An Efficient method for predicting stock market trend using Semi supervised Machine Learning

Milind Kolambe
Smt Kashibai Navale College of Engineering,
Cummins College of Engineering for Women,
Pune, India
Milind.kolambe@cumminscollege.in

Dr.Sandhya Arora
Cummins College of Engineering for Women,
Pune, India
e-mail: sandhya.arora@cumminscollege.in

Abstract— Before making any kind of financial investment, stock market analysis is frequently conducted using the common methods of fundamental analysis and technical analysis. Machine learning algorithms have extremely limited application in fundamental analysis, although they are widely employed in technical analysis. When performing technical analysis, the majority of investors use a methodical approach that involves plotting Japanese candles and obtaining important information. When plotted individually as a time series data, some of these candles provide useful information, but not all candles. The question is if certain candles aren't useful when used individually, will they still be useful when used in groups? When plotted in groups, certain known groups of candles can offer useful information. More patterns would definitely need to be found. It would also be intriguing to learn whether these trends tend to repeat themselves. Finally, regardless of the type of candle, generalized rules must be discovered. A semi-supervised approach was the best option to study this. Following this investigation, we actually discovered several helpful guidelines that provide investors with greater knowledge.

Keywords- Stock Market; Machine Learning; Apriori Algorithm; Japanese Candlesticks; Trend Prediction; Candlestick Patterns

I. INTRODUCTION

In the stock market, technical analysis is done before making any investments. The investigation is crucial in any subject. For an example, suppose there are two options if one of the best restaurants has to be chosen for a dinner date with a friend that you want to go to an unfamiliar location. The first option is to sample cuisine from every restaurant before inviting your friend to join you on the designated day. Another alternative is to choose the restaurant that will be busiest that day. The second alternative is the most effective explanation of how technical analysis is used in the stock market to create different divisions where attractive market chances must be identified. Technical analysis definitely saves time, but every strategy has advantages and disadvantages. Line charts, bar charts, and Japanese Candles are three common chart styles used in technical analysis.

When displayed in charts, the form used in the Japanese method of price charting resembles a candle. The Japanese Candlestick technique is one of the most popular ways that modern traders who use technical tools examine charts. The candlestick works in this way. For a certain period of time, the price at each candlestick's open, high, close, and low price points is shown. The high and low price points in the price range are shown by the length of the candles. The thickness of the region depicts the spread in price between the opening and closing prices of the market. If the price at market closing is higher than the price at market opening, the candlestick rectangle turns white. This indicates an optimistic outlook for the stock market. If the price at market closing is lower than the price at market opening, the thick portion of the candlestick turns black. This shows that the market is unhappy at the moment.

The candlesticks have individual names and exist in a range of forms, dimensions, and designs. For instance, one of the most fundamental and well-known candlestick patterns is the "Hammer" pattern. The candle in question has a long, thicker bottom and a shorter, thinner top. One candle makes up The Hammer, a type of bullish reversal candlestick. Although noise and trend components in the stock market constantly have varying degrees of predictability, all such sorts of candles signal some anticipated movement [1].

Since it is known that the data recorded about the stock price values, i.e. price at which market opens, price at which market closes, the highest price, and the lowest price, are the best options [2, 3] for the trading action for the given period, Charts might be used to display this information in the most accessible way. There are four data points per trading day. It is known as the OHLC. For instance, 20 points must be plotted on a chart for five days. It provides us with inspiration to perform some lengthy calculations.

The process is made easier with charting even if it is claimed that time series in the stock market are more complicated when compared to other types of series because of their long duration, cyclical fluctuations, seasonal changes, and price change in irregular patterns [4]. Short and lengthy trends can occasionally be combined at random [5, 6]. Signals are shown on charts. Given that market players are constantly correcting prices, Price change can be seen throughout the entire day [7]. Despite these price changes, it can still be viewed as signals [8]. Undoubtedly similar patterns can be uncovered that are influenced by a few factors [9]. The elements of the time series are predictable [10]. For time series trend analysis, it might be extremely problematic when the final data needed to make candles must be collected [11]. The accuracy of the prediction is occasionally impacted by other factors. Typically, this is the outcome of some significant causes, such as certain news or occurrences [12, 13, 14, 15]. Finding out which one would be tough, thus it is important to remember that understanding time series characteristics is difficult. [16].

Only one of the most popular charts, the line chart, can be utilized for technical analysis. Four data points must be considered, not just one, which is the reason. Few commonly used patterns are Marubozu, The spinning top, Doji, The Paper Umbrella, Hammer, Hanging Man, Shooting star.

II. Literature Survey

There is a long list of authors who have written about the modeling and study of time series in the financial sector. The currently used patterns and approaches are insufficient to provide acceptable accuracy [17]. Because the majority of these systems did not take into account the effects of many aspects including economics, politics, and psychology investments [19], available methods do not produce exceptional outcomes for long-term prediction [18].

Stock chart images were used by Lee, Kim, Koh, and Kang [20] with deep Q-Network and neural networks to forecast pricing on the worldwide stock market. Recently, a strategy was suggested for predicting stock crises by Naik and Mohan [21]. The hybrid feature selection algorithm was employed. Additionally, technical indicators were used like moving averages and the RSI. In literature, there is a long list of authors who have discussed the modeling and investigation of financial sector time series. The methods and patterns now in use are insufficient to provide good accuracy [17]. Because most of the existing algorithms did not take into account the effects of many aspects including economics, politics, and psychological investing, they did not produce exceptional outcomes for long-term prediction [18].

In order to forecast global stock market pricing utilizing stock chart images, deep Q-Network and neural networks were used by Lee, Kim, koh, and kang [20]. A strategy for predicting stock crises was recently put forth by Naik and Mohan [21]. An algorithm was employed for hybrid feature selection. Technical tools like moving averages and the RSI were also used. Although used for a long time in Japan, Japanese candlesticks were discovered to be less effective outside of Japan. The traders in the West knew nothing about it. Later, candlesticks were discovered by Steve Nison, and this technique was revealed to everyone. In addition, a book was written by him titled "Japanese Candlestick Charting Techniques" that is popular among traders. Japanese names are assigned to almost all designs. Using this strategy, a specific candle can be categorized as bullish or bearish. white (bullish) or black (bearish) candle (bearish).

The Japanese candlestick charts were used by a very few researchers to forecast the stock market. To forecast the market trend, these candlesticks were combined with technical indicators and machine learning techniques by Lin, Liu, Yang, and Wu [26]. A candlestick-based prediction system was created by Birogul, Temur, and Kose [27] and aids in deciding whether to purchase or sell a security. Clustering technique was developed by Inventors Ding and Luo [28]. This work's main motivation is to offer a novel technique for forecasting stock market movement and to assess the suggested candlestick model's predictability using the apriori algorithm's notion of "confidence."

It is understood that daily "open price," "close price," "high value," and "low value" play crucial roles in determining the price for the following day in the stock market. The parameters that determine the contract's current quotation, in addition to the stock market index value, are the highest quote, lowest quote, closing quote, volume, total number of contracts exchanged, and the day the contract was first traded. This study also has the goal of encouraging investors with long-term investment ideas that will optimize returns.

III. Proposed System

It is highly hoped to obtain some common item sets like Japanese candlestick patterns in the case of the stock market. The many stock market data patterns that have been identified over time have already been covered. Some patterns can be easily spotted through close observation but to find out complex patterns, an algorithmic approach must be chosen while employing a vast dataset to identify more intriguing patterns. It might provide us with more intriguing regular patterns that could produce encouraging outcomes.

Nifty time series data was obtained spanning 10 years to conduct this experiment (2006-2015). Most recent data was purposefully not chosen as the effects of the pandemic crisis may be apparent. There are numbers in this dataset. Seven numerical figures are displayed per day in each row. Final three columns were eliminated from the dataset since only four different stock values were needed while using Japanese candlestick data in time series. This dataset is less than 50 KB in size with four specified attributes per row.

For this experiment, eleven of the most widely used candle forms were used. As a general illustration of Japanese candlestick patterns, a few of these patterns can be seen in figure 1.

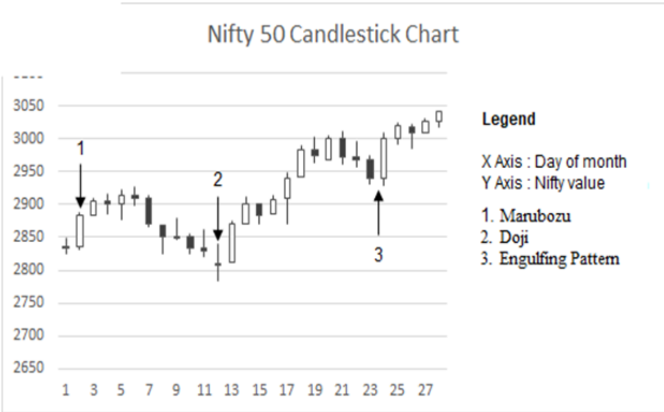

Figure 1. Japanese candle-stick patterns.

There are more than 50 of these patterns or shapes that can be employed, but 12 popular single and multi-candle Japanese candlestick patterns were chosen that are displayed here as Bullish Marubozu, Bearish Marubozu, The spinning top, Doji, Hammer, Hanging man, Shooting Star, Engulfing pattern, Bearish Harami, Bullish Harami, Morning star, Evening star.

A small number of candles give investors some crucial information, but not all. In order to provide information if the individual candles are not relevant, it would be interesting to determine whether the candles are associated in any way in groups.

Futures and options are a popular market category on NSE INDIA. After gaining sufficient stock market expertise, it could be said that practically every stock market investor invests in this sector. There are several contracts to purchase or sell in this segment. These contracts have various life expectancies or expiration dates. The monthly contract is the most traditional and often utilized type. This contract's expiration date is stated as being one month after it becomes trading. Every month's final Thursday is the expiration date for contracts. It is assumed that the support and demand ratio may be impacted by this expiration date. The duration of each item set for a month was chosen taking this supposition into account. This item set is used in apriori algorithm and will be taken into consideration up to this point. The number of patterns found this month is displayed in each row. When the specific Japanese candlestick shape is present in the provided item set, those were encoded as categorical values as "Y" and "N" at the appropriate point of the row to reduce the size of the dataset. Noting that all of these additions to the new dataset result in a 10x reduction in the size of the data set as a whole. Then, as shown in figure 2, the apriori technique was used to identify common item sets. As demonstrated in figure 2, even our own patterns can be added to improve accuracy.

Now that expected candle patterns can be forecasted, it might be used to guide more profitable stock market decisions. Any association rules that might exist between the appearances of particular patterns in each item set might be looked for. Results demonstrate that some rules are clearly identified as frequent item sets. For example, frequent item sets can be found that suggest a bullish trend followed by a state of uncertainty, followed by a bearish trend and another negative indicator that confirms the long bearish trend as a market correction.

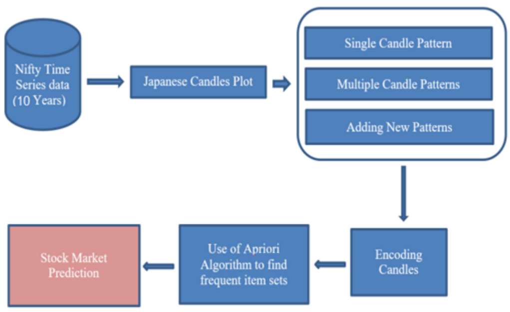

Figure 2. Proposed System

A. The apriori algorithm

The apriori algorithm consists of a series of steps that must be taken in order to identify the most frequent itemset in the given database. This data mining method repeats the join and prune processes up until the most common itemset is reached. The problem specifies or the user presupposes that there is a minimum support threshold.

1) Each item is treated as a candidate for a 1-itemset in the algorithm's initial iteration. The algorithm will count how many times each item appears.
2) Allow some minimal support, min sup (k). Determined is the set of 1 - itemsets whose occurrence is satisfying the min sup. The remaining candidates are trimmed, and only those who score greater than or equal to min sup are advanced to the following iteration.
3) Next, frequent 2-itemset items are found. For this, a group of two items is created in the join phase by joining the objects with one another.
4) The min-sup threshold value is used to trim the 2-itemset candidates.
5) The algorithm is stopped when the most frequent itemset is achieved.

IV. Results and Discussion

The results of the first experiment clearly demonstrated that the candles indeed have some association. Bullish and bearish candles were found to exhibit certain common patterns. Following that, the findings were attempted to be generalized based on their type, function, and placement, such as hammer and hanging man. Here are better results obtained.

The following are the rules that were identified when the Apriori algorithm was applied to historical data going back ten years.

A. Marubozu, Hammer/Hanging Man >> Marubozu (Confidence = 0.83).

Bearish marubozu could be witnessed with a hammer in January 2006. According to figure 3, Marubozu will be bullish in the coming week. With confidence=0.81, this pattern is repeated frequently.

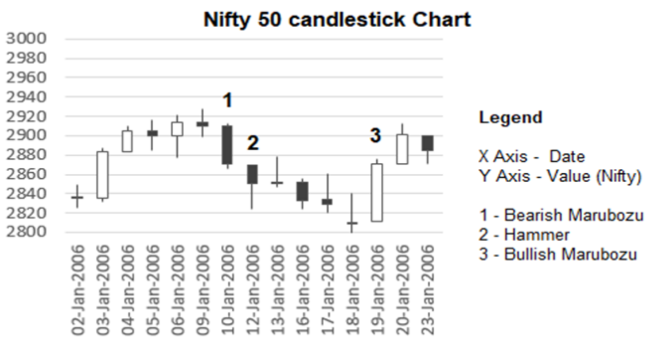

Figure 3. Prediction of bullish Marubozu.

B. Engulfing, doji >> Hammer/Hanging Man (Confidence = 0.81)

This pattern, which is regularly repeated and has a confidence level of 0.8, can be seen in May 2007 as illustrated in figure 4.

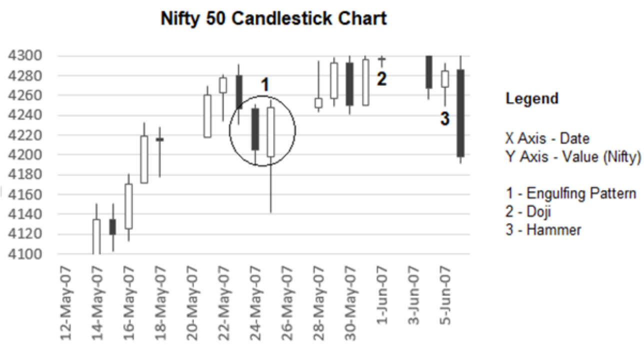

Figure 4. Prediction of Hammer

C. Morning Star, Harami >> Marubozu (Confidence = 0.67)

This pattern, which is depicted in figure 5, can be seen in February 2009. With confidence = 0.68, it is discovered to be repeated regularly once more.

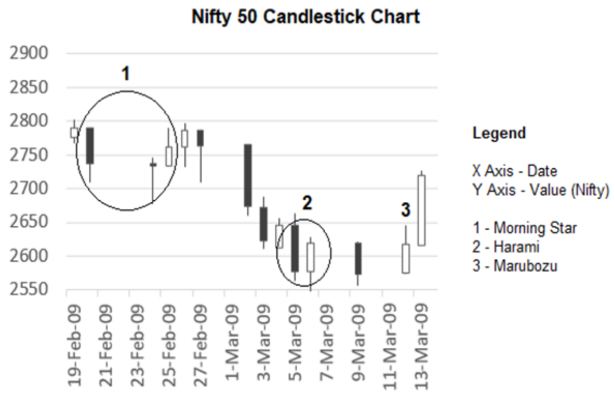

Figure 5. Prediction of bullish Marubozu

This concept was applied to two years of data (June 2020 to June 2022) for testing from the same market to confirm all previously found patterns. Findings were verified and it was discovered that using machine learning and Japanese candlesticks did indeed boost an investor's profit.

The following are the rules that were identified when the Apriori algorithm was applied to historical data going back ten years, and are found to be repetitive in 2 years testing data.
Marubozu (Confidence = 0.76), Hammer/Hanging Man >> Marubozu

D. Bearish marubozu could be witnessed with a hammer in May 2022.

According to figure 6, Marubozu will be bullish in the coming week. With confidence=0.76, this pattern is repeated frequently.

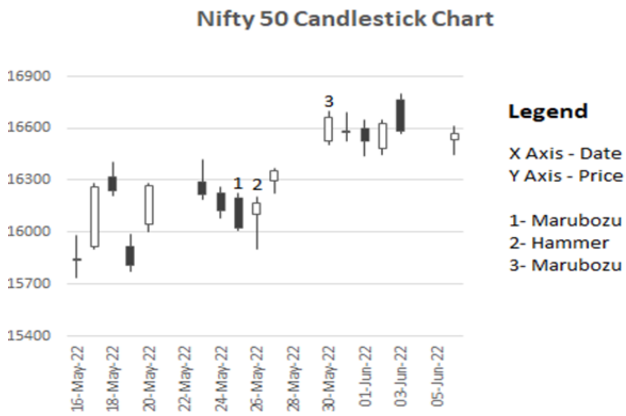

Figure 6. Prediction of bullish Marubozu

E. Engulfing, doji >> Hammer/Hanging Man (Confidence = 0.74)

This pattern, which is regularly repeated and has a confidence level of 0.74, can be seen in Jan 2022 as illustrated in figure 7.

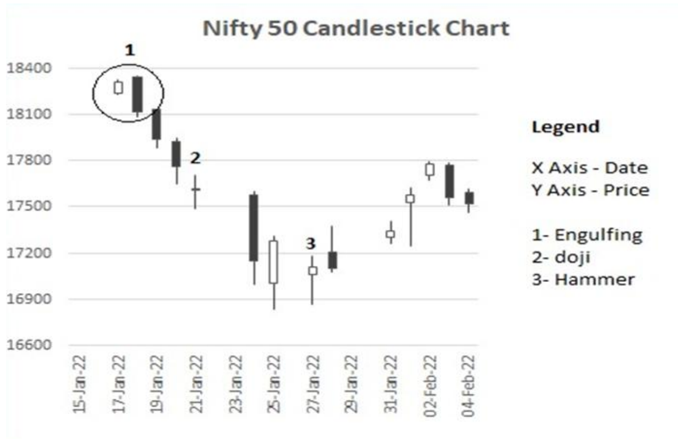

Figure 7. Prediction of Hammer

F. Morning Star, Harami >> Marubozu

This pattern, which was found to be repetitive in 10 years of data, was absent as per the minimum level of support in 2 years of data. It was understood that this pattern could be found in a large dataset only.

The use of this technique on a huge dataset has produced additional intriguing common patterns that may yield encouraging outcomes. The investors will benefit in several ways from the discovery of these regulations.

V. Conclusion and Future Scope

Through the use of time series data from Japanese candlestick charts and the apriori algorithm, a semi supervised method to detect commonly occurring patterns of Japanese candlestick shapes is proposed in this paper. The findings unequivocally show that there are certain useful association rules between various candlestick patterns. The investor will be able to predict the trend's course and will thus have the opportunity to boost their profit margin. The prediction algorithm will function more quickly since the dataset size is much smaller than that of algorithms that use the raw time series data. After refining our experiment, it became clear that generalizing the patterns yields somewhat better results. The apriori algorithm has some shortcomings, despite producing acceptable results for detecting frequent item sets. Future studies will concentrate on employing different learning methods to forecast stock market movements using candlestick patterns, such as neural networks [29] and support vector machines [30]. Additionally, the selection of the item set would be optimized in light of the many factors affecting the index value. Another potential application for the future is the adaptation of this methodology to other stock market segments, such as stocks, mutual funds, etc. The apriori algorithm's dependence on transaction size is one of this experiment's limitations; going forward, it will be possible to achieve better results by using an algorithm that is independent of transaction size.

References

[1] E. Maiorino, F. M. Bianchi, L. Livi, A. Rizzi, and A. Sadeghian, ‘‘Data driven detrending of nonstationary fractal time series with echo state networks,’’ Information Sciences, vols. 382–383, pp. 359–373, Mar. 2017.

[2] W. Bao, J. Yue, and Y. Rao, ‘‘A deep learning framework for financial time series using stacked autoencoders and long-short term memory,’’ PLoS ONE, vol. 12, no. 7, 2017,

[3] F. Ye, Z. Liming, Z. Defu, F. Hamido, and G. Zhiguo, ‘‘A novel forecasting method based on multi-order fuzzy time series and technical analysis,’’ Information Sciences, vols. 367–368, pp. 41–57, Nov. 2016.

[4] A. K. Rout, ‘‘Forecasting financial time series using a low complexity recurrent neural network and evolutionary learning approach,’’ Journal of King Saud University – Computer and Information Sciences, vol. 29, no. 4, pp. 536–552, 2017.

[5] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, ‘‘Modeling long-and short- term temporal patterns with deep neural networks,’’ in Proc. 41st Int. ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 95–104, 2018

[6] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. J. Amaratunga, ‘‘Empirical Mode Decomposition based ensemble deep learning for load demand time series forecasting,’’ Applied Soft Computing, vol. 54, pp. 246–255, May 2017.

[7] N. Nava, T. Di Matteo, and T. Aste, ‘‘Financial time series forecasting using empirical mode decomposition and support vector regression,’’ Risks, vol. 6, no. 1, p. 7, 2018.

[8] L. B. Godfrey and M. S. Gashler, ‘‘Neural decomposition of time-series data for effective generalization,’’ IEEE Transactions on Neural Network Learning Systems, vol. 29, no. 7, pp. 2973–2985, Jul. 2018.

[9] S. Jeon, B. Hong, and V. Chang, ‘‘Pattern graph tracking-based stock price prediction using big data,’’ Future Generation Computer Systems, vol. 80, pp. 171–187, Mar. 2018.

[10] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control. Hoboken, NJ, USA: Wiley, 2015.

[11] X. Dai and M. Bikdash, ‘‘Trend analysis of fragmented time series for mHealth apps: Hypothesis testing based adaptive spline filtering method with importance weighting,’’ IEEE Access, vol. 5, pp. 27767–27776, 2017.

[12] X. Ding, Y. Zhang, T. Liu, and J. Duan, ‘‘Deep learning for event-driven stock prediction,’’ in Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.

[13] X. Ding, Y. Zhang, T. Liu, and J. Duan, ‘‘Knowledge-driven event embed- ding for stock prediction,’’ in Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics, pp. 2133–2142, 2016

[14] L. Zhang, C. Aggarwal, and G.-J. Qi, ‘‘Stock price prediction via discovering multi-frequency trading patterns,’’ in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2141–2149, 2017

[15] T. Fischer and C. Krauss, ‘‘Deep learning with long short-term memory networks for financial market predictions,’’ European Journal of Operational Research, vol. 120, no. 2, pp. 654–669, 2017.

[16] Z.-K. Gao, S. Li, W.-D. Dang, Y.-X. Yang, Y. Do, and C. Grebogi, ‘‘Wavelet multiresolution complex network for analyzing multivariate nonlinear time series,’’ International Journal of Bifurcation and Chaos, vol. 27, no. 8, Jul. 2017, Art. no. 1750123.

[17] Z.-K. Gao, M. Small, and J. Kurths, ‘‘Complex network analysis of time series,’’ IEEE Access, vol. 7, Pages 40220 - 40229, 2019

[18] Jui-Sheng Chou,Thi-Kha Nguyen, “Forward Forecast of Stock Price Using Sliding-Window Metaheuristic-Optimized Machine-Learning Regression,” IEEE transactions on industrial informatics, vol. 14, no. 7, july 2018, pp. 3132 – 3142

[19] Dawei cheng, Ye liu, Zhibin niu, Liqing zhang, “Modeling Similarities Among Multi-Dimensional Financial Time Series,” IEEE Access, vol. 6, pp. 43404 – 43413, 2018

[20] Jinho lee, Raehyun kim, Yookyung koh, and Jaewoo kang, “Global Stock Market Prediction Based on Stock Chart Images Using Deep Q-Network”, IEEE Access, Vol.7, pp.167260 – 167277, 2019.

[21] Nagaraj naik, (Graduate student member, IEEE), and Biju r. mohan, “Novel Stock Crisis Prediction Technique—A Study on Indian Stock Market”, IEEE Access, Vol.9, pp. 86230 – 86242, 2021.

[22] Brenda a. januário; Arthur emanuel de o. carosia, Ana estela a. da silva, guilherme p. Coelho, “Sentiment Analysis Applied to News from the Brazilian Stock Market”, IEEE Latin America Transactions, Vol. 20, pp. 512 – 518, 2022.

[23] Zeynep hilal kilimci and Ramazan duvar, “An Efficient Word Embedding and Deep Learning Based Model to Forecast the Direction of Stock Exchange Market Using Twitter and Financial News Sites: A Case of Istanbul Stock Exchange”, IEEE Access, vol. 8, pp. 2169-3536, 2020.

[24] Suman saha (graduate student member, ieee), Junbin gao and Richard gerlach, ‘‘Stock Ranking Prediction Using List-Wise Approach and Node Embedding Technique’’, IEEE Access, vol. 9, pp. 88981 - 88996, 2021.

[25] Sheikh mohammad idrees, M. afshar alam and Parul agarwal, ‘‘A Prediction Approach for Stock Market Volatility Based on Time Series Data’’, IEEE Access, vol. 7, pp. 2169-3536, 2019.

[26] Yaohu lin, Shancun liu, Haijun yang (member, ieee), and Harris wu, ‘‘Stock Trend Prediction Using Candlestick Charting and Ensemble Machine Learning Techniques With a Novelty Feature Engineering Scheme’’, IEEE Access, vol. 9, pp. 2169-3536, 2021

[27] Serdar birogul, Günay temür and Utku kose, ‘‘YOLO Object Recognition Algorithm and ‘‘Buy-Sell Decision’’ Model Over 2D Candlestick Charts’’, IEEE Access, vol. 8, pp. 2169-3536, 2020

[28] Ding fengqian and Luo chao, “An Adaptive Financial Trading System Using Deep Reinforcement Learning With Candlestick Decomposing Features”, IEEE Access, vol. 8, pp. 2169-3536, 2020

[29] Min Wen, Ping Li, Lingfei Zhang, Yan Chen, “Stock Market Trend Prediction Using High-Order Information of Time Series,” IEEE Access, vol. 7, pp. 28299 – 28308, 2019

[30] J.-S. Chou and A.-D. Pham, “Nature-inspired meta heuristic optimization in least squares support vector regression for obtaining bridge scour information,” Information Sciences, vol. 399, pp. 64–80, 2017