Websites:
http://www.jofamericanscience.org
http://www.sciencepub.net

Emails:
editor@sciencepub.net
sciencepub@gmail.com

Journal of American Science

MARSLAND PRESS
Multidisciplinary Academic Journal Publisher

Forecasting the Trend of Taiwan Stock Price by Using of Deep Learning Techniques

Yi-Lang Ye[1]; Shen Cherng[1]; Hsien-Chiao Teng[2]; Ke-Yu Lee[1*]

[1]Department of Computer Science and Information Engineering, Chengshiu University
Niaosong, Kaohsiung City, Taiwan, ROC
[2]Department of Electrical Engineering, ROC Military Academy,
Fengshan, Kaohsiung City, Taiwan, ROC
*Corresponding author email: k0085@gcloud.csu.edu.tw

Abstract: In recent years, the stock market TWSE in Taiwan has been growing quickly. The explanatory factor of predicting stock price seems hard to be recognized. However, the forecasting trend of stock price is significant in time series, even though too many confounding factors may affect the forecasting. This study presents a framework to predict the trend of TWSE. The deep learning via Decision Tree and Random Forest is proposed for the analysis. The results demonstrated that stock prices of top twenty domestic companies can be forecasted with the accuracy of 63%.

[Yi-Lang Ye; Shen Cherng; Hsien-Chiao Teng; Ke-Yu Lee. **Forecasting the Trend of Taiwan Stock Price by Using of Deep Learning Techniques.** *J Am Sci* 2021;17(3):24-27]. ISSN 1545-1003 (print); ISSN 2375-7264 (online). http://www.jofamericanscience.org. 3. doi:10.7537/marsjas170321.03.

Key Words: A\Deep Learning; Decision Tree; Random Forest; TWSE

Introduction

In Taiwan stock market, the Formosa Stock Index was created by converting the existing Taiwan Stock Exchange (TWSE) and T

Meanwhile, the concern of the impact of U.S. stocks on the financial services industry (FSI) has been increasing year by year [2], same for TWSE, which affected by many external factors making it difficult to forecast. Deep learning is a tool that can be applied to predict the stock price [3] but seems with the low forecasting accuracy [4]. Recently, scientists reported that decision trees, random forests, and neural networks as well as other methods appeared to be able to deal with the stock market forecasting [5-7]. However, so far, no specific model can successfully predict the stock price due to the complicated external impact factors [8]. This study used of ten years data of stock prices of domestic top 20 companies in Taiwan, 20 global currencies, and 20 foreign companies to predict the trend of TWSE via artificial intelligence of decision tree and random forest analysis.

Back to 1974, scientists compared series data to determine forecasting accuracy [9]. The first effort that we know of forecasting the trend of time series data was by Spyros Makridakis [10]. Several convolutional neural networks (CNN) have been published to predict stock market prices [7, 11] with poor predicted benefits. In this study, we developed a deeding learning model to identify the features of time series of the TWSE stock prices. Usually available data are often huge and complicated, the time and cost for data processing has been increased a lot in artificial intelligence (AI) applications. One of the method is to reduce the high-dimensional features [12]. As more input features often make a predictive modeling task more challenging, dimensionality reduction can be used in applied deep learning to simplify the classification of dataset to fit a predictive model better [12, 13]. In deep learning, decision tree is a prediction algorithms, especially in data mining. Decision tree is also a frequently used technology to draw out the features in dataset. It is a classification technology based on the structure of dataset [14]. The advantage of using decision tree for decision-making rules is easily understood and applied. As the features of the dataset are judged and selected through the decision tree, the structure design via the logic rules of the tree should be carefully concerned [15]. Random forest (RF) is also used in deep learning applications like decision tree. The capabilities of RF are very good and easy to use. It belongs to a nonlinear artificial intelligence classifier model, combines multiple decision tree and output category. The key is random sampling and majority voting [13].

Methods

The information was obtained from Yahoo Finance and the Taiwan Stock Market public database from January 1, 2005 to January 1, 2015. As the object of this study is to find out the major features in dataset that may affect the trend of TWSE, we categorized the dataset into three parts, the first part is the stock prices of domestic top twenty companies, the second part is the global exchange rate of top twenty currencies, and the third part is referred to the stock prices of twenty foreign companies. Via the decision trees and random forests, we defined the index of the classification for the dataset and used the daily closing stock price of the Taiwan Weighted Stock Market Index to build a new binary index with the daily price of today, yesterday, and the day before yesterday. 0 was coded as price up and 1 as price down for the index label in the model analysis.

As the inconsistent data collection, we constructed a standard via the TWSE. Therefore, we used the quartile-range (q-tile) and fifth-range (f-order) to normalize the eigen-values **[16]**. The q-tile is to divide the data into four equal parts on average, and the value of the three split points as the quarter-tile. The first split point is called the first four-quarters and represented by Q1. The second split point, called the second four-quarters, is the median. The third split point is called the third four-quarters **[17]**. The f-order is used to denoise the data being sorted randomly. It divides a raw data into 5 levels, which are maximum of the data, average of the data, minimum of the data, high-mean and low-mean of the data. High-mean of the data is defined as the (mean+$\frac{max-min}{2}$) and the low-mean equals to (mean-$\frac{max-min}{2}$). In addition to use q-tile splitting, f-order can convert the features of the data set from continuous to discrete values including 1= low mean, 2 = high mean, 3 = min and 4= max. We also used the principal component analysis (PCA) to reduce the dimension of the dataset to get the eigenvalue and eigenvector on the principal axis with the smallest variance **[18]**. The algorithm types of decision tree learning comprise ID3, C4.5 and CART. These three are the most commonly used algorithms. The algorithm used in this study is CART for discrete label classification decisions. Gini is the process of building decision trees. Impurity increases the amount of information in the decision tree model **[19].**

Results

The data classified are categorized to be q-tile and f-order. Via PCA, the first three Eigen Vectors formed a new dataset. Through the decision tree and random forest training, the forecasting accuracy accounted for 10, 20, 30, and 40% of the model were listed in Table 2 and Table 2.

Table 1

Both f-digit and q-tile demonstrated top three Eigenvector with the variance at 10%, domestic companies have the highest amount of variation weighted, foreign companies are the second highest, and the global currency smallest.

Table 1. Both f-digit and q-tile demonstrated top three Eigenvector with the variance at 10%, domestic companies have the highest amount of variation weighted, foreign companies are the second highest, and the global currency smallest.

Parameters	Eigenvector 1		Eigenvector 2		Eigenvector 3	
	q-tile	f-order	q-tile	f-order	q-tile	f-order
D (Domestic Company)	38.9	44.1	20.7	22.1	17.9	14.0
G (Global Currency)	44.7	49.6	19.1	16.4	11.1	10.3
F (Foreign Company)	54	56.2	22.1	21.3	0.5	0.5

Table 2. Results of Random Forest Analysis with q-tile algorithm

NN-Data Processing	Test	Train	Test	Train	Test	Train	Test	Train
	1114	123	991	246	968	369	745	492
Parameter (q-tile)	Accuracy Rate							
D (100 trees)	59.6		57.6		58.0		55.7	
D (500 trees)	56.4		54.8		55.1		55.7	
D (1000 trees)	62.9		56.0		54.0		52.9	
G (100 trees)	58.8		53.3		52.6		51.9	
G (500 trees)	54.8		55.2		52.9		52.5	
G (1000 trees)	57.2		57.6		54.3		54.1	
F (100 trees)	52.5		55.2		52.6		53.3	
F (500 trees)	54.1		58.5		56.8		55.6	
F (1000 trees)	61.6		54.3		53.7		57.3	

Table 3. Results of Random Forest Analysis with f-order algorithm

NN-Data Processing	Test	Train	Test	Train	Test	Train	Test	Train
	1114	123	991	246	968	369	745	492
Parameter (f-order)	Accuracy Rate							
D (100 trees)	47.5		50.0		52.6		53.1	
D (500 trees)	54.8		51.6		51.8		49.0	
D (1000 trees)	55.6		53.6		53.7		51.9	
G (100 trees)	55.2		54.4		52.9		54.7	
G (500 trees)	54.0		51.4		53.2		53.5	
G (1000 trees)	57.2		56.0		54.0		55.3	
F (100 trees)	55.0		56.9		55.9		53.9	
F (500 trees)	51.6		55.6		56.2		54.1	
F (1000 trees)	56.6		53.9		50.5		56.0	

The forecasting accuracy rate via the decision tree analysis is 62.9% for our data q-tile data of domestic companies. Meanwhile, the forecasting accuracy rate is 55.2% for the q-tile data of foreign companies and 54.8% for the global currencies analysis. The results analyzed by Random Forest were 62.9 per cent for domestic companies, the second highest for foreign companies 61.6 % and, the smallest for global currencies at 58.8 % as well.

Conclusion

Using AI decision trees and random forests algorithms, we obtained the results of that Taiwan's stock market is more likely being affected by the stock prices of domestic top 20 companies. The highest forecasting accuracy rate of stock prices is about 63% via our model analysis. Although today's artificial intelligence is so progressive, the application to the stock market is still being limited accuracy rate, because of the randomized factors. Same with the human evaluation, fortune is the only predictor for forecasting stock market prices.

References

1. Tseng-Chung Tang, su-Tong Deng, and L.-C. Chi, *Textual Analysis of Stock Market Prediction Using Financial News Articles and Google Search Queries.* International Journal of Innovative Research in Science, Engineering and Technology, 2015. 4(8): p. 6992-6994.
2. Adrien, P., *THE FINANCIAL SERVICES INDUSTRY: EMERGING TRENDS IN THE OECS CREDIT UNION SECTOR /.* Savings and Development, 1997. 21(3): p. 275-294.
3. *Harvard Business Review tells us that nearly 85 percent of a company's performance is dependent upon external factors.*
4. Hall, J.L. and P.B. Tacon, *Forecast accuracy and stock recommendations.* Journal of Contemporary Accounting & Economics, 2010. 6(1): p. 18-33.
5. Gao, T. and Y. Chai, *Improving Stock Closing Price Prediction Using Recurrent Neural Network and Technical Indicators.* Neural Comput, 2018. 30(10): p. 2833-2854.
6. Jiang, W., *Applications of deep learning in stock market prediction: recent progress.* Statistical Finance, 2020. arXiv preprint arXiv:2003.01859.
7. Yang, C., J. Zhai, and G. Tao, *Deep Learning for Price Movement Prediction Using Convolutional Neural Network and Long Short-Term Memory.* Mathematical Problems in Engineering, 2020. 2020: p. 1-13.
8. Tiberius and Lisiecki, *Stock Price Forecast Accuracy and Recommendation Profitability of Financial Magazines.* International Journal of Financial Studies, 2019. 7(4).
9. P. NEWBOLD and C.W.J. GRANGER, *Experience with Forecasting Univariate Time Series and the Combination of Forecasts.* Journal of the Royal Statistical Society. Series A (General, 1974. J. R. Statist. Soc. A, (2): p. 131-165.
10. MAKRIDAKIS, S., *FORECASTING ACCURACY AND SYSTEM COMPLEXITY.* Recherche opérationnelle/Opérations Research, 1995. 29(3): p. 259-283.
11. Hao, Y. and Q. Gao, *Predicting the Trend of Stock Market Index Using the Hybrid Neural Network Based on Multiple Time Scale Feature Learning.* Applied Sciences, 2020. 10(11).
12. Alsenan, S.A., I.M. Al-Turaiki, and A.M. Hafez, *Feature Extraction Methods in Quantitative Structure–Activity Relationship Modeling: A Comparative Study.* IEEE Access, 2020. 8: p. 78737-78752.
13. RAFAEL MUÑOZ TEROL, et al., *A Machine Learning Approach to Reduce Dimensional*

Space in Large Datasets. Digital Object Identifier, 2020. 10(1109): p. 148181-148192.
14. Vijh, M., et al., *Stock Closing Price Prediction using Machine Learning Techniques.* Procedia Computer Science, 2020. 167: p. 599-606.
15. Živko, I. and M. Bošnjak, *Time Series Modeling of Inflation and its Volatility in Croatia.* Notitia, 2018(3): p. 1-9.
16. Jolliffe, I.T. and J. Cadima, *Principal component analysis: a review and recent developments.* Philos Trans A Math Phys Eng Sci, 2016. 374(2065): p. 20150202.
17. Daniel J. Lewis, et al., *Measuring Real Activity Using a Weekly Economic Index*. 2020.
18. Young Kyung Lee, Eun Ryung Lee, and B.U. Park, *PRINCIPAL COMPONENT ANALYSIS IN VERY HIGH-DIMENSIONAL SPACES.* Statistica Sinica, 2012. 22(2012): p. 933-956.
19. Tangirala, S., *Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm*.* International Journal of Advanced Computer Science and Applications, 2020. 11(2).

3/10/2021