

Stock Market Prediction Using Support Vector Regression to Account for a Modern Approach

Christopher M. Frutos and Andrew B. Garcia

A. Abstract

As of recently, the stock market has been heavily affected by recent events, such as the war in Ukraine, fluctuation in crypto currency, and the coronavirus pandemic. For investment companies, such as Fidelity or Vanguard, knowing the difference between what stocks will perform the best, and the ones that won't, is invaluable to their business plan. The research in this paper will be using publicly available data like stock market prices, commodities, and cryptocurrency. Our approach to achieving stock price prediction is a continuous problem, which is why we are choosing to implement a Support Vector Regression (SVR) model.

B. Specific Aims

The main goal of this paper will be a novel approach to designing a month over month prediction model of the S&P 500 based on key factors like developing technology and pop-culture trends. We use the S&P 500 as our target predictor because it is widely considered to be a representative of overall market health. Related data, such as the number of COVID cases, price of commodities, cryptocurrency, and inflation will all be key factors integrated into our data set for training and testing. For our model, we will be using Support Vector Regression (SVR), which is a type of non-linear classification algorithm, that will help us predict future stock market prices.

C. Background

A Support Vector Machine (SVM) is a type of algorithm that classifies data by finding the hyperplane in an N-dimensional feature space that distinctly classifies the data points [1]. In

order to determine the best hyperplane on a feature set, as shown in **Figure 1**, the algorithm iterates over all possible separations and picks the one with the largest margin represented as epsilon (ϵ).

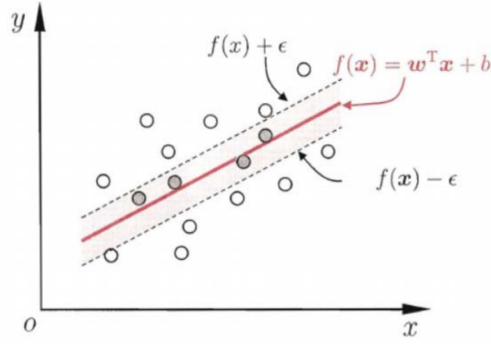


Fig. 1: Implementation principle of Support Vector Machines [2]

SVM's are powerful tools, however, our application is not linear which is why we have to use a support vector regression (SVR). SVR takes the concept of classification and adds Lagrange multipliers to maximize its objective function as seen in **Equation 1**.

$$(1) \quad Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(x_i, x_j)$$

, where $K(x_i, x_j)$ represents the nonlinear inner-product kernel, and α represents the Lagrange multipliers. There are two inner-product kernels that we plan to test in this paper, (1) polynomial and (2) radial-basis function, to see if there is any significant difference for our problem.

D. Research Design and Method

In our machine learning model, we will be using stock market data and splitting it into training data and test data for feature vectors. The output will produce a continuous value

reflecting the predicted stock market price for each month. We will attempt to adjust parameters such as how far the stock will be predicated into the future, removal and addition of different feature vectors, and exploring different models where most applicable.

E. Datasets and Experiments

In this experiment we will be using various datasets as feature vectors to determine the predictability of the model. We have determined that the most prominent data determining the health of the stock market are the price index for silicon, steel, oil and gas, and cryptocurrency.

Below is a table of the kind of datasets we will use in a .csv (comma separated values) file format and the source sites for each data set. The data has been modified to include the last 5 years of data, as this should be sufficient data to train the model. The reason we also choose to only use the last 5 years, is that perhaps data from the past may not have significant influence on the present state of the market; hence why we chose to also use modern indicators such as crypto and silicon.

Dataset Table

S&P500	https://www.kaggle.com/datasets/camnugent/sandp500?resource=download&select=all_stocks_5yr.csv
Oil Index	https://markets.businessinsider.com/commodities/oil-price?type=brent
Bitcoin	https://www.blockchain.com/explorer/api
Gas index	https://markets.businessinsider.com/commodities/natural-gas-price
Steel index	https://www.wsj.com/market-data/quotes/index/XX/STEEL/historical-prices
Silicon index	https://fred.stlouisfed.org/series/PCU334413334413A

F. Timeline

Before mid-November we plan to build our model in python and have our data organized and revised for proper components. After the end of November we plan to have already established appropriate results and compose our analysis.

G. Key References

- [1] Gandhi, R. (2018, July 5). *Support Vector Machine - introduction to machine learning algorithms*. Medium. Retrieved October 14, 2022, from <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- [2] Chen, J. (2021). Application of support vector machines and Holt-Winters Model in local finance forecast. *2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)*. <https://doi.org/10.1109/acctcs52002.2021.00076>
- [3] Zhu, Y. (2021). Research on financial risk control algorithm based on machine learning. *2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*. <https://doi.org/10.1109/mlbdbi54094.2021.00011>
- [4] Nugent, C. (2018, February 10). *S&P 500 stock data*. Kaggle. Retrieved October 14, 2022, from https://www.kaggle.com/datasets/camnugent/sandp500?resource=download&select=all_stocks_5yr.csv

- [5] *Blockchain developer apis*. (n.d.). Retrieved October 15, 2022, from <https://www.blockchain.com/api>
- [6] Business Insider. (n.d.). *Natural gas price today | natural gas spot price chart | live price of natural gas per ounce | markets insider*. Business Insider. Retrieved October 14, 2022, from <https://markets.businessinsider.com/commodities/natural-gas-price>
- [7] *Producer price index by industry: Semiconductors and related device manufacturing: Other semiconductor devices, including parts such as chips, wafers, and heat sinks*. FRED. (2022, October 12). Retrieved October 14, 2022, from <https://fred.stlouisfed.org/series/PCU334413334413A>
- [8] Journal, W. S. (n.d.). *Steel | NYSE American Steel Index Historical Prices - WSJ*. The Wall Street Journal. Retrieved October 14, 2022, from <https://www.wsj.com/market-data/quotes/index/XX/STEEL/historical-prices>