the dataset. The stock price of company A is used as the dependent variable and the stock prices of all other listed companies as the independent variables to calculate correlation coefficients using SSA. SSA allows the user to forward shift periods (lag). The independent variable can be forwarded (i.e. 10 days) to see if there is any predictive power in the indicator.

The correlation coefficients are calculated based on an index of price movements rather than the actual price of the stock. This allows comparison of high priced and low-priced stocks possible. SSA identifies the companies with the highest degree of correlation and sorts them. Real world company data is provided and the accuracy of the prediction is evaluated. The model correctly predicted the downward price trend thus providing an opportunity for sell signals. It also predicting correct upward movements thus making short term capital gains.

Stock selection based on data clustering method

The K-means algorithm is used to create a set of k clusters of the stocks.We arbitrarily choose K objects from the dataset as the initial cluster centers. Each object is assigned to the cluster to which the object is the most similar to, based on the mean value of the objects in the cluster. The cluster means are then updated, i.e. calculating the mean value of the objects for each cluster. This process is repeated until there is no more change.

5 financial indicators are chosen:

- Primary earnings per share
- Net asset value per share
- Total assets turnover ratio
- Principle business growth rate
- Liquidity ratio

Firstly, 7 categories are made by using combinations of 2 parameters. Then,3 leading stocks with great weight and stable performance from the first seven categories are chosen, which join the option stock making the eighth category.The obtained stock after the selection with the cluster analysis method returns a greater profit as compared to the stocks without the cluster analysis.
The cluster analysis method thus gave us a stock investment decision,which improved the success rate and yield.

Stock data analysis based on BP neural network

1. Neural Network –

Artificial neural network is a large broad network with a number of processing units (neurons/nodes) connected. It is an abstract, simplified and simulation to human brain, and reflects the basic characteristics of the human brain using this form to solve a variety of problems. Generally, the neural network is the multi-layered network topology, including the input layer, multiple hidden layers and output layer. The Hopfield network and BP network are most commonly used.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

Hopfield network is the most typical feedback network model, it is one of the models which are most commonly studied now. The Hopfield network is the monolayer constituted by the same neuron, and is also a symmetrically connected associative network without learning function. It can implement the restriction optimization and associative memory.

BP network is the back-propagation network. It is a multi-layer forward network, learning by minimum mean square error. It is one of the most widely used networks. It can be used in the field of language integration, identification and adaptive control, etc. BP network is semi supervised learning.

2. Technical indicators in economics as input of BP neural network –

Moving average (MA) Moving average line is a statistical mean, which sums up stock price of certain days and gives out an average, and then connects them into a line to observe the price trend. The function of moving average is obtaining the average cost during a certain period, and we can use the average cost curve and the movement of daily closing price changes in the line analysis of the change of bull and bear to study and to determine possible changes in stock.

Random indicator (KDJ) There are a total of three lines standing for random indicators in the stock,namely K line, D line and the J line. Random indicator not only considers the highest price,the lowest price in the calculation period, but also takes into account of the random amplitude in the course of the fluctuation of stock price. Therefore, researchers always think that random indicator can more truly reflect the volatility of stock price,and it plays an important role in prompting.

Moving Average Convergence/Divergence (MACD) The principle of MACD is to use the functions of the signs for aggregation and separation of fast-moving average and slow-moving average, in addition to double smoothing operation in order to study and determine the timing of buy and sell.

Relative Strength Index (RS1) Relative Strength Index is an indicator comparing the average of closing high and the average of closing low. It can be used to analyse the market and its strength in order to forecast the future of the markets current trend.

On Balance Volume (OBV) OBV is the degree of the active investors in the stock market. If there are a lot of buyers and sellers, the stock prices and the volume of trade will rise, the atmosphere of the stock market is warm, commonly known as the bull market; if there are a few of buyers and sellers, the stock prices and the volume of trade will decline. It can be seen that the impact of the rise and fall of stock prices and the volume of trade is the OBV of stock. The volume of shares and stock price can also reflect the degree of the rise and fall of popularity of stock.

BIAS is the ratio between the application index and the moving average. Base on BIAS,we can observe the degree which the stock price deviate from the moving average price to decide to buy or sell.

Increase scope= (the stock market closing price of today vs the stock market opening price of today) / the stock market opening price of today. in order to avoid the situation that the dimensions of larger data have larger influence in the results than that of smaller data, we normalized the original data.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

VII.CONCLUSION

Findings

The financial sector and the stock market have changed a lot in recent years and have adopted new methods for prediction of their various fields including many forms of data mining which has proven highly beneficial.

Results

Correlation analysis measures the inter-relationship between the stock. A BP neural network model that is used gives results with an error of less than 1% proving to be an effective form of prediction along with its great customizability.

REFERENCES

[1]H. Zhang, Y. Li, C. Shen, H. Sun and Y. Yang, "The Application of Data Mining In Finance Industry Based on Big Data Background," 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems, New York, NY, 2015, pp. 1536-1539.
[2]I. Gusti Ngurah Narindra Mandala, Catharina Badra Nawangpalupi, Fransiscus Rian Praktikto, Assessing Credit Risk: An Application of Data Mining in a Rural Bank, Procedia Economics and Finance, Volume 4, 2012, Pages 406-412,
[3]J. Zhang and F. Shao, "Stock Data Analysis Based on BP Neural Network," 2009 Third International Symposium on Intelligent Information Technology Application Workshops, Nanchang, 2009, pp. 288-291.
[4]L. Zhao and L. Wang, "Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm," 2015 IEEE Fifth International Conference on Big Data and Cloud Computing, Dalian, 2015, pp. 93-98.
[5]A. Srisawat, "An application of association rule mining based on stock market," The 3rd International Conference on Data Mining and Intelligent Information Technology Applications, Macao, 2011, pp. 259-262.
[6]S. Wang, "Design and implementation of enterprise financing decision model based on data mining," MSIE 2011, Harbin, 2011, pp. 1049-1052.
[7]P. Kumar and A. Bala, "Intelligent stock data prediction using predictive data mining techniques," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, 2016, pp. 1-5.
[8]Carol Hargreaves ; Yi Hao, "Does the use of technical & fundamental analysis improve stock choice? : A data mining approach applied to the Australian stock market," 2012 International Conference on Statistics in Science, Business and Engineering, 2012, pp. 1 - 6.
[9] Cicil Fonseka ; Liwan Liyanage "A Data mining algorithm to analyse stock market data using lagged correlation," 2008 4th International Conference on Information and Automation for Sustainability 2008 pp. 163 - 166.
[10] Ruizhong Wang "Stock Selection Based on Data Clustering Method," 2011 Seventh International Conference on Computational Intelligence and Security 2011 pp. 1542 - 1545.
[11] Li Sun, Kaile Zhou, Xiaoling Zhang, Shanlin Yang, "Outlier Data Treatment Methods Toward Smart Grid Applications", Access IEEE, vol. 6, pp. 39849-39859, 2018.
[12] Zhao, L., & Wang, L. (2015). Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm. 2015 IEEE Fifth International Conference on Big Data and Cloud Computing.