ISSN: 0974-5823 Vol. 7 No. 5 May, 2022

International Journal of Mechanical Engineering

Stock Price Prediction Using Machine Learning: An Extensive Review

Utsav Dhall, Satyam Sharma, Aradhya Aggarwal

B.Tech ,B.Tech , B.Tech

Computer Science and Engineering.

Lovely Professional University, Phagwara, India.

Abstract:

With stock price prediction, we aim to predict the future state of a financial stock using the history of behavior of the said stock. We used Python as our programming language to build the models. These models will identify patterns in the behavior of the stock during recent history and train a model based on the data. The model will use insights from these patterns to predict future values with reasonable accuracy.

Keywords: *Regression, Patterns, Support Vectors.*

Introduction

In earlier time periods, professional stock traders used to deal with the stock market without the interactive tools we have at our disposal today, including but not limited to, Machine learning algorithms. Billions of datasets are generated every day from the stock market around the world , increasing the “volume”, “velocity”, “variety” and “veracity” of stock market data, and making it extremely difficult to analyze. Stock market is very non-linear and unpredictable [15]. Accurate stock prediction systems turn the high-risk high-reward nature of stock trading into low-risk high-reward because the predictions are known to be very accurate, of course considering the market is not affected by extraordinary circumstances (E.g., Pandemics) that the algorithm can’t account for , These exogenous factors increase the risk of using these models[9].

In current times, advanced techniques based on fundamental analysis of the stock market are used for predicting future stock values. The data size of the stock market analysis is extremely large and non-linear. To deal with this vastness, efficient systems are required that can identify patterns and relations. Our algorithms/Techniques have greatly improved. The commonly used models for long-term stock market prediction are the commonly known Capital Asset Pricing Model and Arbitrage Pricing Theory.[12]

The authenticity of the data that will be used to train the model must be accurate, even minuscule changes in the dataset can have something similar to a butterfly effect on the outcome of the prediction. In this project, we use the yahoo finance library that’s available for python. There are many ways of predicting and visualizing the stock market data but one of the most efficient and accurate ways we could come up with is the use of Dash. We will be using Dash (a python framework) and some machine learning models which will show company information and stock plots adjacent to the stock code given by the user. For historical price and volume, the first choice should be the popular Yahoo Finance dataset, which provides free access to data.

Using the Dash library, we can plot dynamic plots of financial data of specific companies by using the tabular data, and also the data which will be used in plotting is provided by the Yfinance library in python.Yfinance is a library that allows the users to fetch financial data of a company since its listing on the stock market from its stock code directly. The main website structure will be made by using mainly Dash HTML components and dash core components. The site’s UI is enhanced by styling using CSS.

Gunicorn and XML libraries will be used for applications development to host this app on the targeted server. Sklearn and scikit-learn are the tools used in the development of machine learning models. The inputs

required from the user will be like stock code, date range selector, number of days of forecast, and buttons. These are the components with which the user will interact.

Related Work

Ishita et al. employ regression and long short-term memory networks on the yahoo finance datasets. The study concludes that LSTM based models perform better for this problem.[1]

Ayaz et al. propose a method using a Long Short term memory network, while also using an autoregressive fractionally integrated moving average (ARFIMA), The paper concludes that an ARFIMA-LSTM hybrid performs better than ARFIMA independently.[2]

Kan et al. employ a combination of long short-term memory networks and generative adversarial network paired with a multi-layer perceptron. Their model performs better than ANNs and independent LSTM models. [3]

M Umer Ghani et al. employ Linear Regression, 3 Months moving average and Exponential Smoothing, they conclude that Exponential smoothing gives the best results [4]

Adil et al. use LSTM recurrent neural networks to approach this problem, they experiment on two stocks from the New York stock exchange. Their results show promising results.[5]

Uma et al. explore a modern approach to this problem; they use Ensemble machine learning methods concluding that EML is more accurate compared to other single ML algorithms. They also review most of the existing popular methods for stock prediction.[6]

Ashish et al. propose an interesting approach of combining multiple existing techniques to make a hybrid model. They analyzed the news sentiment with respect to the prediction value and found that prediction value is good when the sentiment is good and vice versa.[8]

Tejas et al. employ social sentiment analysis using tweets collected from the Twitter api and Support vector regression. In their analysis they find that SVR model is more feasible in comparison to sentiment analysis [7]

Comparative Analysis

Author.	Year	Algorithms/Technique	Observations	Merits/Scope
Ishita et al.[1]	2018	Regression, LSTM	LSTM provides more accurate results.	The model can be improved by utilizing a much bigger dataset.
Ayaz et al.[2]	2020	LSTM, ARFIMA	LSTM-ARFIMA hybrid performs better than independent ARFIMA	The addition of external input of dependent variables into the hybrid model increases the accuracy.
Kan et al.[3]	2019	LSTM, GAN and Multi-Layer Perceptron.	Proposed model works better than standard ANNs and independent LSTMs	Plan on exploring extraction of more valuable insights to increase precision of prediction

M Umer Ghani et al.[4]	2019	Linear regression, 3 months moving average and Exponential smoothing	Exponential smoothing is the best of the three approaches	More recent data is weighted differently than previous data, it ends up providing better insight into the behavior of the stock market.
Adil et al.[5]	2020	Long short-term memory networks	Training with more epochs and less data gives better and promising results with their model.	Their model is capable of understanding the pattern in the change in opening prices.
Uma Gurav et al.[6]	2018	Ensemble Machine Learning	EML methods are more accurate than single ML methods.	They decrease error variation and end up improving the accuracy of automated systems.
Tejas et al.[7]	2018	Social Sentiment Analysis and Support vector regression.	Classification of tweets as positive, negative and neutral gives a good overview of public mood	SVR outperforms sentiment analysis and gives better accuracy.
Ashish et al.[8]	2019	Sentiment Analysis	Improving upon the training data's scale and timeframe can result in better prediction.	The proposed model can successfully recommend the best stocks for investment.

Methodology

The stock market is one of the most volatile places where the price of stocks is very different from the last second. In this environment, predicting a stock price for a short period is extremely difficult, there are machine learning models which can predict the stock price but accuracy is quite low to implement on a large scale. Whereas predicting stock price over a long period is fairly easy because almost all businesses go through the same stages of growth and revenue generation which means almost all businesses show patterns, it is one of the things that machine learning is extremely good at recognizing. Though, it depends upon the kind of machine learning algorithm being used. Some researchers have considered the optimization techniques to boost the efficiency of existing machine learning algorithms to predict the stock market with improved accuracy [16]

There are hundreds of algorithms that can be used to recognize the stock price pattern and forecast the stock price. For linearly inseparable data, it uses a nonlinear mapping to transform the original data into a different dimension. Within this new higher dimension, it searches for an optimal hyperplane that separates the two classes. Here, we are using one of the fastest and most commonly used algorithms for such purposes that come under support vector machine algorithms, usually, it is referred as support vector regression (SVR). To

obtain an optimal regressor function for a given set of training data SVR was introduced by Vapnik [10]. The SVR adopts the concepts of SVM but there is only a small difference between a SVM and SVR.[17] SVMs can perform a little better than techniques like CNN [19]. Even though the logistic regression algorithm is possibly the most widely used supervised learning model of this era [20]. It is not always the best performing in all cases.

Support vector regression uses a similar principle as a Support vector machine to classify data points but is a bit different in the giving results, SVM returns classification values whereas SVR returns continuous values and acknowledges the presence of non-linearity in the data, It aims to formulate a decision function which is the best approximation of a set of insights, keeping in mind the common ground between a good power of generalization and stabilized behavior.[11]

Since finance data is large and hard to get, we have used open-source data from Yahoo finance and fetched the stock prices for the last 60 days using the yfinance Python library. We used the commonly used library Plotly to plot the data and analyze the key data points to select the optimum features for training the model. We selected five variables, Open, High, Low, Close and Volume then split the dataset into 9:1 ratio for training and testing respectively.

Training a model that can give the best results and performance is quite a time-taking task, thus, we used grid search to stipulate values for hyperparameters. Grid search is the process of performing hyperparameter tuning in order to determine the optimal values for a given model. This is significant as the performance of the entire model is based on the hyperparameter values specified.

Then, we trained the SVR model with the training data and grid search which tests your model's performance on each iteration by using metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) on the training data. It selects the best parameters that result in the highest performance model. Even though the model is most optimized, usually it fails to perform as expected in real-time data therefore we tested our models using similar training metrics on the testing data, which provides information about a model performance similar to real-world data.

Conclusion

In this paper, we propose a framework for stock market prediction using support vector regression. We built, fine-tuned, and tested this model using the yahoo finance data on stock market history. Predicting the movement of the stock market is a challenging task due to consistently changing stock values which are dependent on multiple parameters which form complex patterns.

The raw data was suitably cleaned, the appropriate variable transformation was done, and the predictors and the response variable were identified for building the classification and regression-based predictive models. We find that support vector regression gives reasonable results with decent accuracy.

References

I. Parmar et al., "Stock Market Prediction Using Machine Learning," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018.

A. H. Bukhari, M. A. Z. Raja, M. Sulaiman, S. Islam, M. Shoaib and P. Kumam, "Fractional Neuro-Sequential ARFIMA-LSTM for Financial Market Forecasting," in IEEE Access (2020).

Kang Zhang, Guoqiang Zhong, Junyu Dong, Shengke Wang, Yong Wang, Stock Market Prediction Based on Generative Adversarial Network(2019)

Umer, M., Awais, M., & Muzammul, M. (2019). Stock Market Prediction Using Machine Learning(ML)Algorithms. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence

Adil Moghar, Mhamed Hamiche, Stock Market Prediction Using LSTM Recurrent Neural Network,Procedia Computer Science (2020)

Gurav, U., Sidnal, N. (2018). Predict Stock Market Behavior: Role of Machine Learning Algorithms. In: Bhalla, S., Bhateja, V., Chandavale, A., Hiwale, A., Satapathy, S. (eds) Intelligent Computing and Information and Communication. Advances in Intelligent Systems and Computing.

T. Mankar, T. Hotchandani, M. Madhwani, A. Chidrawar and C. S. Lifna, "Stock Market Prediction based on Social Sentiments using Machine Learning," 2018 International Conference on Smart City and Emerging Technology (ICSCET) (2018)

Pathak, A., Shetty, N.P. (2019). Indian Stock Market Prediction Using Machine Learning and Sentiment Analysis. In: Behera, H., Nayak, J., Naik, B., Abraham, A. (eds) Computational Intelligence in Data Mining. Advances in Intelligent Systems and Computing

Dash, R.K., Nguyen, T.N., Cengiz, K. et al. Fine-tuned support vector regression model for stock predictions. Neural Comput & Applic (2021)

Gupta D, Pratama M, Ma Z, Li J, Prasad M (2019) Financial time series forecasting using twin support vector regression.

Yaohao Peng, Pedro Henrique Melo Albuquerque, Jader Martins Camboim de Sá, Ana Julia Akaishi Padula, Mariana Rosa Montenegro, The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with Support Vector Regression, Expert Systems with Applications.(2018)

X. Yuan, J. Yuan, T. Jiang and Q. U. Ain,(2020) "Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market".

Nti, I.K., Adekoya, A.F. & Weyori, B.A. A comprehensive evaluation of ensemble learning for stock-market prediction. J Big Data 7, 20 (2020)

Sidra Mehtab, Jaydip sen (2020) -"A Time Series Analysis-Based Stock Price Prediction Using Machine Learning and Deep Learning Models". International Journal of Business Forecasting and Marketing Intelligence .

A. Sharma, D. Bhuriya and U. Singh, "Survey of stock market prediction using machine learning approach," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), 2017.

Das, S., Sahu, T.P., Janghel, R.R. et al. Effective forecasting of stock market price by using extreme learning machine optimized by PSO-based group oriented crow search algorithm. Neural Comput & Applic (2022).

Deepak Kumar, Pradeepta Kumar Sarangi, Rajit Verma, A systematic review of stock market prediction using machine learning and statistical techniques. Materials Today: Proceedings, 2022.

Weiwei Jiang, Applications of deep learning in stock market prediction: Recent progress, Expert Systems with Applications, 2021.

Wu, J.MT., Li, Z., Herencsar, N. et al. A graph-based CNN-LSTM stock price prediction algorithm with leading indicators. Multimedia Systems (2021).

Karimuzzaman, M., Islam, N., Afroz, S. et al. Predicting Stock Market Price of Bangladesh: A Comparative Study of Linear Classification Models. Ann. Data. Sci. (2021).