Breakout Stocks Identification using Machine Learning Approaches

Md. Siam Ansary

Abstract—Stock market offers a platform for people to engage in trading. It contributes to the growth of nation. Decision making regarding investments needs to be done very carefully so that an investor does not suffer massive loss. Since the share market is susceptible to experience huge change at any given moment, with the probability of profit comes huge risks of losing a fortune. In our research, we have worked on prediction of breakout stocks. If identified properly, it can help one to invest efficiently. We have used multiple machine learning approaches as ML models can offer more effective predictions compared to other methods due to the ability to learn and adapt from dataset information. In our experiment, the models have yielded very good results.

***Keywords*—machine learning, breakout stock, stock market, stock.**

Nomenclature

ML	Machine Learning.
AI	Artificial Intelligence.
RFE	Recursive Feature Elimination.
SVM	Support Vector Machine.
SVR	Support Vector Regression.
KNN	K Nearest Neighbor.
RF	Random Forest.
DT	Decision Tree.
MLP	Multilayer Perceptron.
AdaBoost	Adaptive Boosting.
API	Application Programming Interface.

I. Introduction

Since stock values are always fluctuating, working with stock market data can be exceedingly difficult. Complex circumstances are always present. Predicting the nature of stock prices is rarely simple. While buying and selling stocks, investors and businesses take constant risks, but this is what keeps the economy expanding. The Stock Exchange's dynamic nature aids in the prosperity and economic improvement of a nation. Investing in stocks in the interests of minimizing risk and loss and maximizing gain and profit is a difficult task. Any investment could quickly suffer a significant loss if it is not done correctly.

A breakout stock is one whose price, along by higher volume, varies from a clearly defined support or resistance level. In technical analysis of the stock market, support and resistance are established price levels of a security where it is believed that the price will tend to stop and reverse. Abdullah and Rahaman [1] discussed that based on breakouts, buying and selling decisions should be made. Yu and Li [7] pointed out that the breakout point is a crucial trading point that is more likely to be a high margin opportunity based on the actual experience of human investors.

With enough data, machine learning (ML) approaches may make predictions quite well. When a suitable model is trained on the right data, it may effectively determine if a stock is a breakout candidate or not. The main reason of using ML techniques such as statistical models is that Stock Market is highly volatile and ML can adapt to different scenarios through training better compared to any other method.

The issue regarding investment is that no one wants to suffer any loss. Hence, no decision should be made hastily and all relevant factors should be taken into account. Investments might be quite profitable if a breakout stock can be predicted efficiently. Therefore, we have determined that such a topic requires the use of machine learning techniques to be investigated. Despite reasonable experimentation with other parts of stock markets, we have found that there has been very little work with breakout stocks and there is much scope for progress. Therefore, we have done this work and with classification models, we have achieved satisfactory outcomes.

In the following sections, our research work has been described in details.

II. Related Works

Researchers have done various experiments with stock market data. In different works, distinct aspects have been focused on.

Ballings et al. [2] used information acquired from publicly traded European companies to forecast the direction of stock prices. The primary goal of their research was to determine which kind of algorithm performed better when comparing single learners and ensemble learners. Logistic Regression, Neural Networks, K Nearest Neighbor, Support Vector Machine, Random Forest, AdaBoost, and Kernel Factory were the methods employed in this study. Random Forest outperformed the other methods. For cross validation, they had employed five times two folds. In this experiment, many financial indicators, including those that measure liquidity, solvency, and profitability, were used as features.

Manuscript received November 4, 2022; December 20, 2022.

Md. Siam Ansary is with the Computer Science and Engineering Department, Ahsanullah University of Science and Technology, Dhaka, BANGLADESH (e-mail: siamansary.aust@gmail.com).

Digital Object Identifier (DOI): 10.53907/enpesj.v2i2.173

Data from the National Stock Exchange of India was used by Naik and Mohan [3] to classify stock price movement. As features, indicators such the Relative Strength Index, Simple Moving Average, Exponential Moving Average, Momentum Indicator, Stochastic Oscillator, Moving Average Convergence Divergence, William R, Accumulation Distribution Index, and Commodity Channel Index were employed.

Using different ML models, Islam et al. [4]. predicted the daily closing price of equities. They used data from the Dhaka Stock Exchange. Support Vector Regression and K Nearest Neighbor Regression are the models that were applied in the experiment. SVR performed better at work. In their best outcome, KNN Regression's R Squared Score was 96.39% and Linear SVR's R Squared Score was 97.04%.

Maheswari and Anbalagan [5] worked using data from the Bombay Stock Exchange to make short-term investing selections. The work has tried with the fuzzy graph technique. The analysis made use of the Simple Moving Average, Exponential Moving Average, Moving Average Convergence Divergence, and Relative Strength Index.

Xu et al. [6] performed research on feature selection for trend prediction of stock trading prices. The fluctuations in stock prices can be due to a variety of variables. Thus, it is crucial to pinpoint the most critical attributes when performing machine learning analysis. Two recursive feature elimination techniques, SVM-RFE and RF-RFE, based on Support Vector Machine and Random Forest, respectively, were utilized in the study. Data from the Shanghai Stock Exchange were utilized for the project. The study concluded that while RF and SVM can both forecast trends effectively, SVM may do so more effectively, and RFE may be unneeded for SVM while it is required for RF.

Breakout points were employed by Yu and Li [7] as crucial trading indicators to locate large margin trading opportunities. They talked about how a breakout happens when the trading volume picks up and a stock price breaks through a predetermined resistance level. After the stock price breaks through resistance, human traders typically take a long position. Volatility tends to rise if a stock trade breaks through a price barrier, and prices often move in the direction of a breakout. Breakouts are significant because they serve as the launching pad for significant price moves in the future and, frequently, major price trends. Typically, channel breakthroughs and price breakthroughs, such as triangles, flags, and heads and shoulders, are what cause the most spectacular price movements.

Wen et al. [8] worked in trading decision support system using Support Vector Machine and Oscillation Box where importance was put on breakout point. They discussed that if the strength of buying or selling beat that of the another, the price will effectively break out the upper bound or the lower bound and the time of the breakout is supposed to be the best time to buy or sell the stock. If the current price effective breaks out the upper bound, it means the price will start an uptrend. The same is true for lower bound breakout.

We have observed that there have been definitely lots of works with stock market and machine learning. But, only in a few of them, the concept of breakout stocks have been utilized and we have not found any work where prediction of breakout stocks have been done. With our research, we have addressed this issue.

III. Methodology

Stocks can be classified as potential breakout candidates or not using machine learning techniques. Stock data over time are required for this purpose. A stock might be considered a breakout candidate if consolidation happens for it for an extended length of time and its closing value is either higher or lower than the minimum or maximum closing value from the previous two weeks.

A. Dataset Preparation

1— Data Collection: Firstly, stock data has been collected using Yahoo! Finance's API [9]. This collected data has the Open, High, Low, Close, Volume and Adjusted Close values of each of the stocks. Initially, we have collected 5630 samples which are stock data from 01 January 2000 to 20 May 2022.

2— Calculation of New Features: From the Close values of the stocks, for each stocks we have calculated two more features. They are *Maximum Close Value from Last Two Weeks* and *Minimum Close Value from Last Two Weeks*. These two values are considered as Resistance and Support values.

3— Stock Labelling: The highest and lowest closing prices during the previous two weeks for a stock have been identified as characteristics. The occurrence of consolidation is then verified. Consolidation is stated to have taken place if the lowest closing value from the previous two weeks is more than 98% of the highest closing value from the previous two weeks. If consolidation takes place, a stock is considered a breakout candidate if its closing value is either higher or lower than the highest or lowest closing price during the previous two weeks. In this manner, each stock is then classified as either a breakout candidate or not.

4— Balancing the Dataset: It is observed that the number of breakout candidate stocks is less compared to the stocks that are not breakout candidates. Random under-sampling is done to make the dataset balanced. The balanced dataset has 432 samples.

B. Use of Machine Learning Models

1— Feature Scaling: Machine learning classification models can be used to classify stocks as breakout candidates or not after the dataset has been produced. Scaling has been completed on the prepared dataset before applying the classification models. The features can initially be in different ranges, but the scaling procedure puts all the features in the same range. Such a procedure guarantees the absence of any unwarranted bias. Standardization has been used for scaling purpose.

2— Train Test Split: We have divided the dataset into two portions. 80% of the data is used for training the machine learning models and the rest 20% is used for testing purposes. After training the ML models using the 80% of the dataset, we have applied the models on test data and the results are illustrated through the Figures.

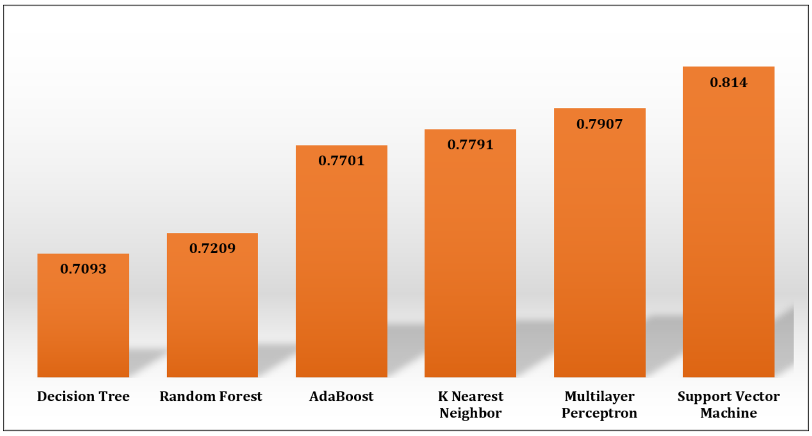

Fig. 1: Accuracy values of the ML Models for Predicting BreakoutStocks

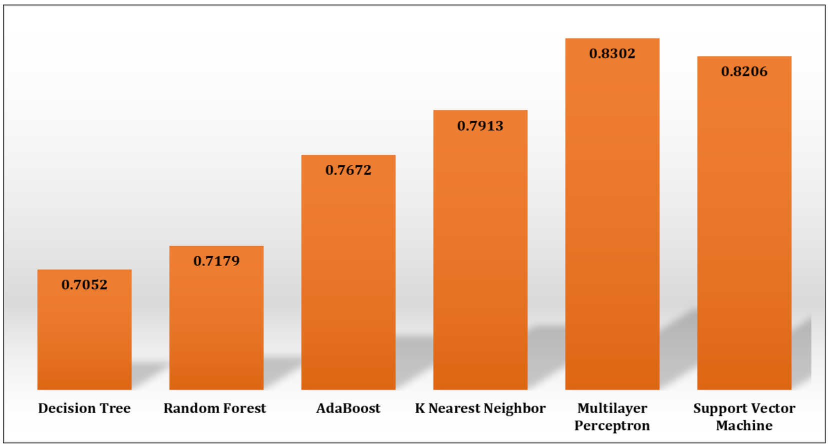

Fig. 2: Precision values of the ML Models for Predicting BreakoutStocks

3— Machine Learning Models: We have used six machine learning models in our work for the classification task. Before applying the models, we have observed the dataset for outliers or null values but prepared dataset had no issues of these kinds. The applied ML models are as the following.

- Decision Tree: A Decision Tree uses a tree-like structure to develop models. It incrementally develops an associated decision tree while segmenting a dataset into smaller and smaller sections. The outcome is a tree containing leaf nodes and decision nodes.

- Random Forest: Random Forest is an ensemble-learning model. The Decision Tree model's shortcomings are sought to be fixed by the Random Forest model. Given that many decision trees are aggregated, the Random Forest technique is far more reliable than the Decision Tree approach.

- Adaptive Boosting: AdaBoost is an ensemble boosting classifier that, by combining weak classifiers, improves their performance. A number of underperforming classifiers are combined through an iterative process to produce a strong classifier with high accuracy. To make effective predictions of unexpected observations, the basic principle is to train the data sample and set the classifier weights in each iteration.

- K Nearest Neighbor: The K Nearest Neighbor is an effective classification technique. KNN maintains all of the examples that are accessible and categorizes additional cases using a similarity metric. A majority vote from its neighbor classes determines the classification of an object (a new instance). The object is grouped into the K closest neighbors' most prevalent class, which is mostly determined by a distance function.

- Multilayer Perceptron: A three-layer feed-forward neural network is called a Multilayer Perceptron. The input layer, hidden layers, and output layer are these three. From input to output, data in an MLP flow in a forward manner. The neurons are trained using back propagation.

- Support Vector Machine: For classification, Support Vector Machine is frequently employed. The algorithm takes data as input and, if possible, creates a line that separates those classes. The technique's goal is to identify the best decision boundary or line for categorization. A hyper-plane is the name given to this optimal choice boundary.

IV. Performance Analysis

The performances of the six classification models are illustrated in the figures. The used evaluation metrics are Accuracy, Precision, Recall and F1 Score values of the models.

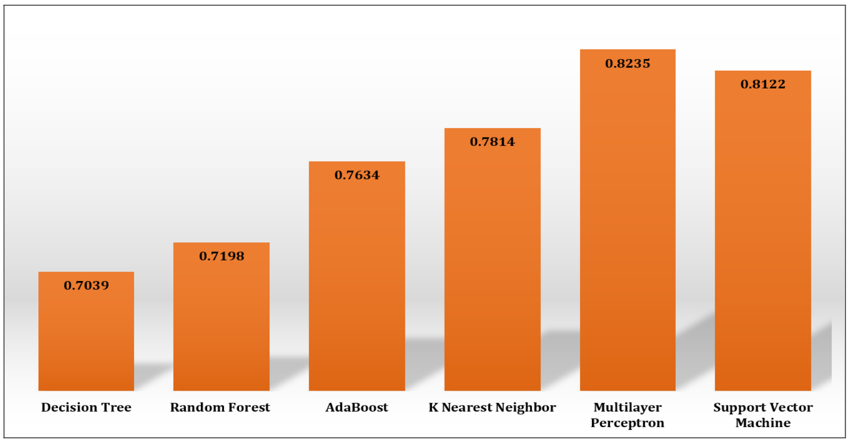

Fig. 3: Recall values of the ML Models for Predicting Breakout Stocks

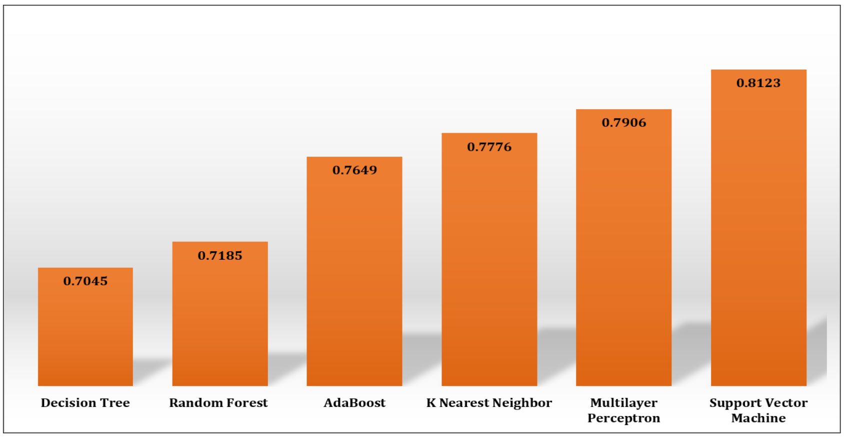

Fig. 4: F1 Scores of the ML Models for Predicting Breakout Stocks

Figure 1 shows the Accuracy values, Figure 2 displays the Precision scores, Figure 3 illustrates the Recall values and Figure 4 represents the F1 Scores of the trained models' performances on the Test portion of the dataset. Among these six models, Decision Tree has had the poorest outcomes while Multilayer Perceptron and Support Vector Machine are the leading ones.

V. Conclusion and Future Works

In our work, we have observed that machine learning models can very efficiently identify breakout stocks. Such prediction can help an investor to make wise decisions so that minimal loss and maximum gain can be achieved. Support Vector Machine and Multilayer Perceptron algorithms have shown the most promising results while all the models have achieved satisfactory outcomes. In future, we intend to expand our research and do experiments with more datasets targeting different stock exchanges from various countries and want to work with Deep Learning models.

References

[1] Abdullah, Sheikh Shaugat, and Mohammad Saiedur Rahaman. "Stock market prediction model using tpws and association rules mining." In *2012 15th International Conference on Computer and Information Technology (ICCIT)*, pp. 390-395. IEEE, 2012.

[2] 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. https://doi.org/10.1016/j.eswa.2015.05.013

[3] 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.

[4] Islam, Sharmin, Md Sikder, Md Hossain, and Partha Chakraborty. "Predicting the daily closing price of selected shares on the Dhaka Stock Exchange using machine learning techniques." SN Business Economics 1, no. 4 (2021): 1-16. DOI:10.1007/S43546-021-00065-6

[5] Anbalagan Thirunavukarasu, and S. Uma Maheswari. "Classification and prediction of stock market index based on fuzzy metagraph." *Procedia Computer Science* 47 (2015): 214-221. https://doi.org/10.1016/j.procs.2015.03.200

[6] Xu, Yanru, Zhengui Li, and Linkai Luo. "A study on feature selection for trend prediction of stock trading price." In *2013 International Conference on Computational and Information Sciences*, pp. 579-582. IEEE, 2013.

[7] Yu, Xinpeng, and Dagang Li. "Important trading point prediction using a hybrid convolutional recurrent neural network." *Applied Sciences* 11, no. 9 (2021): 3984. https://doi.org/10.3390/app11093984

[8] Wen, Qinghua, Zehong Yang, Yixu Song, and Peifa Jia. "Intelligent stock trading system based on SVM algorithm and oscillation box prediction." In *2009 International Joint Conference on Neural Networks*, pp. 3341-3347. IEEE, 2009.

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

Md. Siam Ansary has completed Bachelor of Science (BSc) in Computer Science and Engineering (CSE) from Ahsanullah University of Science and Technology (AUST), Dhaka, BANGLADESH. His undergraduate thesis work was on Extractive Summarization of Newspaper Articles. In the year of 2021, he joined the Department of CSE as a Lecturer. His research interests are on the various fields of Machine Learning and Artificial Intelligence.