IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

STOCK PRICE PREDICTION AND FORECAST USING HYPER PARAMETER TUNED MACHINE LEARNING ALGORITHMS

Mansha Rapria

Amity International School,
Power Grid Complex, Sector 43, Gurugram
Haryana, (122001) India

Abstract - ***Stock market is the direct indication of the economy of the nation, the financial market derived by many factors such as earning, interest rate, consumer spending and more. Stock market involves holdings from promoters, financial institutions and retail investors. Stock price prediction is challenging due to its volatility nature. Thus this needs a highly computational intelligence system. Nowadays, artificial intelligence proving its computational efficacy in various domain, financial domains will benefit through the Machine learning (ML) and deep learning (DL) techniques. The proposed work is stock price prediction based on machine learning models. The proposed work also projects the stock price for upcoming days, this forecast is based on machine learning models. The proposed work used regression analysis as the dependent value is continuous in nature. The algorithms implemented are Decision Tree, Random Forest and K-Nearest Neighbor models and these are implemented as regression models. To improve the model and make the stock price predictions more accurate, the algorithm is hyper parameter tuned with the given search space. Grid search cross validation (GSCV) technique is used for validating the dataset with the given search space for finding the best fit parameter. Experimental results show that Random forest predicted the stock price with minimum MSE loss.***

Key Words: Artificial Intelligence, Machine Learning, Deep learning, Decision Tree, Random forest, Regression analysis, Grid Search Cross Validation

1.INTRODUCTION

Development of the economy has given great opportunities for financial markets, stock markets have highly impacted and went through high growth from the last century. Capital market is also expanding with speed, which also highly impacts stock market growth. Not only are the companies beneficial with financing by listing stock, the investors also benefiting and improving their quality of life. Due to the growth of internet and online services of stock, the retail investors are also growing in pace every year. The availability of the data influencing the stock price prediction, the limitations in accessing the data is impacting the price predictions. Apart from data availability, the major factor influences the price prediction is volatility of prices, the stock becomes more volatile when it is driven by news, events, company financial results, economic crisis, economic growth and more. Thus price prediction is a highly challenging task for financial institutions and retail investors.

There are traditional stock price prediction methods using statistical model, moving average, relative strength indicator, Fibonacci techniques, pivot point calculation and more tools are available to indicate the trend and price predictions. However, these techniques fail more times, as they are hard to understand and highly computational involved and not simple to use for users. These tools work on historic prices of stocks. Historic price movements only cannot be considered for price predictions as there are other factors like volume traded, news flows will affect the prices.

Retail traders or institutions look for short-term price movements for trading, whereas investors look for very long term holding of stocks. Short term price prediction is the problem and it's more challenging. This work predicts the price and forecasts the price, which is more suitable for short term traders and investors. The existing methods are more complex to use and understand, moreover the accuracy of the system is very less. Thus stock price prediction needs an effective model. There are many factors drives the selection of technique, the proposed work used regression analysis as the stock prices are continuous values, this technique makes predictions simple and interpret the price predictions, this technique also finds the feature importance automatically while training, as well as regression models can effectively interpolate the know data points and extrapolate unobserved ranges. As the stock price is a time series dataset, the historical patterns are identified by the regression models easily. Stock price projections are also implemented with a regression algorithm, this projection for short term like 5 to days can be done with ML models.

Long term investors rely on financial performance of the company like quarterly results and price earnings ratio, company debt on equity and more to choose their preferable

stocks. Capital asset price model is the one used for selecting the stock for long term investments.

The main contributions of this paper are

- To implement stock price prediction based on regression ML models and study their performance.
- Define the search space for cross validation to build a best fit parameter for the proposed regression models.
- Implement DT, RF and KNN regression models to compare the result to find the better model for stock price prediction.
- To implement and analyse the price forecast to help the financial institutions and retail investors for finding very short term projections.

This paper is organized as Chapter 2 provides literature survey and related work in stock price prediction. Chapter 3 focuses on proposed models and the novelty of the proposed work. Chapter 4 gives detailed insight on results and analysis derived from implemented models. Chapter 5 draws conclusions based on this work and further enhancement possibilities.

2. RELATED WORK

Existing studies on stock price prediction through mathematical techniques and artificial intelligence models are vastly studied. Machine learning and deep learning models are highly researched due to its scalability and performance. Existing models analyzed stock trend prediction based on social media post arriving sentiment analysis, however, this can approach trends and price cannot be arrived at using these models. Thus there is a need for a conventional method, which can be practiced without high computation by the user and with high accuracy prediction is a challenging problem, is approached in this paper. This work also aimed at projecting the stock price for the short term, which helps retail traders, investors and institutions for considering the stock positions.

Genetic algorithm, one of the feature selection techniques, is used by Chen et al. in [1], this paper, they trained the China Construction bank dataset with Long Short Term Memory (LSTM) model. R squared co-efficient is used as a fitness function for GA. The larger the R-squared value, highly able to define X to Y. LSTM is one of the deep learning models for supporting time series analysis implemented. Experimental results showed that GA and LSTM model has achieved less MSE value. There are other feature selection based models studied including Yual et al.[2] feature selection based on Recursive feature elimination and Support Vector Machine (SVM) and Random forest models were used, this results showed that the RF has achieved high accuracy of 52.7%.

Sentiment analysis from social media posts has gained importance as the social media content highly influences stock markets. The posts are analyzed for positive, negative or neutral sentiment, accordingly the price predictions are followed. The work [3] represented deep learning based sentiment analysis for stock price prediction. This work implemented Sparrow Search Algorithm (SSA) and LSTM model with multi source data with sentiment included. Experimental results showed that the less MSE value arrived with the SSA - LSTM model. Some existing work reviewed different feature extraction and reduction models [6]. The work [4] proposed textual feature extraction through different techniques n-gram, lagged technique from twitter data. The other feature selection techniques like RFE and feature reduction Principal component analysis (PCA) were used. Experimental results showed RFE with Random forest has achieved the highest accuracy of 0.58.

Crypto currency price prediction is also more volatile in nature studied with the help of sentiment analysis in [5] for Litecoin and Bitcoin were analyzed. Along with price, sentiment data also given to the training model with VADER sentiment, twitter data is extracted for this sentiment analysis. Experimental results showed that MAE loss is 0.087 by combining daily prices and sentiment to the model.

The literature survey on stock price prediction using various techniques is studied in this chapter. It is inferred from this study that, machine learning and deep learning techniques are extensively used as its performance is incompatible with other traditional techniques. The existing ML and DL models with various algorithms and feature selection and reduction techniques have been adopted. Sentiment analysis based work is also used for price prediction, this is not reliable most of the time other than news based or events based sentiment. Thus proposed work focused on optimizing the machine learning model with hyper parameter tuning as it can increase the performance of the ML models.

3. PROPOSED WORK

In this chapter, the proposed work, stock price prediction using regression ML models are discussed. The work proposed implementation of Decision Tree, Random forest and K-Nearest Neighbor algorithms for stock price prediction. To improve the novelty and to get an accurate price prediction model, we proposed hyper parameter tuning on training parameters. Cross validation technique is used, which takes training set and train and performs validation, gets the error metrics. Finally the cross validation finds the best fit parameter based on the low loss values.

DATASET DETAILS

Dataset used for this study is live and real time data extracted from Yahoo finance API. The API gets a stock symbol and starting date, end date as input to collect the data. Dataset can be collected from any duration, for this work we extracted 23 years data for training. The price values of stock like open, high, low, close price and volume traded are attributes used for this work. The data can be extracted for any NSE stock, thus users can go with their choice of stock picks.

The following table represents dataset attributes extracted from yahoo finance.

Names	Attribute Details
Date	Trading Date
Symbol	Listed stock symbol in NSE
Series	Equity or Currency
PrevClose	Stock's previous date close price
Open	Stock's current date open price
High	Stock's current date high price
Low	Stock's current date low price
Close	Stock's current date close price
Volume	Current date volumes traded on stock
VWAP	Current date Volume weighted average price

Table 1: Extracted stock data from Yahoo finance

METHODOLOGY

This proposed work implemented Decision Tree regressor, Random forest regressor and K-NN regressor models for stock price prediction. The dataset is collected real time till date data, the user can give the stock symbol to extract the data with API from Yahoo finance. Any stock data can be extracted and trained, predicted and forecasted. There are more than 7500 stocks listed in NSE. The extracted data is stored in comma separated values in a file and plotted using matplotlib for visualizing the trend of stock for the past 23 years. The data collection date start from January 2000. The following figure shows the system architecture of the proposed work.

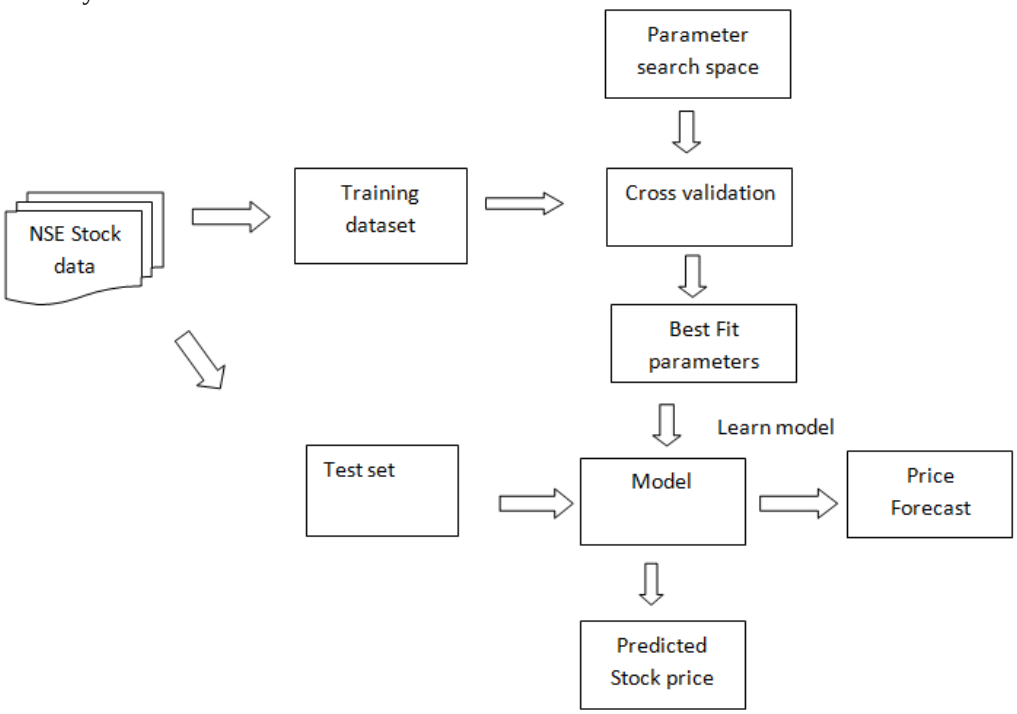

Figure 1: System architecture of proposed Stok prediction using Hyper tuned ML models

Exploratory Data Analysis (EDA) is performed on the dataset by the data scientists to understand better on the data patterns. This work projects the stock price for 20 years data as a single visualization plot with 1000 ticks shows how the stock behaved for the past two decades. This gives insight about the stock performance. The following plot shows the stock price symbol INFY price movement for 20 years. The stock can be any National Stock Exchange (NSE) stock.

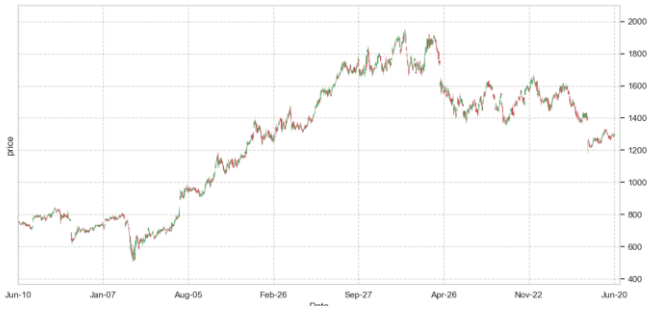

Figure 2: Stock Price Visualization for stock-INFY

HYPER PARAMETER TUNING

Hyper parameter tuning is performed for the machine learning or deep learning models to choose the best parameter from the search space. There are a number of parameters used for machine learning algorithms and these parameters vary for each algorithm. For example Decision Tree uses Splitter, max_depth and minimum sample leafs are the parameters, whereas KNN uses number of neighbors and weights. Thus the parameter given should be optimum value so that the learning and prediction is more accurate. One of the most important hyper parameter tuning models is Grid Search Cross Validation (GSCV). The cross validation is performing the validation on a dataset to arrive the evaluation metric and choose the best one according to the low loss or high accuracy. The following table shows the parameter tuned in each algorithm and its search space cross validated.

Algorithm	Parameter tuned	Search Space value
Decision Tree	Splitter	Best split, random split
	Max depth	1,3,5,7,9
	Samplea leaf	2,4,6,8
	Max features	auto,log2,sqrt
Random forest	Depth	2,4,6
	Number of estimators	64,128,256
KNN	Number of neighbors	2,3,4,5
	weights	uniform, distance

Table 2: Hyper parameter used for the study

Cross validation works to minimize the loss and maximize the accuracy of the system, thus controlling the learning model by choosing the best parameters. There are minimum two parameters used for tuning in each machine learning models as stated in the above table. Grid search cross validation works based on brute force is more effective in minimizing the error and find best parameters.

Grid Search Cross Validation (GSCV)

The prediction error is computed by the equation given in (1), where 'err' is the prediction error, 'm' is the mean and 'b' the

number of folds cross validation taken. The minimum the error value, the effective model.

$$m = \frac{\sum_{i=1}^{b} err_i}{b} \quad \text{--(1)}$$

Mean squared error (MSE) loss is used as scoring function for cross validation, the MSE loss is computed for each search space on the given parameter on the test set. The other metrics which can be used for cross validation scoring is Mean absolute error (MAE) and accuracy. The below equation shows the MAE computation.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |x_i - x| \quad \text{--- (2)}$$

Where xi, x is the absolute error and 'n' is the total number of observations.

4. RESULTS AND DISCUSSIONS

The proposed stock price prediction through machine learning regression models used a 23 years dataset from 2000 to till date. The training data 80% and test data 20% is split and evaluated. The hyper parameter tuning is implemented for all ML models DT, RF and KNN. The number of search space given on each parameter is minimum two. The validation data is cross validated with the scoring function mean squared error and computed the best fit parameter. The hyper tuned best fit model is trained and predicted the stock prices. The results are discussed in this section. Experimental results are computed for stock prediction with error metrics MAE, MSE, R2 score and RSME error.

The below plot shows the five days price forecast for INFY stock by decision tree regression model. The price projections can be computed for customized number of days, the short the number of days, the accuracy is high.

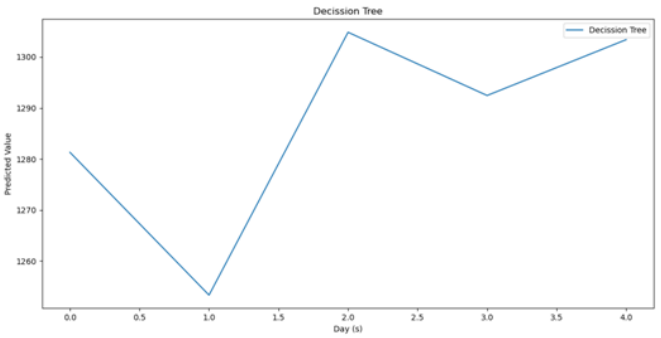

Figure 3: Price projection by Decision Tree for next five days for Stock-INFY

The following plot shows the price projection done for INFY stock by Random forest regression model. It is also observed from forecast, that the DT and RF given highly similar price projections.

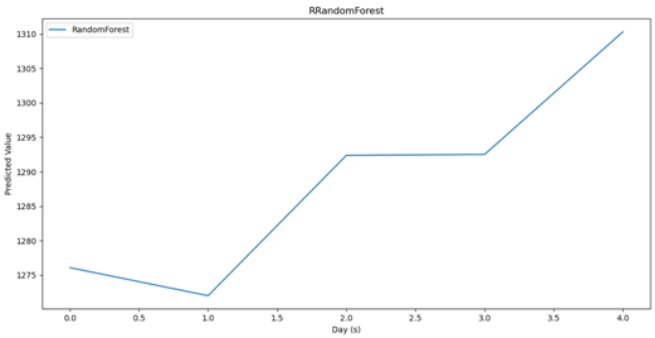

Figure 4: Price projection by Random Forest for next five days for Stock-INFY

The below plot represents the evaluation metric scores computed by traditional and hyper parameter tuned Decision Tree (HT-DT) regression models are compared. Blue colour indicates the normal model and orange colour indicates hyper tuned model.

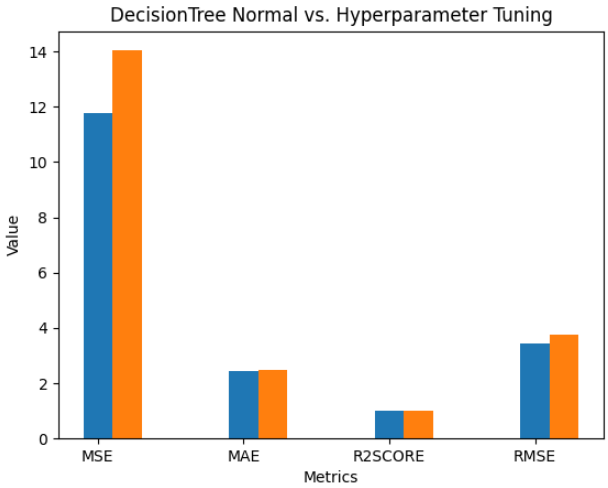

Figure 5: Normal and hyper tuned DT comparison

The below plot represents evaluation metrics of normal and hyper tuned KNN (HT-KNN) regression model.

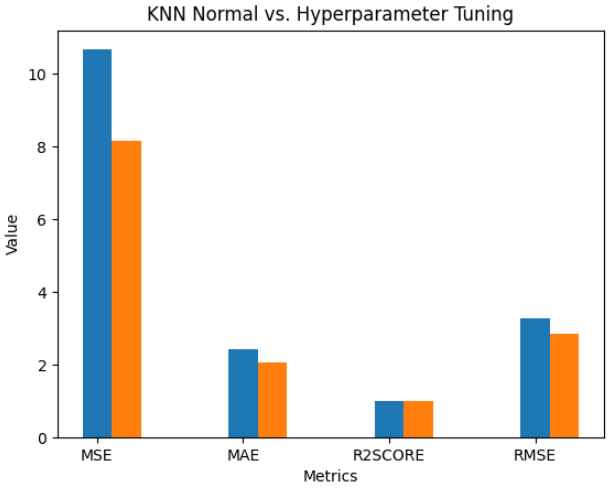

Figure 6: Normal and hyper tuned DT comparison

The below plot represents evaluation metrics of normal and hyper tuned Random forest (HT-RF) regression model.

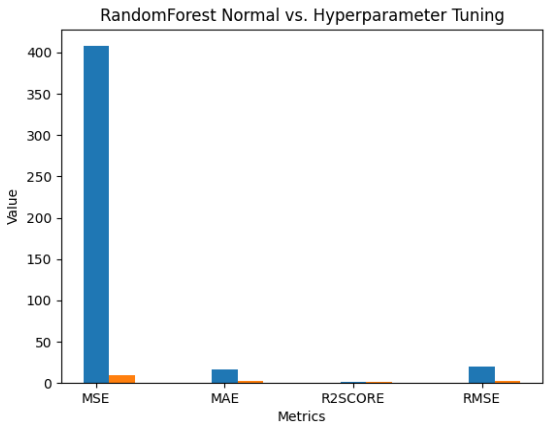

Figure 7: Normal and hyper tuned DT comparison

ML models	MSE	MAE
HT-DT	14.62	2.15
HT-RF	12.32	2.12
HT-KNN	9.22	2.02

Table 3: Error metrics computed for stock price prediction by hyper tuned ML models

The error metric MAE, MSE are computed for proposed hyper tuned ML models, the table shows the results arrived, it is observed that the error value is less in hyper tuned K-nearest neighbor algorithm. The mean squared error is 10.27, which is lowest comparing decision tree and random forest. Similarly, mean absolute error is 2.02, which is lowest comparing other two models.

5. CONCLUSIONS

Stock prices are highly volatile and need challenging computation techniques to predict the prices. Though there are existing techniques on statistical models and machine learning models are available, the accuracy of price prediction is less. Thus to predict the stock prices and forecast the prices with less MSE losses, the proposed system used hyper tuning parameters to built the ML models. The regression models are build with best fit parameters, the proposed method optimizes and improves the regression models. The search space containing minimum four parameters is considered for cross validation. Experimental studies showed that the Hyper parameter tuned K-Nearest Neighbor Regression model has achieved the lowest MAE and MSE error values.

As the future enhancement, the price prediction can be extended for crypto-currencies also, as they are emerging with new technology called block-chain. The stock price prediction can be performed with deep learning models or transfer learning models. To include the volatility nature, the system can be built for index funds and transfer the learning for stock price predictions.

REFERENCES

[1] S. Chen and C. Zhou, "Stock Prediction Based on Genetic Algorithm Feature Selection and Long Short-Term Memory Neural Network," in IEEE Access, vol. 9, pp. 9066-9072, 2021, doi: 10.1109/ACCESS.2020.3047109.

[2] X. Yuan, J. Yuan, T. Jiang and Q. U. Ain, "Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market," in IEEE Access, vol. 8, pp. 22672-22685, 2020, doi: 10.1109/ACCESS.2020.2969293.

[3] G. Mu, N. Gao, Y. Wang and L. Dai, "A Stock Price Prediction Model Based on Investor Sentiment and Optimized Deep Learning," in IEEE Access, vol. 11, pp. 51353-51367, 2023, doi: 10.1109/ACCESS.2023.3278790.

[4] S. Bouktif, A. Fiaz and M. Awad, "Augmented Textual Features-Based Stock Market Prediction," in IEEE Access, vol. 8, pp. 40269-40282, 2020, doi: 10.1109/ACCESS.2020.2976725.

[5] R. Parekh et al., "DL-GuesS: Deep Learning and Sentiment Analysis-Based Cryptocurrency Price Prediction," in IEEE Access, vol. 10, pp. 35398-35409, 2022, doi: 10.1109/ACCESS.2022.3163305.

[6] F. G. D. C. Ferreira, A. H. Gandomi and R. T. N. Cardoso, "Artificial Intelligence Applied to Stock Market Trading: A Review," in IEEE Access, vol. 9, pp. 30898-30917, 2021, doi: 10.1109/ACCESS.2021.3058133.

[7] P. K. Aithal, U. D. Acharya, M. Geetha and P. Menon, "Building a Calendar of Events Database by Analyzing Financial Spikes," in IEEE Access, vol. 9, pp. 114192-114206, 2021, doi: 10.1109/ACCESS.2021.3102495.

[8] X. Li, H. Xie, R. Y. K. Lau, T. -L. Wong and F. -L. Wang, "Stock Prediction via Sentimental Transfer Learning," in IEEE Access, vol. 6, pp. 73110-73118, 2018, doi: 10.1109/ACCESS.2018.2881689.

[9] T. Kabbani and E. Duman, "Deep Reinforcement Learning Approach for Trading Automation in the Stock Market," in IEEE Access, vol. 10, pp. 93564-93574, 2022, doi: 10.1109/ACCESS.2022.3203697.

[10] C. -H. Chen, P. Shih, G. Srivastava, S. -T. Hung and J. C. -W. Lin, "Evolutionary Trading Signal Prediction Model Optimization Based on Chinese News and Technical Indicators in the Internet of Things," in IEEE Internet of Things Journal, vol. 10, no. 3, pp. 2162-2173, 1 Feb.1, 2023, doi: 10.1109/JIOT.2021.3085714.

[11] M. R. Hossain, M. T. Ismail and S. A. B. A. Karim, "Improving Stock Price Prediction Using Combining Forecasts Methods," in IEEE Access, vol. 9, pp. 132319-132328, 2021, doi: 10.1109/ACCESS.2021.3114809.

[12] C. -H. Chen, P. Shih, G. Srivastava, S. -T. Hung and J. C. -W. Lin, "Evolutionary Trading Signal Prediction Model Optimization Based on Chinese News and Technical Indicators in the Internet of Things," in IEEE Internet of Things Journal, vol. 10, no. 3, pp. 2162-2173, 1 Feb.1, 2023, doi: 10.1109/JIOT.2021.3085714.

[13] Z. -R. Lai, P. -Y. Yang, L. Fang and X. Wu, "Reweighted Price Relative Tracking System for Automatic Portfolio Optimization," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 50, no. 11, pp. 4349-4361, Nov. 2020, doi: 10.1109/TSMC.2018.2852651.

[14] Y. Li, S. Wang, Y. Wei and Q. Zhu, "A New Hybrid VMD-ICSS-BiGRU Approach for Gold Futures Price Forecasting and Algorithmic Trading," in IEEE Transactions on Computational Social Systems, vol. 8, no. 6, pp. 1357-1368, Dec. 2021, doi: 10.1109/TCSS.2021.3084847.