Comparison of SVM and ARIMA Model in Stock Market

Yuxuan Zhu[1,*]

[1] *Rowe School of Business, Dalhousie University, Kenneth C. Rowe Management Bldg, 6100 University Ave, Halifax, NS, Canada, B3H4R2*
[*] *Corresponding author. Email: yx977729@dal.ca*

ABSTRACT

Forecasting the price of one certain asset is always a research hotspot in the financial area. This paper studies which of SVM model and ARIMA model is more suitable for short-term stock forecasting. This paper use SVM model and ARIMA model to predict the stocks of Tesla, Apple, Meta and Amazon, respectively. Next, the paper compares the accuracy of the two models and test which of the two models is suitable for the short-term prediction of the stock market. For the four companies studied in this article, ARIMA model is more accurate than SVM model in short-term stock price prediction. The results in this paper benefit the related investors in financial markets.

Keywords: *machine learning, support vector machine (SVM), Autoregressive Integrated Moving Average model (ARIMA), Stock price movement prediction.*

1. INTRODUCTION

Machine learning is a subject that studies how to build a computer system that can be improved automatically through experience [1]. It provides many evidence-based decisions for data-intensive industries such as financial modelling, education and manufacturing [1]. Machine learning is very suitable for processing a large amount of data, and because of its characteristics of autonomous learning, it can better predict the stock market. Thus, in the following research, machine learning is used for stock prediction.

Accurate stock forecasting has always been of great significance for decision-making in the financial industry [2]. Therefore, for all those who work in the financial industry, it is essential to accurately predict the stock trend and future stock price, which will play a vital role in their decision-making.

At present, there are many models used to predict the future trend of stocks, such as support vector machine (SVM) [3], Artificial Neural Network (ANN) [4], Autoregressive Integrated Moving Average model (ARIMA) [5]. Most of these models are based on machine learning. These models have been used to predict stocks. Kim once applied SVM to the prediction of stock price index, and discussed the possibility of its application to financial prediction [6]. Wanjawa & Muchemi studied the use of artificial neural network to predict the typical stock market [7]. They selected the data from the national I Olympic markets such as Nairobi stock exchange and New York Stock Exchange to do this [7]. Angadi & Kulkarni Used historical stock market data and ARIMA model, a model for predicting stock market trend is proposed on the basis of technical analysis [8].

However, the stock market is affected by economic, political, financial and social factors, and stock price characteristics have different performance in different time periods [9]. For example, the outbreak of COVID-19 in 2019 will seriously affect domestic demand in almost all countries [10]. Therefore, the purpose of this paper is to find the most suitable model among these models to predict stocks.

How to accurately predict the stock market has always been a very important problem. Therefore, in order to ensure that the model selected by the model is more accurate, SVM and ARIMA are compared, rather than just studying one model. Having a better model through comparison will make the stock price prediction more rigorous. Furthermore, few SVM and ARIMA models are used to predict the performance comparison of specific stocks in the short term. Therefore, the purpose of this study is to compare which model is more suitable for applying these two methods to short-term stock forecasting. From the results, ARIMA model is

Y. Jiang et al. (Eds.): ICEDBC 2022, AEBMR 225, pp. 928–934, 2022.
https://doi.org/10.2991/978-94-6463-036-7_137

more accurate in short-term stock prediction than SVM model. In terms of model accuracy, ARIMA model is more suitable for short-term stock prediction of the company. The specific data will be shown in the "results" part of this paper.

In order to know which model is more suitable for predicting the stock market, SVM and ARIMA are used to predict the stocks of four randomly selected U.S. stock market listed companies to determine which method is better in the specific application of short-term application investment. First, the stock data of Tesla, Amazon, apple and Meta are collected from Yahoo Finance and visualized by it. Second, use time series and support vector machine model to predict the future trend of the stocks of the four companies. Third, ARIMA model is used to predict the future trend of these four stocks. Fourth, compare the accuracy of the two models to obtain a more suitable model for stock forecasting. The accuracy comparison table of ARIMA and SVM is shown in the results of this paper. The results show that ARIMA model has higher accuracy and more accuracy than SVM model in terms of stock price prediction of these four companies in the short term.

The paper is structured is as follows: the second part introduces the methods used in this paper, the third part is empirical research, and the fourth part summarizes this paper.

2. DATA&ANALYSIS

The data comes from the historical data of Yahoo Finance (http://dinance.yahoo.com) from March 2015 to March 2022. The data collected is stock price data. However, the purpose of this study is to compare the future daily return $(r_t)$of the four companies. Therefore, in order to convert the price data into continuous compound income data, the following formula is used to calculate the daily rate of return:

$$r_t = \ln(1 + P_t) \quad (1)$$

Among them, (P_t) is the price of each stock on that day. Since four companies are involved in our study, the above formula will be reused in the study.

The specific annual data of Tesla, apple, Meta and Amazon are collected to compare the specific company data of the technology industry. These four companies were selected because they are all innovative enterprises. At the same time, these four companies are several companies with relatively good business capacity among innovative enterprises.

Founded in 2003, Tesla is committed to producing pure electric vehicles and clean energy power generation and storage products [13]. Tesla's mission is to accelerate the world's transition to sustainable energy [13]. Tesla has been rated by Forbes as one of the most innovative companies in the world [14].

Apple is a company dedicated to innovation. It has many functions and is the first in the world. For example, portal lighting (2017), and night mode (2019) are but a few of the improvements [15].

Amazon has usually been among the most innovative companies [16]. The company started from e-commerce, but it is not limited to e-commerce. It has developed in household goods, electronic equipment and even film investment, maintaining a rapid growth rate [16].

Meta Platforms, Inc. is committed to the research of the popular concept of meta universe in recent years. Their vision is to bring the concept of meta into life [17].

These four companies are based on innovation in different fields and have made great achievements, so this paper chose the specific data of these four companies as the research data. At the same time, they all belong to the companies covered in S & P 500, which shows that they have a certain popularity in the United States.

Table 1. Descriptive statistics of the Data

	AAPL	TSLA	AMZN	FB
Mean	65.74	220.07	1736.01	186.30
Variance	1956.71	89248.32	1029582.16	5597.14
Maximum	182.01	1229.91	3731.41	382.18
Minimum	22.58	28.73	366.37	77.46

It can be seen from the table that the average closing prices of AAPL, TSLA, AMZN and FB are 65.74, 220.07, 1736.01 and 186.30, respectively. The maximum value of AAPL is 182.01 and the minimum value is 22.58, the maximum value of TSLA is 1229.91 and the minimum value is 28.73. The maximum value of AMZN is 3731.41 and the minimum value is 366.37, the maximum value of FB is 382.18 and the minimum value is 77.46. The difference between the maximum value and the minimum value is large. It can also be seen from its variance that AAPL, TSLA AMZN The variances of FB are 1956.71, 89248.32, 1029582.16 and 597.14 respectively, which are large, indicating a high degree of dispersion of the data.

3. MATHOD

3.1 Support Vector Machine (SVM)

Support vector machine first appeared in 1992 and was introduced by Boser, Guyon and Vapnik in colt-92. Support vector machine uses machine learning theory, classification and regression prediction tool and maximizes the prediction accuracy through machine learning theory [18].

When using support vector machine, it is noticed that there are many linear classifiers to separate data, but only

one of them achieves the maximum separation. Therefore, the maximum edge classifier is an obvious solution. This is a simple example of maximum separation:

$$\text{margin} = \arg\min_{X \in D} d(x) = \arg\min_{X \in D} \frac{|\mathbf{x} \cdot \mathbf{w} + b|}{\sqrt{\sum_{i=1}^{d} w_i^2}} \quad (2)$$

However, in real-world problems, it is unlikely to get a completely separate data line in space. At this time, this paper introduce Lagrange variables to punish large relaxation

$$\min L = {}^{1}/_{2}\, w'w - \sum \lambda_k (y_k (w'^{x_k} + b) + s_k - 1) + \alpha \sum s_k \quad (3)$$

Accounting to HEO and Yang, the prediction results of support vector machine have good predictability [19].

3.2 Autoregressive Integrated Moving Average model (ARIMA)

ARIMA is an important method to study time series. According to Ariyo and Adewumi [20], ARIMA model has strong short-term prediction potential and can compete with the existing stock price prediction technology. It studies the characteristics of data by analyzing data sets and predicts the future value of the series according to the characteristics [21]. ARIMA model models the time series based on the box-Jemkins method, and regards the data series formed by the prediction object over time as a random sequence. ARIMA model can be expressed as:

$$(1 - \sum_{i=1}^{p} \emptyset_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^{q} \emptyset_i L^i)\varepsilon_t \; d \in Z, d > 0 \quad (4)$$

Where, L is the lag operator, p represents the lags of the time series data used in the prediction model, also known as the Auto-Regressive term, d represents the order difference differentiation required for the time series data to be stable, also known as the Integrated term, and q represents the lags of the prediction error used in the prediction model, also known as the moving average term.

4. RESULTS

In order to explore which of support vector machine and ARIMA model is more suitable for stock forecasting, this paper describes the relationship from both theoretical and empirical aspects. Theoretically, based on the existing research results, this paper makes a theoretical analysis of stock prediction from two aspects: the uncertainty of the financial market itself and the influence of COVID-19 on the financial market. In the empirical aspect, this paper selects the stock data of Tesla, Amazon, apple and Meta in recent years for data analysis. At the same time, these two models are used to predict the stock prices of these four companies [see figures 1,2,3,4,5,6,7,8]. The accuracy comparison of the two models is attached in Tables 1 and 2.

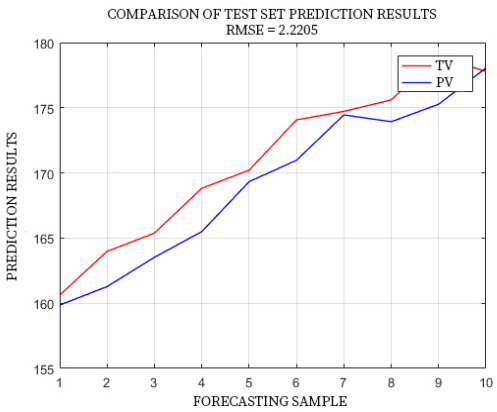

Figure 1 Prediction of AAPL stock price using SVM model

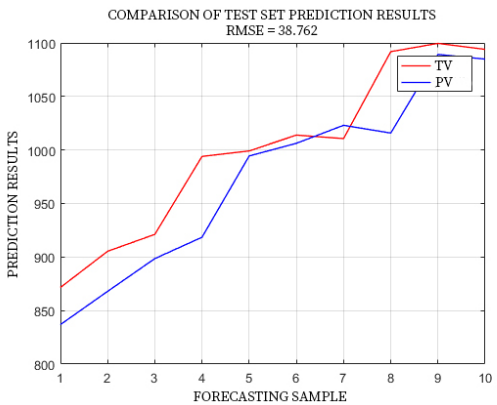

Figure 2 Prediction of TSLA stock price using SVM model

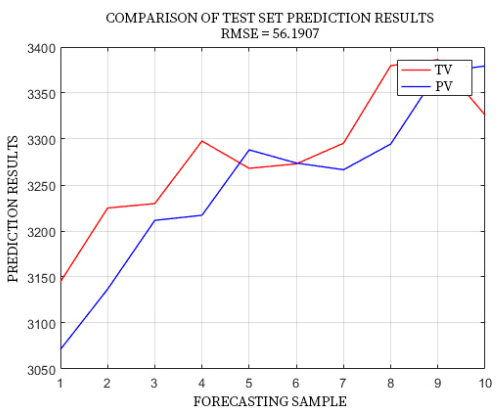

Figure 3 Prediction of AMZN stock price using SVM model

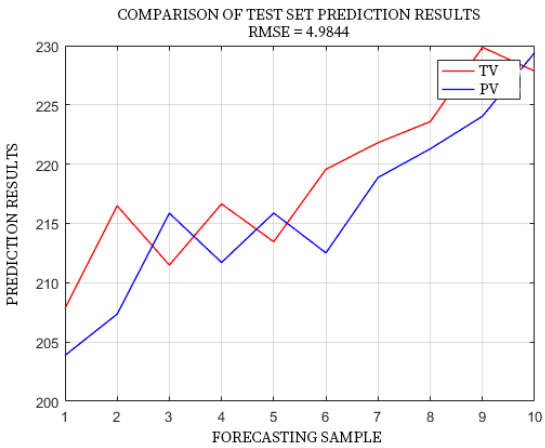

Figure 4 Prediction of FB stock price using SVM model

These four digures (figure 1, 2, 3, 4) are the comparative broken line charts of the closing price true value (TV) and predicted value (PV) of the stocks of tsla, AAPL, AMZN and FB using SVM model. It can be seen from these four figures that there is a large difference between the real value and the predicted value, and the short-term prediction of SVM model may not be perfect. Table 2 shows the error data of this model.

Table 2. Comparison of mean square error indicators of different companies in SVM model

	AAPL	TSLA	AMZN	FB
R^2	0.858	0.341	0.447	0.746
MSE	2.221	56.191	4.984	38.762

It can be seen from the table that the R^2 of AAPL, TSLA, AMZN and FB are 0.858, 0.341, 0.447 and 0.746 respectively, the MSE of AAPL is 2.221, the MSE of TSLA is 38.762, the MSE of AMZN is 56.191, and the MSE of FB is 4.984. It can be seen from the R^2 that APLL and TSLA are closer to 1, and the regression line fits the observed values better. It can be seen from MSE that the prediction models of APLL and FB have better accuracy. The next four figures (Figure 5, 6, 7, 8) will show the ARIMA model's prediction of the stocks of these four companies

Figure 5 Prediction of TSLA stock price using ARIMA model

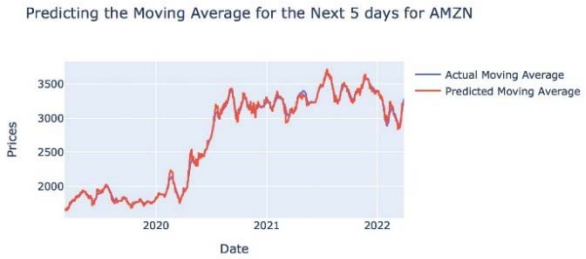

Figure 6 Prediction of AMZN stock price using ARIMA model

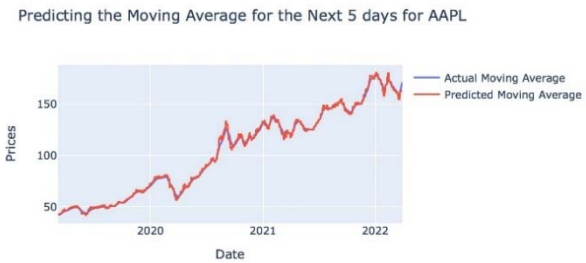

Figure 7 Prediction of AAPL stock price using ARIMA model

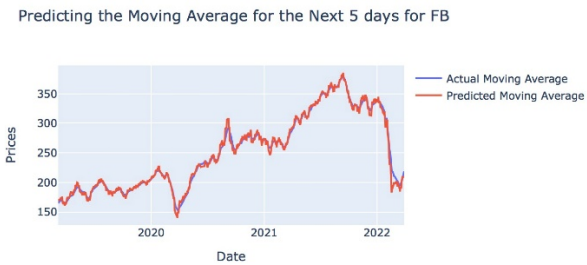

Figure 8 Prediction of FB stock price using ARIMA model

These four figures (Figures 1, 2, 3 and 4) are broken line charts comparing the real value (TV) and predicted value (PV) of the closing prices of tsla, AAPL, AMZN and FB by using ARIMA model. Compared with the prediction made by SVM model (Figure 1,2,3,4), the broken line of the predicted value made by ARIMA model is obviously more consistent with the actual value, and there is little difference between the actual value and the predicted value. This shows that ARIMA model is more suitable for short-term prediction of these four stocks than SVM model. Table 3 shows the error data of the model.

Table 3. Comparison of mean square error indicators of different companies in ARIMA model

	AAPL	TSLA	AMZN	FB
R^2	0.998	0.996	0.992	0.995
MSE	1.464	29.862	3.883	13.917

It can be seen from the table that the R^2 of AAPL, TSLA, AMZN and FB are 0.998, 0.995, 0.996 and 0.992 respectively, the MSE of AAPL is 1.464, the MSE of

TSLA is 13.917, the MSE of AMZN is 29.862, and the MSE of FB is 3.883. Their R^2 are all close to 1, but it is obvious that AAPL and AMZN are closer to 1 than the other two companies, it means that the regression line fits the observed values better. It can be seen from MSE that the prediction models of APLL and FB have better accuracy.

5. CONCLUSION

The purpose of this study is to compare which SVM model and ARIMA model are more suitable for stock forecasting. This study uses SVM model and ARIMA model to predict the stocks of Tesla, apple, meta and Amazon. After the prediction, the predicted value is visualized, compared with the actual closing price of the stock, and the broken line chart is used to make the comparison trend more obvious. After visualization, this study uses two evaluation methods, R^2 and mean square error, to evaluate the accuracy of the two models. The results show that in terms of short-term prediction, the prediction accuracy of ARIMA model is higher than that of SVM model, which shows that ARIMA model is more suitable for short-term prediction of stock price than SVM model. When financial personnel need to use this model to predict stocks in the short term in the future, they can give priority to ARIMA model for prediction, which will be very useful for financial practitioners to make model decisions.

This paper may have two defects. Firstly, this paper only compares the short-term stock prediction accuracy of SVM model and ARIMA model, and does not compare other models. This makes the conclusion of this paper only applicable to the comparison between the two models. Second, this paper only compares the short-term prediction performance of SVM model and ARIMA model in the stock market, and the results are only applicable to the short-term prediction within 10 days. If users want to know which model can more accurately predict the long-term stock value of the stock market, other studies are needed to compare.

REFERENCES

[1] M. I. Jordan, & T. M. Mitchell. (2015). Machine learning: trends, Perspectives, and prospects. Machine learning: Trends, perspectives, and prospects. Retrieved April 21, 2022, from https://www.cs.cmu.edu/~tom/pubs/Science-ML-2015.pdf

[2] Xiao, J., Zhu, X., Huang, C., Yang, X., Wen, F., & Zhong, M. (2019). A new approach for stock price analysis and prediction based on SSA and SVM. International Journal of Information Technology & Decision Making. Retrieved April 23, 2022, from https://www.worldscientific.com/doi/abs/10.1142/S021962201841002X

[3] Lin, Y., Guo, H., & Hu, J. (2013). An SVM-based approach for stock market trend prediction. IEEE Xplore. Retrieved April 23, 2022, from https://ieeexplore.ieee.org/abstract/document/6706743

[4] Vui, C. S., Soon, G. K., On, C. K., Alfred, R., & Anthony, P. (2013). A review of stock market prediction with Artificial Neural Network (ANN). IEEE Xplore. Retrieved April 23, 2022, from https://ieeexplore.ieee.org/abstract/document/6720012

[5] Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock price prediction using the ARIMA model. IEEE Xplore. Retrieved April 23, 2022, from https://ieeexplore.ieee.org/abstract/document/7046047

[6] Kim, K. J. (2003). Financial time series forecasting using support vector machines. Neurocomputing. Retrieved April 23, 2022, from https://www.sciencedirect.com/science/article/abs/pii/S0925231203003722

[7] Wanjawa, B. W., & Muchemi, L. (2014). Ann model to predict stock prices at stock exchange markets. arXiv.org. Retrieved April 23, 2022, from https://arxiv.53yu.com/abs/1502.06434

[8] Angadi, M. C., & Kulkarni, A. P. (2015). Time Series Data Analysis for stock market prediction using data mining techniques with R. ResearchGate. Retrieved April 23, 2022, from https://www.researchgate.net/publication/286890497_Time_Series_Data_Analysis_for_Stock_Market_Prediction_using_Data_Mining_Techniques_with_R

[9] Fenghua, W. E. N., Jihong, X. I. A. O., Zhifang, H. E., & Xu, G. O. N. G. (2014). Stock price prediction based on SSA and SVM. Procedia Computer Science. Retrieved April 23, 2022, from https://www.sciencedirect.com/science/article/pii/S1877050914004864

[10] Goodell, J. W. (2020). Covid-19 and finance: Agendas for Future Research. Finance Research Letters. Retrieved April 23, 2022, from https://www.sciencedirect.com/science/article/pii/S1544612320303974

[11] Makala1, D., & Li1, Z. (2021). Prediction of gold price with ARIMA and SVM. Journal of Physics: Conference Series. Retrieved April 23, 2022, from https://iopscience.iop.org/article/10.1088/1742-6596/1767/1/012022/meta

[12] Ming, W., Bao, Y., Hu, Z., & Xiong, T. (2014). Multistep-ahead air passengers traffic prediction with Hybrid Arima-SVMS models. The Scientific World Journal. Retrieved April 23, 2022, from https://www.hindawi.com/journals/tswj/2014/567246/

[13] Tesla. (2022). About Tesla. Tesla. Retrieved April 23, 2022, from https://www.tesla.com/about

[14] Dyer, J., Gregersen, H., & Furr, N. (2015). Tesla's Secret Formula. Forbes India. Retrieved April 23, 2022, from https://www.forbesindia.com/article/cross-border/teslas-secret-formula/41057/4

[15] Podolny, J. M., & Hansen, M. T. (2020). How Apple is organized for innovation. empowerment.ee. Retrieved April 23, 2022, from https://empowerment.ee/wp-content/uploads/2021/03/How-Apple-Is-Organized-for-Innovation.pdf

[16] Robischon, N. (2017). Why Amazon is the world's Most Innovative Company of 2017. publicaffairs-sme.com. Retrieved April 23, 2022, from https://publicaffairs-sme.com/PatriotFamily/wp-content/uploads/2015/01/Why_Amazon_is_most_innovative_company_13_Feb_2017.pdf

[17] Meta. (2022). Welcome to meta: Meta. Welcome to Meta | Meta. Retrieved April 24, 2022, from https://about.facebook.com/meta/

[18] Jakkula, V. (2006). Tutorial on support vector machine (SVM). course.ccs.neu.edu. Retrieved April 23, 2022, from https://course.ccs.neu.edu/cs5100f11/resources/jakkula.pdf

[19] Heo, J., & Yang, J. Y. (2016). Stock Price Prediction Based on Financial Statements Using SVM. gvpress.com. Retrieved April 24, 2022, from https://gvpress.com/journals/IJHIT/vol9_no2/5.pdf

[20] Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock price prediction using the ARIMA model. IEEE Xplore. Retrieved April 24, 2022, from https://ieeexplore.ieee.org/abstract/document/7046047

[21] Mondal, P., Shit, L., & Goswami, S. (2014). Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. researchgate.net. Retrieved April 23, 2022, from https://www.researchgate.net/publication/276197260_Study_of_Effectiveness_of_Time_Series_Modeling_Arima_in_Forecasting_Stock_Prices