2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE) | 979-8-3503-1156-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/WIECON-ECE57977.2022.10150983

Prediction of Stock Price Direction with Trading Indicators using Machine Learning Techniques

Md. Siam Ansary
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
siamansary.aust@gmail.com, siamansary.cse@aust.edu

Amina Brinto
Dhaka Community Medical College
Dhaka, Bangladesh
{aminabrinto, brintoamina}@gmail.com

***Abstract*—Stock market provides a platform where people can engage in trading. Such tradings contribute to the growth of economy. It is a very intriguing research topic to predict the direction of a stock price i.e. if a stock is neutral or going upwards or downwards. If such prediction can be done efficiently, people can invest more systematically. In this research, we have addressed this issue using different trading indicators as features. Different classification models such as K Nearest Neighbor, Decision Tree, Random Forest, Support Vector Machine, AdaBoost, Multilayer Perceptron have been experimented with and they have yielded excellent results.**

***Index Terms*—machine learning, stock, classification**

I. Introduction

Working with Stock Exchange data can be very challenging since the prices of stocks are continuously changing. A Stock Exchange is always going through complex situations. It is hardly ever easy to assume the direction of the prices of stocks. Investors and companies are constantly taking risks while buying and selling stocks but this is what keeps the economy growing. The fluid nature of the Stock Exchange helps to reform a country's economy and finances and become prosperous. It is a grueling task to invest in stocks with the hopes of minimal risk and loss and maximal profit and gain. If it is not done properly, any investor can experience a huge loss within a short amount of time.

Machine Learning (ML) techniques can do predictions very efficiently with sufficient data. If an appropriate model is trained with proper data, it can predict if a stock is heading upwards or downwards very efficiently. But, for this purpose, the right features need to be used. Otherwise, the model can become ineffective. Trading indicators are mathematical tools that can help to understand the direction of a stock. Such indicators can be very powerful features for an ML model. In our experiment, we have tried to predict the direction of stock price for investment decision making utilizing different trading indicators. Since Stock Market data can be very volatile, prediction approaches should be able to adapt to such quick changes. ML models are more suitable for such tasks compared to any other techniques for accommodating such situations.

In the subsequent sections, we go through the existing works related to our experiment, our experimental approach and its performance.

II. Related Works

There have been various works regarding stocks. Different researchers have focused on different aspects. It is necessary to observe them to decide upon a methodology for improved predictive approach.

Chen and Hao [1] worked with Shanghai Stock Exchange Composite Index and Shenzhen Stock Exchange Component Index. Their goal was to analyze the Chinese Stock Market. They noted that if a stock's closing price of a day is more than the closing value of a previous day, profit is gained; otherwise loss occurs. They labeled the dataset with this concept. As for features, different trading indicators were used such as Moving Average, Exponential Moving Average, Moving Average Convergence Divergence, Volume Ratio, Relative Strength Index, On Balance Volume, Momentum Index, AR, BR. As for machine learning models, feature weighted Support Vector Machine and K Nearest Neighbor were used. Anbalagan and Maheswari [2] worked on Bombay Stock Exchange data for short term investment decisions. Fuzzy Metagraph approach has been experimented with in the work. Simple Moving Average, Exponential Moving Average, Moving Average Convergence Divergence, Relative Strength Index were used for the analysis. He et al. [3] worked on stock price movement prediction of pharmaceutical companies during the COVID 19 period. They emphasized that while it can be difficult to work with stock data, hand crafted features can help in the predictive research work. They had collected data from multiple sources; The Yahoo! Finance's API [10] was used and also data from Our World in Data [11] was utilized. They noted that use of handcrafted features can help achieve results with greater accuracy. Ballings et al. [4] tried to predict stock price direction with the data that was gathered from publicly listed European Companies. Their main focus of research was to see between single learners and ensemble learners, which type of algorithm performs the best. The techniques used in this research were Logistic Regression, Neural Networks, K Nearest Neighbor, Support Vector Machine, Random Forest, AdaBoost and Kernel Factory. Out of the techniques, Random Forest

979-8-3503-1156-3/22/$31.00 ©2022 IEEE

performed the best. They had used five times two fold for cross validation. Different financial indicators were used in this experiment as features such as liquidity indicators, solvency indicators, profitability indicators. Kara et al. [5] experimented with Artificial Neural Networks and Support Vector Machine for predicting direction of stock price index. In the research, data of the Istanbul Stock Exchange was used. Artificial Neural Networks achieved accuracy of 75.74% whereas Support Vector Machine gained 71.52% of accuracy. Different indicators such as Simple Moving Average, Weighted Moving Average, Momentum, Stochastic K%, Stochastic D%, Relative Strength Index, Moving Average Convergence Divergence, Larry William's R%, Accumulation Distribution Oscillator, Commodity Channel Index were used. Basak et al. [6] tried predicting stock market prices' direction with Random Forest and Gradient Boosted Decision Trees. In the work, Relative Strength Index, Stochastic Oscillator, Williams Percentage Range, Moving Average Convergence Divergence, Price Rate of Change, On Balance Volume were used as features for the ML models. Parmar et al. [7] used Open, High, Low, Close and Volume values to make prediction of stock prices with the Yahoo Finance dataset. 20% of the dataset was used for testing while 80% was made use of for training models. Naik and Mohan [8] worked with data from the National Stock Exchange of India for stock price movement classification. Indicators such as Simple Moving Average, Exponential Moving Average, Momentum Indicator, Stochastic Oscillator, Moving Average Convergence Divergence, Relative Strength Index, William R, Accumulation Distribution Index, Commodity Channel Index were used as features.

III. Methodology

In this work, we have predicted if a stock price is moving upwards, downwards or if it is neutral. Features have been extracted, stocks have been labeled, dataset has been balanced, unimportant features have been identified and eliminated and then, ML models have been used. The steps of the experiment are illustrated in Fig 3.

A. Dataset Preparation

1) Primary Dataset: Firstly, stock data has been collected using Yahoo! Finance's API [10] . This collected data has the Open, High, Low, Close, Volume and Adj Close values of each of the stocks. Initially, we have collected 5630 samples. The samples are of the period from 01 January 2000 to 17 May 2022.

2) Trading Indicators Calculation: We have extracted important features from the dataset. Using Technical Analysis Library [9], based on the Open, High, Low, Close and Volume values of each stock, different trading indicators have been calculated so that they can be used as features. The calculated indicators are as below.

- Moving Average (MA)
- Balance of Power (BOP)
- Money Flow Index (MFI)
- Momentum (MOM)
- Rate of Change (ROC)
- Relative Strength Index (RSI)
- Weighted Close Price (WCP)
- Average Price (AP)
- Median Price (MP)
- Typical Price (TP)
- On Balance Volume (OBV)

3) Labelling of the Dataset: With the MFI and RSI values of each stock, the stocks have been labeled. If the value of MFI is greater than 80 or the value of RSI is greater than 70, a stock is overbought. Overbought condition signifies the market is upwards. If the value of the MFI is less than 20 or the value of RSI is less than 30, a stock is oversold. Oversold signifies that at the current moment, the stock is going downwards. The stocks that are neither overbought nor oversold, are neutral. In such a manner, all the stocks have been labeled.

4) Balancing the Dataset: After labelling the stocks as upwards, downwards or neutral, we have observed 5033 stocks to be neutral, 426 stocks to be of upwards direction and 171 records to be of downwards direction. Hence, for classification task, the dataset is not balanced. Hence, at first, we have used Random Undersampling on Neutral Stocks so that number of neutral records come down to 426 and then, applied Random Oversampling so that number of downwards stocks become 426. The total number of records of the balanced dataset becomes 1278 where each class has same amount of records.

B. Feature Elimination

After making sure the dataset is balanced, we have tried to analyze if all the features of our dataset are important because unimportant features can increase the complexity of the problem and can cause overfitting issue. Our dataset has 17 features, 6 of which are Open, High, Low, Close, Adj Close and Volume from preliminary dataset and the rest 11 features are trading indicators which have been calculated. Along with 17 features, there is 1 target column. We have applied the Pearson Correlation approach on the dataset. The heatmap of the values is in Figure 1. The threshold value has been set to 0.8 for dropping unnecessary features. The features with correlation value above 0.8 have been dropped. On the existing columns, we have applied Mutual Information Gain approach. The values are shown in Figure 2. The threshold value has been set to 0.5 to further drop insignificant features, implying the features whose information gain has been under 0.5, have been eliminated.

C. Application of Machine Learning Models

After feature elimination, we have applied Standard Scaling on the remaining features. Standard Scaling has been done to bring all features to a similar value range because if different

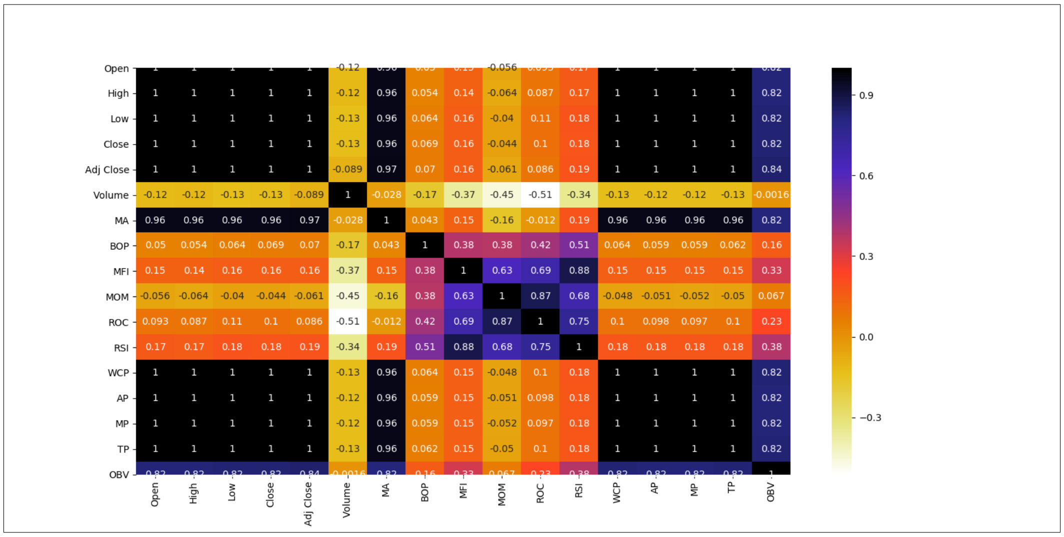

Fig. 1. Pearson Correlation Heatmap of Features

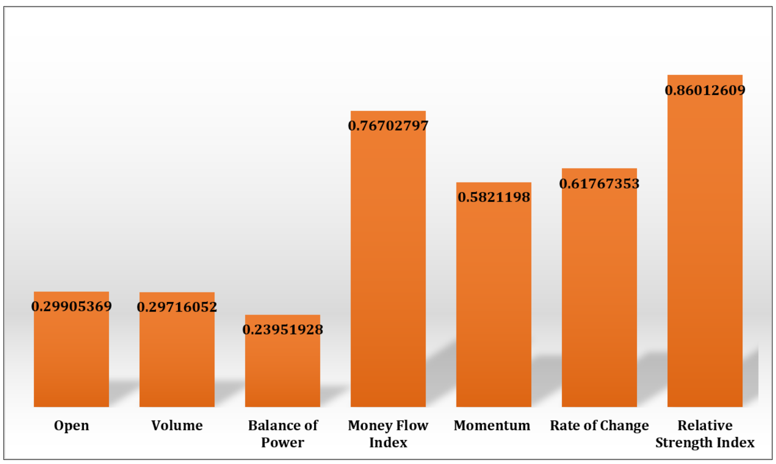

Fig. 2. Mutual Information Gain from Features after Features' Removal through Pearson Correlation Approach

features are of different ranges, unwanted bias can occur. Then, we have split our dataset in two portions. 80% of the dataset has been used for trainging and the rest 20% has been used for testing purpose. We have applied six classification models. They are as followed.

- K Nearest Neighbor Classifier
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine Classifier
- AdaBoost Classifier
- Multilayer Perceptron Classifier

IV. Performance Analysis

The performances of the machine learning models in predicting the direction of stock price are illustrated in Fig 3. The accuracy, precision, recall and f1 score values are displayed. All the models have performed well. Among the models, Random Forest approach has performed the best.

V. Conclusion and Future Works

In our research, the Machine Learning Classifier models have yielded very good results in predicting the direction of prices of stocks. Using trading indicators as features, labelling of stocks based on MFI and RSI, elimination of unnecessary features have contributed to this excellent outcome. In future, we want to experiment with more datasets and classifiers.

References

[1] Chen, Yingjun, and Yongtao Hao. "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction." *Expert Systems with Applications* 80 (2017): 340-355.

[2] Anbalagan, Thirunavukarasu, and S. Uma Maheswari. "Classification and prediction of stock market index based on fuzzy metagraph." *Procedia Computer Science* 47 (2015): 214-221.

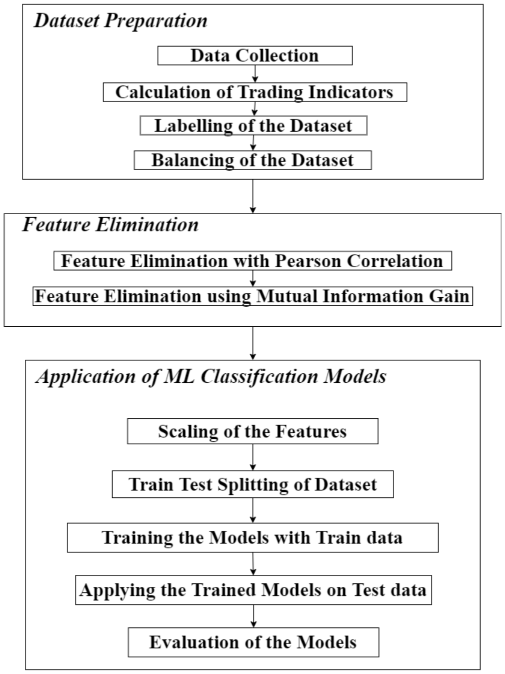

Fig. 3. Steps of the Experimental Approach

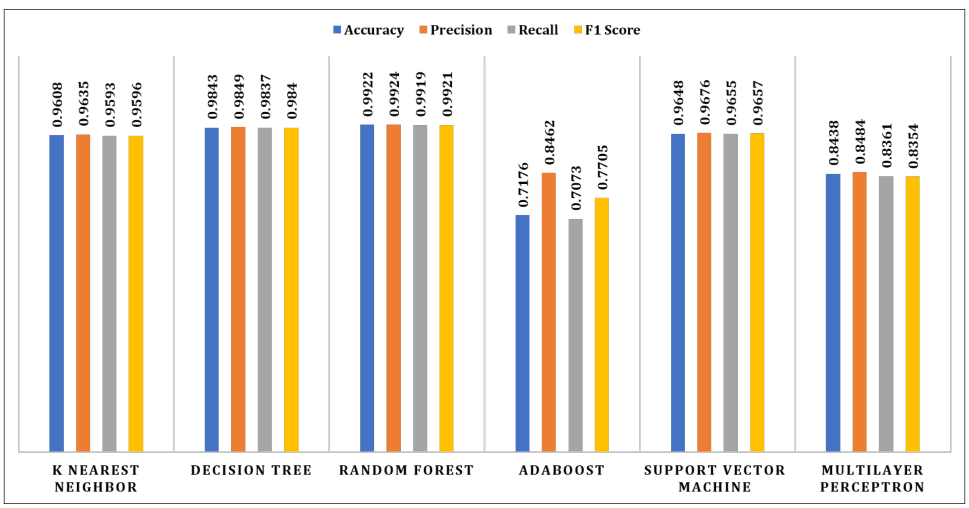

Fig. 4. Performances of the Classification Models

[3] He, Beilei, Weiyi Han, and Suet Ying Isabelle Hon. "A Machine Learning Approach: Enhancing the Predictive Performance of Pharmaceutical Stock Price Movement during COVID." *Journal of Data Analysis and Information Processing* 10, no. 1 (2021): 1-21.

[4] Ballings, Michel, Dirk Van den Poel, Nathalie Hespeels, and Ruben Gryp. "Evaluating multiple classifiers for stock price direction prediction." *Expert systems with Applications* 42, no. 20 (2015): 7046-7056.

[5] Kara, Yakup, Melek Acar Boyacioglu, and Ömer Kaan Baykan. "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange." *Expert systems with Applications* 38, no. 5 (2011): 5311-5319.

[6] Basak, Suryoday, Saibal Kar, Snehanshu Saha, Luckyson Khaidem, and Sudeepa Roy Dey. "Predicting the direction of stock market prices using tree-based classifiers." *The North American Journal of Economics and Finance* 47 (2019): 552-567.

[7] Parmar, Ishita, Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta, Himanshu Dhiman, and Lokesh Chouhan. "Stock market prediction using machine learning." In *2018 first international conference on secure cyber computing and communication (ICSCCC)*, pp. 574-576. IEEE, 2018.

[8] Naik, Nagaraj, and Biju R. Mohan. "Stock price movements classification using machine and deep learning techniques-the case study of indian stock market." In *International Conference on Engineering Applications of Neural Networks*, pp. 445-452. Springer, Cham, 2019.

[9] "Technical Analysis Library", [Online] Available: https://mrjbq7.github.io/talib/, Last Accessed on 18 September 2022.

[10] "Yahoo! Finance's API", [Online] Available: https://pypi.org/project/yfinance/, Last Accessed on 18 September 2022.

[11] "Our World in Data", [Online] Available: https://ourworldindata.org/, Last Accessed on 18 September 2022.