PREDICTING STOCK MARKET TRENDS USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

1.DR. N. JAGADEESHAN, 2. T. NAGA TEJASWI, 3. VATSAVAI SYAMALA GAYATHRI,4. Y. LIKITHA SREE
1.PROFESSOR,2,3&4UG SCHOLAR
DEPARTMENT OF ECE, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN, HYDERABAD

ABSTRACT Stock prices prediction is interesting and challenging research topic. Developed countries' economies are measured according to their power economy. Currently, stock markets are considered to be an illustrious trading field because in many cases it gives easy profits with low risk rate of return. Stock market with its huge and dynamic information sources is considered as a suitable environment for data mining and business researchers. In this paper, we applied k-nearest neighbor algorithm and non-linear regression approach in order to predict stock prices for a sample of six major companies listed on the Jordanian stock exchange to assist investors, management, decision makers, and users in making correct and informed investments decisions. According to the results, the kNN algorithm is robust with small error ratio; consequently the results were rational and also reasonable. In addition, depending on the actual stock prices data; the prediction results were close and almost parallel to actual stock prices

1. INTRODUCTION Recent business research interests concentrated on areas of future predictions of stock prices movements which make it challenging and demanding. Researchers, business communities, and interested users who assume that future occurrence depends on present and past data, are keen to identify the stock price prediction of movements in stock markets (Kim, 2003). However, financial data is considered as complex data to forecast and or predict. Predicting market prices are seen as problematical, and as explained in the efficient market hypotheses (EMH) that was put forward by Fama (1990). The EMH is considered as bridging the gap between financial information and the financial market; it also affirms that the fluctuations in prices are only a result of newly available information; and that all available information reflected in market prices. The EMH assert that stocks are at all times in equilibrium and are difficult for inventors to speculate. Furthermore, it has been affirmed that stock prices do not pursue a random walk and stock prediction needs more evidence (Gallagher and Taylor, 2002; Walczack, 2001; Kavusssanos and Dockery, 2001; Lakonishok et.al, 1994; O'Connor et. al., 1997; Lo and MacKinlay, 1997; Kirt and Malaikah, 1992; Lo and

MacKinlay, 1988). Moreover, various studies were performed to determine stock price predictions (Subha and Nambi, 2012; Qian and Rasheed, 2007; Fama and French, 1992; Cochrane, 1988; Campbell, 1987; Chen, et al. 1986; Basu, 1977). In addition to purchasing and selling stocks and shares in stock markets, each stock is not only characterized by its price, but also by other variables such as closing price which represents the most important variable for predicting next day price for a specific stock. There is a relationship and specific behavior exists between all variables that effect stock movements overtime. Different economic factors, such as political stability, and other unforeseeable circumstances are variables that have been considered for stock price predictions (Ou, P. and Wang, H., 2009; Fama and French, 1993; Cochrane, 1988; Campel, 1987; Chen. et.al.1986). Table 1 summarizes the main variables that affect stock movements used in this article. Data mining technology is used in analyzing large volume of business and financial data, and it is applied in order to determine stock movements. Mining temporal stock markets is required to provide additional capabilities required in cases where the existing data and their interactions need to be observed through time dimension. In stock predictions, a set of pure technical data, fundamental data, and derived data are used in prediction of future values of stocks. The pure technical data is based on previous stock data while the fundamental data represents the companies' activity and the situation of market. Combining data mining classification approaches in stock prediction yields a future value for each unknown entities of companies' stocks values based on historical data. This prediction uses various methods of classification approaches such as neural networks, regression, genetic algorithm, decision tree induction, and k-Nearest Neighbors (kNN). In classification approaches, a data set is divided into training data set and testing set. kNN uses similarity metrics to compare a given test entity with the training data set. Each data entity represents a record with n features. In order to predict a class label for unknown record, kNN selects k recodes of training data set that are closest to the unknown records.

2. LITERATURE REVIEW Financial services companies are developing their products to serve future prediction. There are a large amount of financial information sources in the world that can be valuable research areas, one of these areas is stock prediction and also called stock market mining. Stock prediction becomes increasingly important especially if number of rules could be created to help making better investment decisions in different stock markets. The genetic algorithm had been adopted by Shin et al. (2005); the number of trading rules was generated for Korea Stock Price Index 200 (KOSPI 200), in Sweden Hellestrom and Homlstrom (1998) used a statistical analysis based on a modified kNN to determine where correlated areas fall in the input space to improve the performance of prediction for the period 1987-1996. Both models mentioned were provided in the Zimbabwe stock exchange to predict the stock prices which included Weightless Neural Network (WNN) model and single exponential smoothing (SES) model Mpofu (2004). Clustering stocks approach was provided by Gavrilov et al. (2004) to group 500 stocks from the Standard & Poor. The data represented a series of 252 numbers including the opening stock price. A fuzzy genetic algorithm was presented by Cao (1977) to discover pair relationship in stock data based on user preferences. The study developed potential guidelines to mine pairs of stocks, stock-trading rules, and markets; it also showed that such approach is useful for real trading. Moreover, other studies adopted kNN as prediction techniques such as (Subha et al., 2012; Liao et al. 2010; Tsai and Hsiao 2010; Qian and Rasheed, 2007)

3. RESEARCH METHODOLOGY AND ANALYSIS The kNN algorithm method is used on the stock data. Also, mathematical calculations and visualization models are provided and discussed below.

3.1 K-NEAREST NEIGHBOR CLASSIFIER (KNN) K-nearest neighbor technique is a machine learning algorithm that is considered as simple to implement (Aha et al. 1991). The stock prediction problem can be mapped into a similarity based classification. The historical stock data and the test data is mapped into a set of vectors. Each vector represents N dimension for each stock features. Then, a similarity metric such as Euclidean distance is computed to take a decision. In this section, a description of kNN is provided. kNN is considered a lazy learning that does not build a model or function previously, but yields the closest k records of the training data set that have the highest similarity to the test (i.e. query record). Then, a majority vote is performed among the selected k records to determine the class label and then assigned it to the query record. The prediction of stock market closing price is computed using kNN as follows:

a) Determine the number of nearest neighbors, k.

b) Compute the distance between the training samples and the query record.

c) Sort all training records according to the distance values.

d) Use a majority vote for the class labels of k nearest neighbors, and assign it as a prediction value of the query record.

3.2 MATHEMATICAL CALCULATIONS AND VISUALIZATIONS MODELS This represents an overview of equations that were applied in this article for predicting next day price. The calculations includes error estimation, total sum of squared error, average error, cumulative closing price when sorted using predicted values, k-values and training Root Mean Square (RMS) errors.

a) Root Mean Square Deviation (RMSD) is accuracy estimated values, Y, and the actual values, X. Th measure. RMSD = SQRT$(Y-X)^2$.

b) Explained Sum of Squares (ESS) is computed as follc

$$\text{ESS} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 .$$

Where y_i is the predicted variable, and y *is* the actual v

c) Average Estimated Error (AEE)

AEE is the total sum of RMS errors for all variables records.

$$\text{AEE} = \frac{\Sigma_{i=1}^{n} RMSi}{n}$$

3.3 VISUALIZATION GRAPH To evaluate the performance of kNN learning model, lift graph is applied and drawn for different companies' stock values. The lift chart symbolizes the enhancement that a data mining model offers when distinguished against a random estimation, and the change is expressed in terms of lift score. Through contrasting the lift scores for a variety of parts of the data set and for different models, it can then be decided which model is supreme and which percentage of the cases within the data set would gain from employing the predictions model. Furthermore, using the lift chart assist in distinguishing how accurate predictions are for various models with identical predictable characteristic. The lift graph also shows the ratio between the results obtained using the predictive model or not. The other graph applied is the plot curves to show the relation between the actual and predicted stock price.

4. DATA DESCRIPTION, RESULTS, AND ANALYSIS In this article, data from the Jordanian stock exchange was analyzed and a brief data analysis is presented to provide the reader with the fundamental concepts of data attributes. Also, the obtained results of prediction of the Jordanian stock exchange are provided.

4.1 Data Description The sample data was extracted from the Jordanian stock exchange. The study sample included stock data of five randomly selected companies listed on the Jordanian stock exchange as a sample training dataset from the period June 4, 2009 to December 24, 2009 as shown in table 2. Each of these companies has approximately 200 records with three attributes including closing price, low price, and high price as shown in table 3. A brief data analysis is presented with the fundamental concepts of data attributes. The attributes for each company are included in the data analysis. Closing price is the main factor that affects the prediction process for a specific stock based on kNN algorithm. The kNN algorithm is applied on a 1000 records to estimate predicted values for each stock.

4.2 Analysis And Results The results of the predicted stock price for each individual company used in the sample with graphs for the actual and predicted prices are presented. The results as seen in tables 4.1 to 4.5 and in figures 1 to 11 are those after applying kNN algorithm for each company's closing prices with the residual values which indicates how far away is the predicted values from the actual values; the negative residual value indicates that the predicted value is larger than the actual one. Section 4.3 summarizes the five companies' prediction performance. Tables 4.1 to 4.5 respectively represent the results after applying kNN algorithm of the Arab international for education and investment (AIEI), Jordan steel company (JOST), Arab financial investment (AFIN), Irbid district electricity (IREL), and the Arab potash company (APOT). As depicted in the figures (1-10) below, the line chart of the actual and predicted values for the companies in the sample and after adopting the kNN prediction model, the results show that the predictive value and the actual value were moving in similar manner as seen in figure 2,4,6,8, and 10. Moreover, the lift chart also applied to evaluate the performance of kNN learning model used and proved that the model used is performing well; this can also be seen in figures 1,3,5,7, and 9 representing the lift charts (company's dataset).

2. CONCLUSION

The aim of this research is to improve the statistical fitness of the proposed model to overcome a KNN problem due to its computation approach. The KNN classifier can compute the empirical distribution over the Profit and Loss class values in the k number of nearest neighbors. However, the outcome is less than adequate due to sparse data. The KNN classifier has under fitting issue as it does not cater to generalization of sparse data outside the range of nearest neighborhood.

We have compared a hybrid KNN-Probabilistic model with four standard algorithms on the problem of predicting the stock price trends. Our results showed that the proposed KNN-Probabilistic model leads to significantly better results compared to the standard KNN algorithm and the other classification algorithms.

The limitation of the proposed model is that it applies a binary classification technique. The actual output of this binary classification model is a prediction score in two- class. The score indicates the model's certainty that the given observation belongs to either the Profit class or Loss class. For future work, the knowledge component is to transform the binary classification into multiclass classification. The multiclass classification involves observation and analysis of more than the existing two statistical class values. Additional research will include the application of the probabilistic model to multiclass data in order to provide more specific information of each class value. The newly formed multiclass classification will contain five class labels named “Sell”, “Underperform”, “Hold”, “Outperform”, and “Buy”. In numerical values for mapping purpose, we will convert “Sell” to -2 which implies strongly unfavorable; “Underperform” to -1 which implies moderately unfavorable; “Hold” to 0 which implies neutral; “Outperform” to 1 which implies moderately favorable; and “Buy” to 2 which implies strongly favorable.

REFERENCES

[1] Benjamin Graham, Jason Zweig, and Warren E. Buffett, The Intelligent Investor, Publisher: Harper Collins Publishers Inc, 2003.

[2] Charles D. Kirkpatrick II and Julie R. Dahlquist, Technical Analysis: The Complete Resource for Financial Market Technicians (3rd Edition), Pearson Education, Inc., 2015.

[3] Bruce Vanstone and Clarence Tan, A Survey of the Application of Soft Computing to Investment and Financial Trading, Proceedings of the Australian and New Zealand Intelligent Information Systems Conference, Vol. 1, Issue 1, http://epublications.bond.edu.au/infotech_pubs/ 13/, 2003, pp. 211–216.

[4] Monica Tirea and Viorel Negru, Intelligent Stock Market Analysis System - A Fundamental and Macro-economical Analysis Approach, IEEE, 2014.

[5] Kian-Ping Lim, Chee-Wooi Hooy, Kwok-Boon Chang, and Robert Brooks, Foreign investors and stock price efficiency: Thresholds, underlying channels and investor heterogeneity, The North American Journal of Economics and Finance. Vol. 36, http://linkinghub.elsevier.com/retrieve/pii/ S106 2940815001230, 2016, pp. 1–28.

[6] Lamartine Almeida Teixeira and Adriano Lorena Inácio de Oliveira, A method for automatic stock trading combining technical analysis and nearest neighbor classification, Expert Systems with Applications, http://linkinghub.elsevier.com/retrieve/pii/

S095 7417410002149, 2010, pp. 6885–6890.

[7] Banshidhar Majhi, Hasan Shalabi, and Mowafak Fathi, FLANN Based Forecasting of S&P 500 Index. Information Technology Journal, Vol. 4, Issue 3, http://www.scialert.net/abstract/?doi=itj.2005.2 89.292, 2005, pp. 289–292.

[8] Ritanjali Majhi, G. Panda, and G. Sahoo, Development and performance evaluation of FLANN based model for forecasting of stock markets, Expert Systems with Applications, Vol. 36, Issue 3, http://linkinghub.elsevier.com/retrieve/pii/S095 7417408005526, 2009, pp. 6800–6808.

[9] Tong-Seng Quah and Bobby Srinivasan, Improving returns on stock investment through neural network selection, Expert Systems with Applications, Vol. 17, Issue 4, http://linkinghub.elsevier.com/retrieve/pii/S095 741749900041X, 1999, pp. 295–301.

[10] Halbert White, Economic prediction using neural networks: the case of IBM daily stock returns, IEEE International Conference on Neural Networks, http://ieeexplore.ieee.org/document/23959/, IEEE, 1988, pp. 451–458.

[11] Yauheniya Shynkevicha, T.M. McGinnity, Sonya A. Coleman, and Ammar Belatreche, Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning, Decision Support Systems, Vol. 85, http://linkinghub.elsevier.com/retrieve/pii/S016 7923616300252, 2016, pp. 74–83.

[12] Han Lock Siew and Md Jan Nordin, Regression Techniques for the Prediction of Stock Price Trend, 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), IEEE, 2012.

[13] Chi Ma, Junnan Liu, Hongyan Sun, and Haibin Jin, A hybrid financial time series model based on neural networks, IEEE, 2017.

[14] Lean Yu, Shouyang Wang, and Kin Keung Lai, Mining Stock Market Tendency Using GABased Support Vector Machines, Internet and Network Economics, http://link.springer.com/10.1007/11600930_33, 2005, pp. 336–345.

[15] Fu-Yuan Huang, Integration of an Improved Particle Swarm Algorithm and Fuzzy Neural Network for Shanghai Stock

Market Prediction, 2008 Workshop on Power Electronics and Intelligent Transportation System.[Online].August 2008, IEEE, http://ieeexplore.ieee.org/document/4634852/, 2008, pp. 242–247.