The Optimal Method for Predicting the Trend of the NASDAQ Stock Exchange Index

***Abstract*** - ***Predicting Stock Market prices in today’s era is very valuable and profitable. It is not an easy task though. It can be affected by multiple factors, such as gross domestic product, politics, exchange rates, gold prices, etc. In this paper, we aim to use different machine-learning models to predict the stock index of the NASDAQ Composite to figure out the best model to do so.***

***Keywords - NASDAQ Composite; Logical Regression; Support Vector Regression; Long Short-Term Memory; K-Nearest Neighbors; Random Forest; Artificial Neural Network; Hybrid Models;***

1. Introduction

The NASDAQ Stock Index is a composite of all the stocks listed and traded under the NASDAQ Stock Exchange. In this paper we aim to train machine learning with Logical Regression, Support Vector Regression, Long Short-Term Memory, K- Nearest Neighbors, Random Forest, Artificial Neural Network, and Hybrid Models, so as to predict the trend of the NASDAQ for the next day. We will be using RMSE value, MAE value, R-Squared value to find the best model and Sharpe Ratio, which measures performance of an investment.

1. Machine Learning Models
2. Logical Regression

Logical Regression(LR) is one of the most applied models for real-life situations [1]. LR considers, that the data

Bernoulli Distribution and uses gradient descent to solve for parameters to achieve binary classification. LR models are based on the Signomid function. The Signomid function formula and graph are shown in Figure 1.

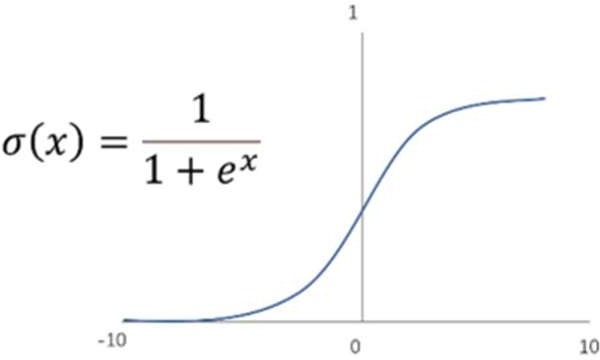


Figure 1: Signomid function[1]

1. Support Vector Regression

SVR is an implementation of SVM for performing regression [1]. SVR is a component of the SVM algorithm that is used to handle regression cases. SVR reduces the limits of generalization errors and can produce good results. It is relatively good for high-dimensional data [3], [5]. A hyperplane is a line that separates the data. In the given figure, the hyperplane is f(x).

The closest

data to the margin is the support vector and the margin is the distance between the hyperplane to the closest data.

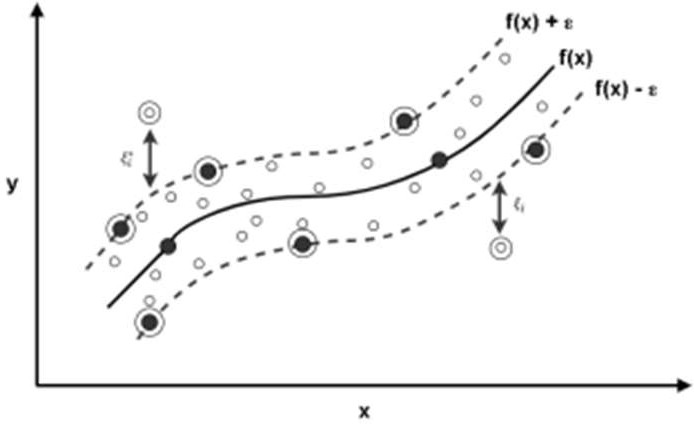


Figure 2: Principle of Support Vector Regression [5]

1. Long Short-Term Memory

Long Short-Term Memory (LSTM) is a gradient-based algorithm introduced to solve the problems in RNN [5]. LSTM works best on time series data [5]. Both LSTM and RNN can be used for storing information for a longer duration [11] as LSTM is designed to hold onto memory in the case of time series [11].

1. KNN Algorithm

K-Nearest Neighbor is one of the most used algorithms in machine learning [6]. KNN algorithm is often used on larger datasets, low training costs, and requires faster data update and quicker prediction due to fast training time.

1. Random Forest

Random Forest(RF) is a machine learning model that can be used to perform both classification and regression. RF’s main principle is to create multiple versions of Decision Trees which are all trained on different features and different samples. The final forecasting would be the average outputs of all Decision Trees [1]. The RF model allows the selection of different data from a single dataset, which increases the variance of the classification model which in turn gives more accuracy to the regression model [2].

1. Neural Network

In this each neural network is connected to various other neurons that allow the signal to navigate from one direction across the network from input to output. With a single hidden layer of sufficient complexity, any unknown parameterized function can be approximated to a certain degree of precision. As a result, the Artificial Neural Network model is consistent for both the input and output layers [7].

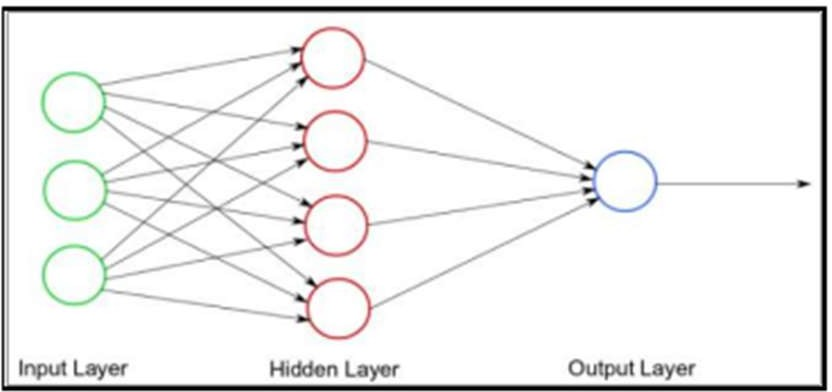


Figure 3: Simple Neural Network[7]

1. Hybrid Model

When combining 2 or more different models is called a hybrid model. And Zichen Fu [8], on combining ANN, SVM, and linear aggression, observed more accuracy from the hybrid model. The hybrid model showed 37% more annual profit margin and approximately 62% accuracy on S&P 500 intraday prediction.

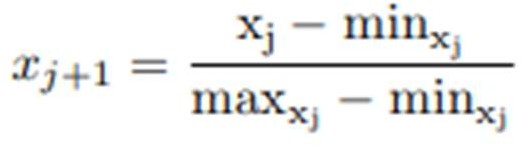
1. Literature Review
2. Data Collection

Several factors impact market performance: Market History, the NEWS, General Public Mood, Commodity price, Interest Rate, and Foreign Exchange. Data collected for this study is spread over 3 months from September 2015 to January 2016 and the data collected from NEWS and Twitter were from the current day closing to the next day’s closing. The data from NEWS and Twitter were processed with the help of OpinionFinder Library from the feed.

1. Data Processing

The collected data was developed so that it can be converted into a form that can be used as the input for the model. The data collected from the mining were processed to be declared as POSITIVE or NEGATIVE. Similarly, the other data were processed and used for the rest of the parameters [10].

While in [5] the data processing is done with the LSTM and SVR model and converted into some range intervals.



Where X’ stands for normalized data and X is the data before normalization, minxj is the minimum value of the data per column and maxxj is the maximum value of the data per column [5].

From the collected data, date format is removed as it is not useful for training the machines. However, the data are arranged in the date order starting from 1 [6].

1. Methodology

In [1], they introduced new variables, mostly different moving averages of stock prices and the difference between the highest/lowest price and opening/closing price of the day. Whereas in [3] they used Principal Component Analysis, therefore they created 10 sets of data. After solving for characteristic root and eigenvectors of the covariance matrices, they obtained 6 sets of data with a contribution rate 90% higher than the original sets.

Sliding Window strategy was used in [4] as they considered that data can get old and needs to be replaced in order to get better accuracy from the model and they needed a dynamic model for predicting the trend of the stock index.

1. Conclusion

It can be concluded that, if and when an appropriate model is trained for predicting stock market index with excellent accuracy, it would decrease the risk factor involved in investing in stock market drastically.

1. Refrences
   1. I. Medarhri, M. Hosni, N. Nouisser, F. Chakroun and K. Najib, "Predicting Stock Market Price Movement using

Machine Learning Techniques," 2022 8th International Conference on Optimization and Applications (ICOA), Genoa, Italy, 2022, pp. 1-5, doi: 10.1109/ICOA55659.2022.9934252.

* 1. J. Xue, "Prediction of NASDAQ Index Based on Stacking Method," 2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE), Hangzhou, China, 2021, pp. 505-510, doi: 10.1109/ICAICE54393.2021.00103.
  2. S. Zhu, M. Zhao, S. Wei and S. An, "Stock Index Prediction Based on Principal Component Analysis and Machine Learning," 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), Bangkok, Thailand, 2020, pp. 246-249, doi: 10.1109/ICBASE51474.2020.00059.
  3. Y. Qian, "The Optimal Model for Predicting the Trend of Stock Index Through the Comparison of Different Classifiers," 2020 2nd International Conference on Economic Management and Model Engineering (ICEMME), Chongqing, China, 2020, pp. 538-543, doi: 10.1109/ICEMME51517.2020.00113.
  4. N. M. Sarumpaet, Indwiarti and A. A. Rohmawati, "Performance Comparison Between Support Vector Regression and Long Short - Term Memory for Prediction of Stock Market," 2022 10th International Conference on Information and Communication Technology (ICoICT), Bandung, Indonesia, 2022, pp. 168- 173, doi: 10.1109/ICoICT55009.2022.9914839.
  5. J. Tan, "Stock Index Forecasting Model based on Short-term Volatility Trend and KNN Algorithm," 2022 8th Annual International Conference on Network and Information Systems for Computers (ICNISC), Hangzhou, China, 2022, pp. 583-589, doi: 10.1109/ICNISC57059.2022.00120.
  6. T. Paul, "Fintech Empowers Prediction of Stock Market Index Using Artificial Neural Network," 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST), Yogyakarta, Indonesia, 2021, pp. 42-46, doi: 10.1109/ICAICST53116.2021.949781 7.
  7. Z. Fu, "Machine Learning Models' Combination for Higher Accuracy of

S&P 500 Index Prediction," 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), Bangkok, Thailand, 2020, pp. 215-218, doi: 10.1109/ICBASE51474.2020.00053.

* 1. V. Moni, "Machine Learning to Predict Annual Stock Market Index - a Genetic Programming Approach," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), Chennai, India, 2019, pp. 1-4, doi: 10.1109/ICIICT1.2019.8741439.
  2. M. Usmani, S. H. Adil, K. Raza and S.

S. A. Ali, "Stock market prediction using machine learning techniques," 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, Malaysia, 2016, pp. 322-327, doi: 10.1109/ICCOINS.2016.7783235.

* 1. D. Bharathi, G. Yamini, and G. K. Chakravarthi, “Long short term memory for forecasting the stock market,” vol. 29, 2020.