**Journal of Computer Technology and Software**

**ISSN:2998-2383**

**Vol. 3, No. 8, 2024**

# Enhancing Stock Price Prediction Accuracy: An Integrated Learning Approach Based on SVR

**Marceline Rocher**
Department of Data Science, University of Technology Sydney, Australia
marceline@uts.edu.au

**Abstract:** The stock market serves as a critical indicator of a nation's economic vitality, reflecting trends and potential risks. With increasing participation from investors seeking high returns, the demand for efficient and accurate stock price prediction methods has intensified. Traditional time-series statistical models such as ARMA, GARCH, and Markov models have been widely applied, yet limitations remain in their predictive performance. This study identifies the insufficient generalization capability of the Support Vector Regression (SVR) model in stock price forecasting and introduces an integrated learning algorithm to address this issue. By combining SVR with the Linear Regression (LR) and K-Nearest Neighbor (KNN) models, the proposed approach capitalizes on the strengths of each method. Experimental results across multiple datasets demonstrate that the integrated model outperforms the standalone SVR model in accuracy and robustness. Furthermore, this paper highlights future research opportunities, including incorporating additional influencing factors and optimizing SVR parameters, to further enhance predictive capabilities.

**Keywords:** Stock price prediction, SVR, Integrated learning.

## 1. Introduction

A stock market serves as a vital indicator of a nation's economic development and is often referred to as a "barometer" or an "early warning device" of economic trends. Its origins can be traced back to 1611, when businessmen in Amsterdam engaged in the trading of shares from overseas trading companies, laying the groundwork for the modern stock exchange. Over the centuries, stock markets have undergone significant evolution, particularly in the 20th century, when the pace of development accelerated. This rapid growth, especially after 1970, was driven by the increasing scale and intensity of global trade as well as advancements in communication network technology, which facilitated faster and more efficient trading.

As living standards improved globally, and as individuals and institutions sought higher returns on their capital, participation in the stock market grew significantly. Despite this growth, the inherent volatility and unpredictable nature of stock prices posed significant risks. For most investors, especially those with a limited capacity to absorb financial losses, these fluctuations could lead to substantial economic setbacks. Consequently, there has always been a pressing demand for more efficient investment theories and strategies that can guide investors in maximizing returns while minimizing risks.

In recent decades, the focus of investment research has increasingly shifted toward leveraging advanced techniques for stock price prediction. Many researchers have sought to use historical market data as a foundation for developing predictive models. By arranging this data in chronological order, they employ various time-series statistical methods, such as the ARMA (Autoregressive Moving Average) model, GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model, Markov chains, and the Moving Average (MA) model, among others. These models help establish quantitative relationships between past and future stock prices, enabling researchers to predict short-term market movements with greater accuracy. Additionally, as computational power and data availability have expanded, more sophisticated approaches, such as machine learning and deep learning algorithms, have started to play an important role in enhancing the precision of stock price forecasts. These advancements represent a pivotal step forward in equipping investors with the tools needed to navigate the complexities of modern financial markets effectively.

## 2. Literature review

Since the 1990s, some researchers began to apply machine learning to stock price prediction. In the literature, Gencay et al. established a forward neural network model. The results show that the effect of the neural network model is better than that of the simple model MA, however, the prediction effect is not up to industrial standards. In the literature, Zhang et al. proposed a method to combine ARIMA and artificial neural networks, and proved that the prediction accuracy of this combination method is higher than that of the ARIMA model in the nonlinear data processing. Some scholars use short-term and short-term memory networks (LSTM) and cyclic neural networks (RNN) to predict stocks.

Support vector machine model (Support Vector Machine, SVM) is another machine learning model that is often used in stock price prediction research. SVM method uses the principle of post-festival risk minimization and has good generalization ability. Support vector regression (support vector regression, SVR) is an important branch of SVM. In 2003,

Kim used the SVR model in stock price prediction and compared it with the BP neural network model,it was found that the accuracy of the model was higher. After that, more and more investors start to use the SVR model to predict the stock price. We used the SVR model to predict the stock price and found that the price forecast of individual stocks (such as Amazon stock, stock code AMZN, and Tianchang group, stock code 2182.HK) was very deviated. Aiming at this problem, we suggested an improved algorithm and will be proposed in this paper.

Two are selected as the analysis resultants of the model predication. One is Amazon(AMZN),the other is Ping An Bank(000001). The data of two model are the historical data of the two stocks fromJune 2016 to June 2020.

Amazon stock(AMZN):It can be found from the data that AMZN mainly shows a increasing trend, rising volatility, while it will fall in a few cases. The range for rising is within 11per cent. However, the range of decline is wide. And in the period of growing, the closing price is higher than opening price. Meanwhile, the closing price is lower than the opening price during the decline.

Ping An Bank(000001):it can be also found from the data that there are great variation for Ping An Bank,up and down. In total, the rise are almost same as the fall. Even ,there are continue drops. The value can increase by up to 13 per cent and decrease by down to 14 per cent. At the same time ,the closing value is higher than opening value in rise progress. At the same time, the closing value is lower than opening value in descending progress.

Short-term securities refer to all kinds of marketable securities that can be realized at any time and held for no more than one year, as well as other investments for no more than one year. I found a fewstocks to have a general study of their trend. The first stock is NIKE, whose overall trend over the past year has been upward. From 90 per share to about 120 per share. However, it has a long period of decline in between. It began to decline from about February 20, 2020, and reached the lowest at the end of March, about 60 per share. This is probably because of coVID-19, which is causing peoplenot to go shopping, and some other reasons, which is causing the stock price to plummet. It did not begin to return to pre-epidemic levels until early June. After that, the stock began to rise until September of this year. The second stock is Boeing. From September last year to the end of February this year, the stock price has been in a very stable, little volatility, and has remained at 340-380, untilthe end of February this year, the stock price began to fall rapidly. From the end of February to the middle of March, the share price fell from 340 to 95 in just half a month. The share price has never been back to where it was. Until a small peak in early June, the stock has remained around 170 sincethen.

SVR (Support Vector Regression) is same as SVC, selects a part of data from the practiced databaseto support vector more positively. And analysis the predicted object through the value of sample.

Linear regression refers to using linear functions to fit samples in vector Spaces. The model takes the comprehensive distance between the actual positions of all samples and the linear function as the loss,and calculates the parameters of the linear function by minimizing the loss. For linear regression, a sample is calculated as long as it does not fall exactly on the linear function as the model.

The model of SVR is also a linear function ‘y= kx + b’, but different from linear regression. It is a more tolerant predictive model than linear regression. The SVR creates an "interval band" on both sides of the linear function. For all samples falling into the interval band, the loss is not calculated. Only those outside the interval band are included in the loss function. Then the model is optimized by minimizing the width of the interval band and the total loss.

## 3. Experiment

### 3.1 Experimental data

We obtained the stock historical data of many domestic and foreign companies through the yfinancemodule. In the experiment part, we mainly focused on the price prediction of individual stocks (such as Amazon stock, stock code AMZN and Tianchang Group, stock code 2182.HK), because the SVR model showed a large deviation in the price prediction of these several stocks.

We mainly obtained the stock prices of Amazon and Tianchang Group during 2000-01-01 to 2018- 12-01, and used the closing price as the target value for training and prediction.

We will get 70% of the data as training data and 30% as testing data.

### 3.2 Experimental results

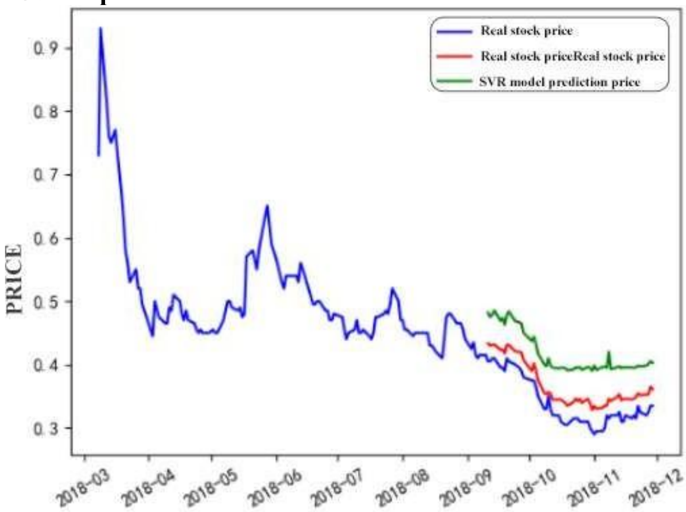


**Figure 1.** Amazon stock forecast results

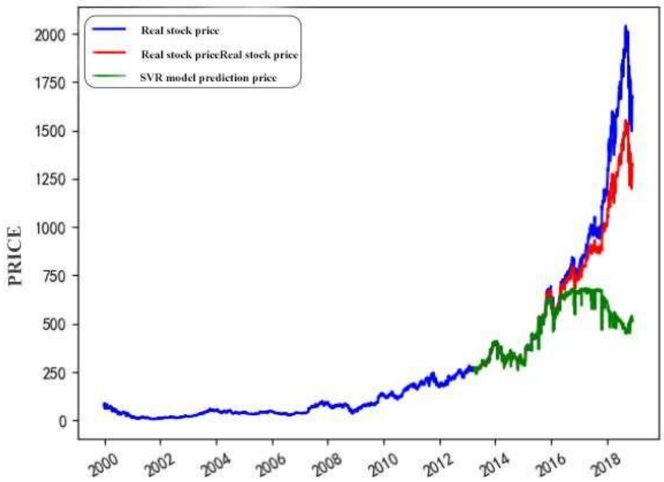


**Figure 2.** Tianchang Group Stock Forecast

We used the SVR model and the integrated model proposed in the paper to forecast the stock prices of Amazon and Tianchang Group respectively, and the experimental results are shown in Figure 1-2below.

It is obvious from the figure that the prediction result of the integrated model is better and more accurate than that of the simple SVR model, no matter for the prediction of

Amazon's stock or the stock price of Tianchang Group.

Table 1 lists the results obtained after the training of stock price data by the two models. In addition to the root mean square error RMSE and determination coefficient introduced before R2, the evaluation index also has a score of 10% discount cross validation. K folding cross validation (k-fold cross-validation) first divides all the data into K sub-samples, one of which is not repeated as the test set, and the other K-1 samples are used for training. Repeat K times and get a score of K times on average. Here we take K=10.

**Table 1.** The results obtained after the training of stock price data

| | SVR Model | Integration Model |
|---|---|---|
| RMSE | 0.0674 | 0.0319 |
| | 273.8249 | 82.44144 |
| R2 | 0.5339 | 0.8987 |
| | 0.5512 | 0.9588 |
| Cross_val_score | 0.1218 | 0.8989 |
| | 0.5367 | 0.9587 |

In Table 1, the shadow table is the data of Tianchang stock, and the blank background is Amazon stock. We can see from the table that the three evaluation indexes of the integrated model are all improved to a certain extent than the single SVR model, which shows that the root mean square error RMSE decreases, the correlation index increases, and the cross-validation score increases on the test set. The evaluation index and cross-validation score have been greatly improved.

## 4. Summary

This paper identifies the issue of limited generalization ability in the Support Vector Regression (SVR) model through prediction experiments conducted on multiple stock prices. While the SVR model is a widely used and robust approach in regression tasks, its performance in stock price prediction reveals shortcomings in capturing the broader variability of stock market dynamics. To address this limitation, the paper proposes an improved algorithm based on the concept of integrated learning. The essence of integrated learning lies in combining multiple weak learners in a manner that leverages their individual strengths and mitigates their weaknesses, ultimately resulting in a more powerful and generalized learner.The integrated model proposed in this study builds on the SVR framework by incorporating two additional components: a simple and effective linear regression (LR) model and a K-nearest neighbor (KNN) model. The integration of these models aims to harness the complementary strengths of linear regression for simplicity and interpretability, KNN for capturing local data patterns, and SVR for its robustness in handling non-linear relationships. By combining these models, the integrated learning approach creates a comprehensive predictive model that can better handle the complexities of stock market data.To validate the effectiveness of the proposed integrated model, the paper conducts extensive experiments using different stock datasets. The experimental results demonstrate that the integrated model outperforms the standalone SVR model in terms of multiple evaluation metrics, such as prediction accuracy and error reduction. These findings highlight the potential of the integrated learning approach in enhancing the predictive capabilities of traditional machine learning models in financial applications.However, the paper also acknowledges areas for further improvement and exploration. One of the limitations identified is that the current model focuses solely on the closing price as the influencing factor. Future studies could incorporate additional factors, such as trading volume, market sentiment, or macroeconomic indicators, to enhance the predictive performance further. Moreover, the SVR model in this study uses default parameter settings, which may not be optimal for all datasets. Conducting a thorough parameter optimization process, such as using grid search or evolutionary algorithms, could potentially yield better results. These insights point to promising directions for future research, aiming to refine the model and improve its applicability in dynamic and complex stock market environments.

## References


[1] Zhao Zhenquan, Jiang Yingkun, Chen Shoudong, Dai Li. Appropriate scale of China's securities market. China in 2001: Analysis and Forecast of Economic Situation, 2001.

[2] Xu Haoran, Xu Bo, Xu Kewen. Overview of the application of machine learning in stock forecasting [J].Computer Engineering and Applications, 2020, 56(12): 19-24.

[3] Vapnik, Vladimir. The nature of statistical learning theory. Springer science & business media, 2013.

[4] Kim, Kyoung-jae. "Financial time series forecasting using support vector machines." Neurocomputing 55.1-2 (2003): 307-319.

[5] Mei, Wenjuan, et al. "Stock price prediction based on ARIMA-SVM model." (2018).