



Stock market analysis using candlestick regression and market trend prediction (CKRM)

M. Ananthi¹ · K. Vijayakumar²

Received: 12 October 2019 / Accepted: 13 March 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Stock market data is a time-series data in which stock value varies depends on time. Prediction of the stock market is an endeavor to assess the future value of a company's stock rate which will increase the investor's profit. The accurate prediction of stock market analysis is still a challenging task. The proposed system predicts stock price of any company mentioned by the user for the next few days. Using the predicted stock price and datasets collected from various sources regarding a certain equity, the overall sentiment of the stock is predicted. The prediction of stock price is done by regression and candlestick pattern detection. The proposed system generates signals on the candlestick graph which allows to predict market movement to a sufficient level of accuracy so that the user is able to judge whether a stock is a 'Buy/Sell' and whether to short the stock or go long by delivery. The prediction accuracy of the stock exchange has analyzed and improved to 85% using machine learning algorithms.

Keywords Machine learning · Stock price · Time-series data · Candle-stick pattern · Regression analysis

1 Introduction

The stock prices are volatile in nature and are hard to predict them. The market price movement changes due to various reasons and the sentiment of the market depends on the current issues and decision taken not only by a single person but also lot more who invest their money in this. All the people who invest in the market should be able to decide where to invest their money. Predicting time series data to forecast stock prices is one of the applications of machine learning techniques. Machine learning is an artificial intelligence technology that provides the system which is able to learn and develop prediction automatically from experience without precise programming. Many machine learning algorithms have already been used to predict the stock market trend. Neural Networks have been used widely by many

research people. Time delay analysis has also been used to predict the market trend prediction. Machine learning algorithms like Support Vector Machine (SVM), Bayesian belief networks and evolutionary algorithms. The k-NN regression algorithm is one of the machine learning algorithm is considered to have more accuracy than the other regression machine learning algorithms such as linear regression and Support Vector Machine. Predicting the stock market is something that has been done before but only for monetary purpose in the name of algorithm trading. Candlestick is a traditional pattern for stock market analysis in which prediction is done based on the previous and historical data of open and close values of stock (Ahmed 2017).

The proposed system has used the classification and regression methods of machine learning technique to foresee the readings of previous and current stock values for prediction. Our system generates signals for buying/selling equity based on a number of parameters. These signals are kept open for everyone to view so that we have the power of the market behind us. The strength of the signal is measured from 3 distinctive parameters so that false signals are not generated. Existing systems include stock insight websites like Moneycontrol.com, Screener.in. They all provide several tools but they lack user friendliness for newbie users. We used webworkers to collect data from various sources

✉ M. Ananthi
ananthi.sairam@gmail.com

K. Vijayakumar
mkvijay@msn.com

¹ Department of Information Technology, Sri Sairam Engineering College, Chennai, India

² Department of Computer Science and Engineering, St. Joseph's Institute of Technology, Chennai, India

over the web like Indian indices, Moneycontrol and other related content providers using the scrapy library. The tools provided in this work helps ease into it. We also provide a holistic outlook on whether to buy or sell stocks by analyzing the price trend.

2 Literature survey

The literature survey of various stock market prediction methods and machine learning algorithms for stock analysis were carried out. Nakov et al. (2016) made predictions on stock price returns using micro blogging sentiment analysis by analyzing the pattern in blogs to predict stock price overturn in near future. Correlation between Initial Public Offering and stock market are also done by gaining the public opinion on stock price movements. Recent research tends to hybridize other AI techniques with ANN. Kim and Shin et al. have proposed a hybrid model of Genetic Algorithms and Neural Networks for optimization of the number of time delays and network architectural factors using GA, to improve the conclusiveness of constructing the ANN model. Sadia et al. (2019) used random forest and support vector machine algorithm for Stock market prediction. Data preprocessing was done before analyzing the data, then stock data was predicted using machine learning algorithms and stated that accuracy improved by using random forest method. Zhao et al. (2019) stated the relationship identified between the e-commerce in social networks and stock market which will be helpful for stock analysts. Sun et al. (2017) stated the relationship between stock price movement and social media sentiment in China, which has large market capitalization and a unique social media landscape. They collect data from several different types of social media sites (microblogs, chat rooms, web forums), and find that data from these sites exhibit distinct characteristics in activity level, post length, and correlation with stock market behavior. They also propose a prediction model that uses chat room sentiment to forecast the market direction, and develop a trading strategy that utilizes the prediction as trading indicators. Their sentiment analysis is clouded by false pretenses by anonymous chat users. Erdogan and Namli (2019) and Ahmed et al. (2016) measured quality of life index using machine learning techniques of artificial neural networks and support vector machine.

Hiransha et al. (2018) performed the stock market prediction using deep learning algorithms. NSE(India) and NYSE(New York) data have analyzed using convolutional neural networks method which outperformed well. Regression based method and Long short term memory network based method are used by Chouhan et al. (2018) to predict stock market analysis. Only one dataset is used for testing the accuracy and prediction which will not considered as an

efficient solution. Kalra and Prasad (2019) predicted stock value using news articles and past historical values. Traditional Naïve Bayesian classification method was used to classify text from articles. It was proved that kNN algorithm was more efficient than Naïve Bayes, SVM, Neural network algorithms. But Naïve Bayesian method was used and prediction was made only based on text values which will not be accurate always.

Various candlestick patterns were represented and all the algorithms were analyzed by Martinsson and Liljeqvist (2017) and Somani et al. (2014) predicted stock analysis using hidden markov model which is compared with traditional algorithms of SVM and Neural network methods. They proved that predicted values were nearer to actual values. Usmani et al. (2016) discussed multi-layer perceptron algorithm to analyze Karachi stock market. 77% accuracy proved for 100 instances of data. Peng (2019) stated stock data is pipelined and preprocessed using the concept of big data apache hadoop which is implemented using machine learning spark tool. The results found were not accurate, still need to be modified with other approaches. Peng (2019); Wen and Li (2019) predicted financial time series data using convolutional neural network method. They proved 5 to 7 percentage improvement in the accuracy compared to traditional algorithms. Vijayakumar and Arun (2017) discussed the risk of identifying automated assessment process using Natural Language Processing and a tool was developed with deep learning technique. Chandar (2019), predicted stock market real-time data using discrete wavelet transform and fuzzy logic. He stated that the tool developed will be useful for financial assistance, stock exchange analysts, etc. Ahmed et al. (2019) has considered the a discussion towards the QoS of data improvisation considered towards medical aspect of interpreting stock data operations. With addon featuring towards appending secure stocking (Ahmed et al. 2020; Kadam et al. 2019)

From this survey, it has been found that machine learning algorithms are suitable for predicting stock market analysis. It is being identified that preprocess the data is better than using raw data but which is difficult to implement for dynamic stock market data. This paper is proposed to preprocess the data and analyze the stock prediction using machine learning algorithm with traditional candlestick pattern which will improve the accuracy of the prediction.

3 Proposed stock market prediction system

Stock market investment is considered to be one of the major forms of the investment, the proposed system helps the user to identify the right stock to invest in by performing pattern detection and by learning the past price movements. The proposed system includes the creation of a portfolio for

users by maintaining their investment pattern on which company they have invested to give an overview of the whole scenario of their investments. The implementation of the regression algorithm is applied on the dataset collected from an open source library called Quandl for obtaining financial and alternative data, then predicting the trend of the market based on various other indicators like MACD, RSI, Bollinger bands and also candle stick pattern detection. The stock prediction analysis is displayed through graphs which consists of signals to the user on which stock to invest.

Figure 1 describes the overview of the proposed system of Stock market prediction analysis. The company details are retrieved by Django templates, stock values are monitored and price movements are analyzed by Portfolio analysis and readings are measured and displayed through MACD graph. Further the stock values are analyzed using kNN regression machine learning algorithm. The prediction of stock market is also analyzed by considering previous stock history, user interest on investment which is collected using SQLite. Finally, Signals are generated and displayed using Quandl Rest API.

The stock prediction is done into three main modules such as,

1. *Trendline Identification* This module consist of representation of candlestick graph and MACD with trade arrows indicating the buying and selling time.
2. *Learning Past Price Movements* K Nearest Neighbor algorithm is to used to learn the past price movements and compare them with the original data.
3. *Pattern formation* The various pattern formation are analyzed and signals are indicated in this module.

3.1 Market trend identification

Market trend has to be identified to analyze the future stock movement. This can be achieved by forming various patterns using past price movement and indicators to examine the selected stock. The 3 parameters using which the price trend of the stock is analyzed and signals are generated are,

1. Pattern formations
2. Learning past price movement
3. Indicators

3.1.1 Pattern formation

Pattern is formed by analyzing the Open, High, Low, Close (OHLC) data over the past few days. The traditional method of Japanese candlestick pattern is used to formulate the

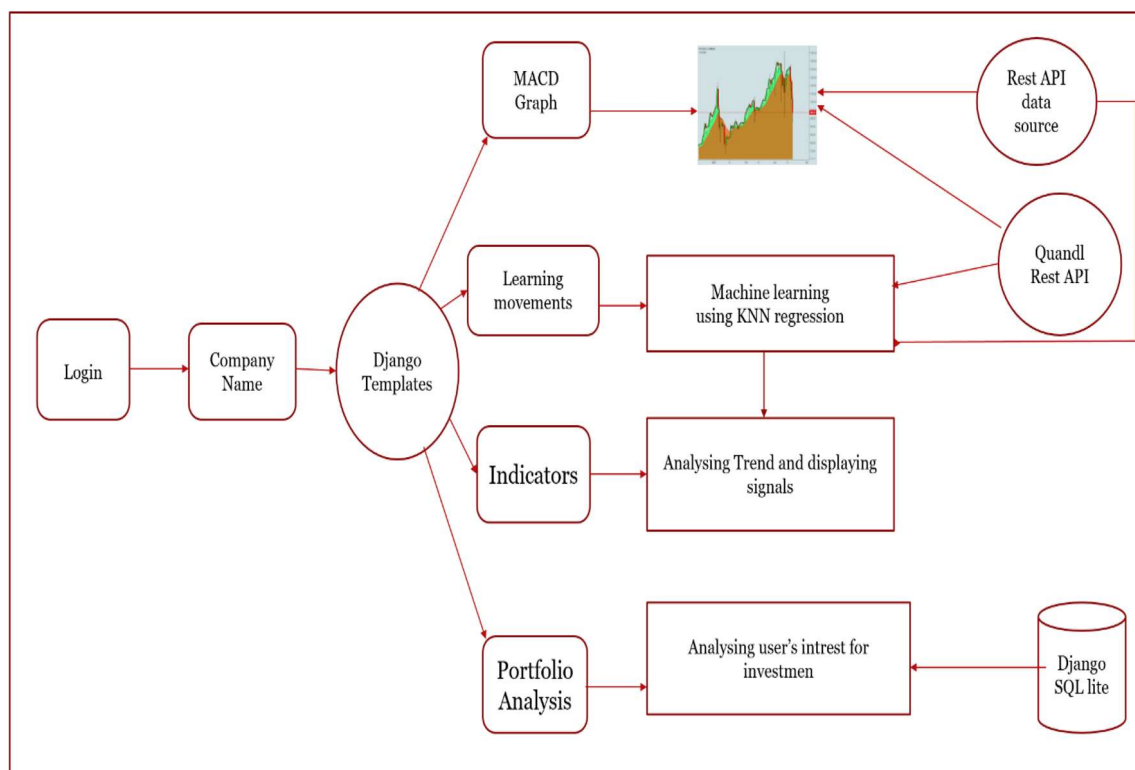
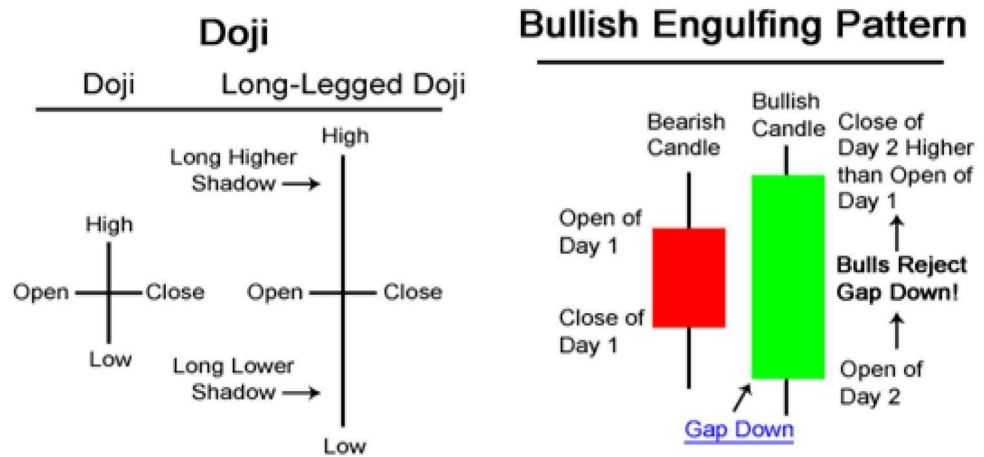


Fig. 1 Stock market prediction overview

Fig. 2 Doji and bullish candlestick patterns [Ref. commodity.com]



pattern. Among many patterns, Bullish and Bearish patterns are used in the system and the pattern is dynamically changed by analyzing several candles. Figure 2 represents Doji and Bullish Engulfing pattern which is taken from commodity.com. The pattern has a bearish candle on the first day and a larger bullish candle on the second day which engulfs the first candle.

Doji is a sign of indecision in the market and when it occurs at the end of the trend cycle, it indicates that there is going to be a trend reversal i.e. the next day price movement will be opposite to the last few days. The prediction is analyzed using these methods to determine if the stock is currently bullish or bearish at-least for the next day.

3.1.2 Learning past price movement

With access to historical data of price movement of stocks, we can train models based on the past price movement. This enables us to get several classifications to which the recent price movement can be tested on. By fit of the test data with the past data, we can predict the upcoming few days as history always tends to repeat itself. For example, if ABC inc. has had its CMP move by 15% in the past week. We can test the data with past scenarios where it surged 15% in a similar time period. Based on the trend after that, we can conclude whether the surge sustained, or the gain was a simple bubble. Also, by usage of knn regression, the movement of supports and resistances can be analyzed effectively. If the price keeps constantly breaking S&R, we can scale up accordingly.

3.2 K nearest neighbors regression (k-NN regression)

Once data set is preprocessed, it is classified using nearest dataset by calculating the distance among themselves. This is implemented using classification and regression

Table 1 Stock values of Tata Consultancy Services

S. no.	Forecast price	Original price
1	2959.3000	2959.30
2	3174.6000	3174.60
3	3121.0500	3102.00
4	3198.8500	3198.85

techniques of machine learning algorithm. K-NN regression method is used in this system whose output is used as the stock property value for the object. The weightage of the data sets are calculated using the distance formula, nearest distance are identified and classified

The algorithm is calculated as follows:

- Identify the dataset, then calculate its distance using the formula specified in Eq. 1. The distance between a point P and a distribution D is calculated using the following equation

$$D_m(a, b) = \sqrt{(a - b)^n(a - b)} \quad (1)$$

where a and b are two random vectors of the same distribution.

- Categorize the similar datasets by increasing the distance.
- Compute root mean square deviation to locate nearby best possible number of datasets.
- Calculate an inverse distance weighted average with k nearest multivariate.

3.2.1 Indicators

Indicators are the last confirmation parameter of signal generation. Based on the results from the previous two

Table 2 Accuracy range of regression algorithms

Method	Accuracy range (in %age)
k-NN regression	90–96
Linear regression	80–95
Support vector machine	60–80

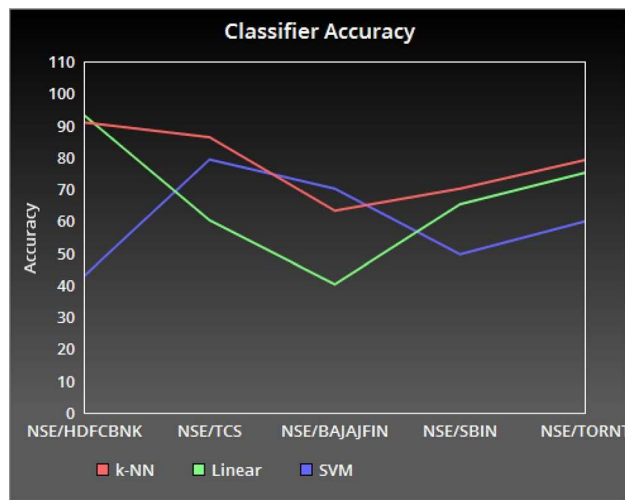


Fig. 3 Classifier accuracy

parameters, concurrence with the indicators can make a signal doubly strong. Even otherwise the analysis still holds and one can follow the signal in hopes of the indicator recovering. We intend to use a culmination of popular indicators like RSI, MACD, Alligator, Supertrend etc.

Based on the results of the three parameters, we can generate an abstract signal between 0 and 100 for buying/shorting of a stock based on pure technical analysis.

Fundamental analysis of stocks cannot be determined with accuracy as perception varies from person to person,

Table 4 Predicted values using machine learning algorithm

Company/machine learning algorithms	kNN regression	SVM	Linear regression
TCS	2948.679	2930.067	1962.287
Infosys	1149.887	1181.541	1921.568
Wipro	302.4868	443.0676	481.3753
HCL	930.1876	839.772	839.011
Oracle Finance	3882.339	4046.162	3192.046
MindTree	785.0703	730.844	686.5661
Hexaware	353.4225	319.8947	323.8736
Tech Mahendra	613.0417	633.1691	664.855
NIITech	850.5223	704.8029	398.9765

but we intend to do a sentiment analysis of each stock on popular social media to correlate with stock movement so that people can catch the current trends.

4 Implementation

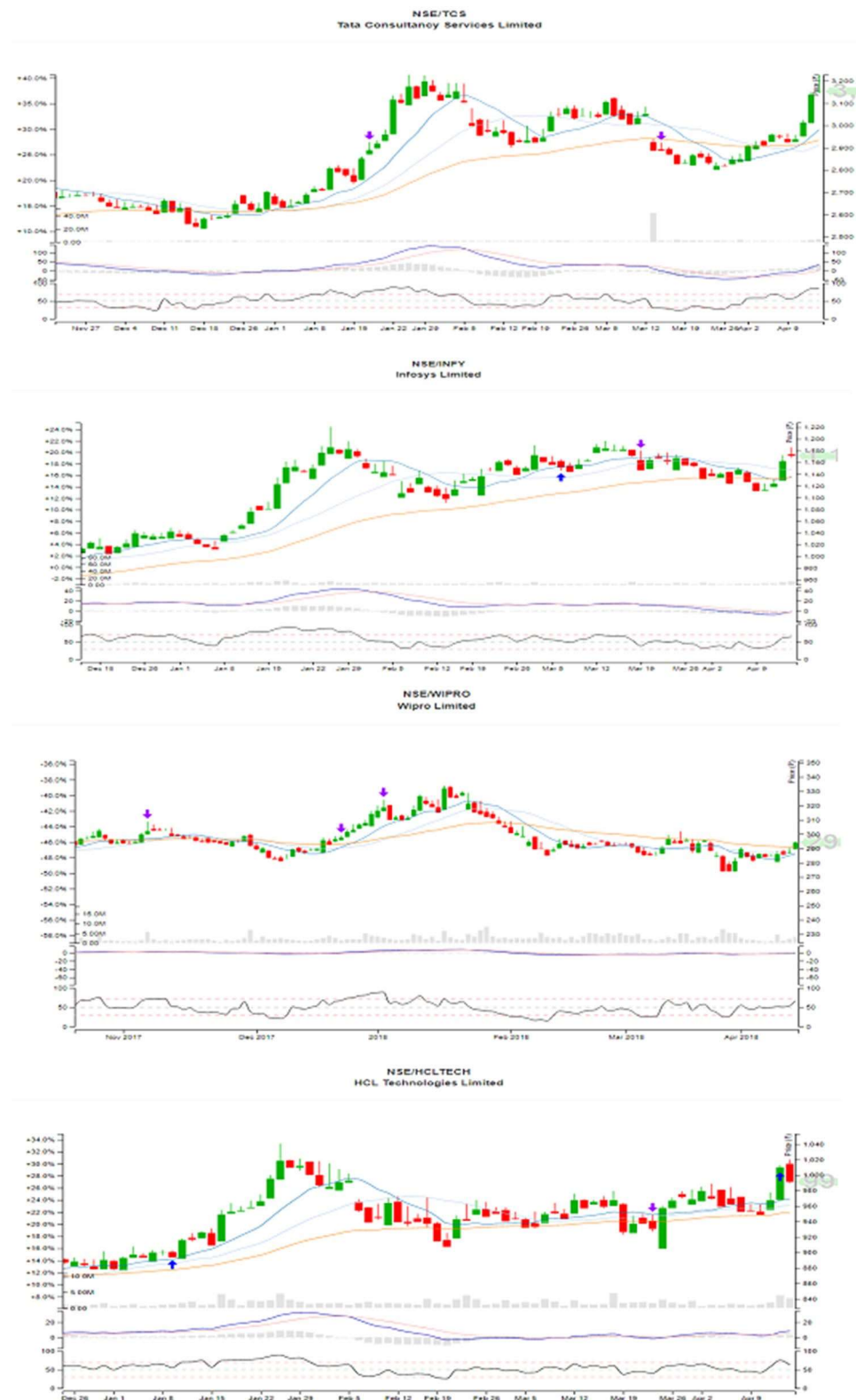
The stock market prediction system has 3 main modules which include the creation of a portfolio for users to give an overview of the whole scenario of user investments. This can be done by maintaining their investment pattern on which company they have invested. The implementation of the regression algorithm is applied on the dataset collected from an open source library called Quandl for obtaining financial and alternative data, then predicting the trend of the market based on various other indicators like MACD, RSI, Bollinger bands and also candle stick pattern detection. The last module consists of displaying results through graphs which consists of signals to the user on which stock to invest.

Different baskets of stocks are curated based on the risk profile of the user. The user can opt to follow the recommendations or maintain their own portfolio. The user can declare the amount of units held in each stock in the portfolio so that its value can be assessed. This module where the sentiment analysis is done in a small scale to give slight edge over the

Table 3 Price movements of various companies

Company	MACD(26,12)	Exp(9)	MACD histogram	RSI
TCS	3.064355	15.52661	− 12.4623	46.87864
Infosys	12.14546	12.70497	− 0.55951	57.33604
Wipro	− 3.13829	− 3.53998	0.401685	42.46668
HCL Technologies	2.35203	0.761203	1.590827	55.69949
Tech Mahindra	5.54565	11.34031	− 1.07144	57.55423
Oracle Financial Software Service Limited	− 48.2571	− 38.3524	− 9.90472	40.43316
Mind Tree	25.23676	26.04664	− 0.80988	60.38304

Fig. 4 Stock price variation with relative strength index



other prediction methods which help the user to gain an idea about their investment over the security.

4.1 k-NN regression

The datasets are trained and tested using machine learning library using python. The accuracy is calculated for different

training sets. The comparison of original data set and the trained dataset are shown below:

The forecast prices of the company Tata Consultancy Services are represented in Table 1 as an example. The price varies as the weighted values and is calculated based on the distance. The accuracy of the k-NN regression algorithm varies between 75 and 95% based the training dataset and they are regressed based on their distance.

4.2 Calculation of exponential moving average (EMA)

The average weight and significance of most relevant data sets are calculated using the formula as specified in Eq. 2.

$$S_{ma} = \sum P_n / n \quad (2)$$

where S_{ma} is simple moving average value, P_n is the total price for n days and n is the number of days. The every-day maximum weightage value is calculated using the close exponential values multiplied with maximum multiplier and previous value of exponential moving average as shown in Eq. 3.

$$E_{ma} = Cma_{n-1} * \left(\frac{2}{n+1} \right) + E_{ma}(n-1) \quad (3)$$

where E_{ma} is an exponential moving average value, Cma is the close Ema values, 2 represents maximum multiplier and $n+1$ is used to calculate the current day max weightage value. MACD line is calculated using the changes in the closing prices of EMA by setting the values 12,26 and 9 depending on stock trade style and goals.

4.3 Calculation of relative strength index

The variation in the stock price is measured using relative strength index based on the closing price of stock values. The variation of profit and loss is represented using the Eq. 4 as

$$R_{si} = 100 - \left(\frac{100}{1 + R_s} \right) \quad (4)$$

$$R_s = \sum_{i=0}^n G / \sum_{i=0}^n L \quad (5)$$

where R_s is relative strength calculated using the average of gain and loss. Relative strength indicator is measured using this relative strength value.

5 Experimental analysis

The market price variation is represented using the OHLC (Open High Low Close) financial chart in the market over time. The chart consists of vertical line where each vertical line displays the price range over a unit time. There are various pattern formations that can be analyzed from the OHLC chart over a period of time. The system is implemented and analysed using the Django web framework. The data source is obtained from a Flask REST API which in turn fetches the live data from an open source data provider called Quandl.

The proposed machine learning system accuracy is calculated by applying the stock market data into k-NN regression, Linear regression and support vector machine algorithms. The accuracy range is represented in Table 2. It has been found that k-NN regression is more accurate compared to linear regression and SVM methods.

NSE stock market values of various companies such as HDFC Bank, TCS, Bajaj Finance, SBI National, etc. are analysed by applying various regression methods. The classifier accuracy is represented in the graph Fig. 3. The graph shows accuracy of k-NN regression is higher compared to Linear regression and SVM algorithms.

The measurements of moving average convergence divergence, exponential growth, relative strength index values calculated for various company stock market prices analyzed and is shown in Table 3. The results range from 0 to 100.

Various machine learning algorithms are applied for various companies stock price values and is shown in Table 4. The stock price original values are compared with the predicted value which yields 90% accuracy.

Stock price movements with its relative strength index of some sample companies are shown in Fig. 4. Time period is represented in the candle stick graph, increase in price is represented in y-axis.

6 Conclusion

In this paper we have done the market trend prediction using the k-NN regression. A set of technical indicators are analyzed from the stock values of various companies is predicted. The result showed a great improvement in the accuracy over the other machine learning algorithms. In future a full-fledged sentiment analysis is to be implemented which help in achieving a greater performance in terms of market trend prediction. There scope of market trend prediction is very wide and in future the trend analysis for a particular type of industry can be analyzed and the production and the revenue of the company can be machine learned. This can be of a great use to various types of industries.

References

- Ahmed ST (2017) A study on multi objective optimal clustering techniques for medical datasets. In: 2017 international conference on intelligent computing and control systems (ICICCS), pp 174–177. IEEE. <https://doi.org/10.1109/ICCONS.2017.8250704>
- Ahmed SST, Thanuja K, Guptha NS, Narasimha S (2016) Telemedicine approach for remote patient monitoring system using smart phones with an economical hardware kit. In: 2016 international conference on computing technologies and intelligent data engineering, pp 1–4. IEEE. <https://doi.org/10.1109/ICCTIDE.2016.7725324>
- Ahmed ST, Sandhya M, Sankar S (2019) A dynamic MooM dataset processing under TelMED protocol design for QoS improvisation of telemedicine environment. *J Med Syst* 43(8):257. <https://doi.org/10.1007/s10916-019-1392-4>
- Ahmed ST, Sandhya M, Sankar S (2020) TelMED: dynamic user clustering resource allocation technique for moom datasets under optimizing telemedicine network. *Wirel Pers Commun*. <https://doi.org/10.1007/s11277-020-07091-x>
- Chandar SK (2019) Fusion model of wavelet transform and adaptive neuro fuzzy inference system for stock market prediction. *J Ambient Intell Humaniz Comput*, pp 1–9 pp. <https://doi.org/10.1007/s12652-019-01224-2>
- Chouhan L, Agarwal N, Ishita P, Saxena S (2018) Stock market prediction using machine learning. In: First international conference on secure cyber puttinging and communications, National Institute of Technology, JALANDHAR, Dec 2018. <https://doi.org/10.1109/ICSCCC.2018.8703332>
- Erdogan Z, Namli E (2019) A living environment prediction model using ensemble machine learning techniques based on the quality of life index. *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-019-01432-w>
- Hiransha M, Gopalakrishnan EA, Menon VK, Soman KP (2018) NSE stock market prediction using deep-learning models. *Elsevier Procedia Comput Sci* 132(2018):1351–1362
- Kadam VJ, Jadhav SM, Vijayakumar K (2019) Breast cancer diagnosis using feature ensemble learning based on stacked sparse autoencoders and softmax regression. *J Med Syst* 43:263. <https://doi.org/10.1007/s10916-019-1397-z>
- Kalra S, Prasad JS (2019) Efficacy of news sentiment for stock market prediction. In: 2019 international conference on machine learning, big data, cloud and parallel computing (Com-IT-Con), India, 14th–16th Feb 2019, pp 491–496
- Martinsson F, Liljeqvist I (2017) Short-term stock market prediction based on candlestick pattern analysis, Thesis
- Nakov P, Ritter A, Rosenthal S, Stoyanov V, Sebastiani F (2016) SemEval-2016 task 4: sentiment analysis in Twitter. In: Proceedings of the 10th international workshop on semantic evaluation, ser. SemEval'16
- Peng Z et al (2019) Stock analysis and prediction using big data analytics. In: 2019 international conference on intelligent transportation, big data & smart City (ICITBS), pp 309–312
- Sadia KH, Sharma A, Paul A, Padhi S, Sanyal S (2019) Stock market prediction using machine learning algorithms. *Int J Eng Adv Technol (IJEAT)* 8(4). ISSN: 2249–8958
- Somani P, Talele S, Sawant S (2014) Stock market prediction using hidden markov model. In: 2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference, 2014, 978-1-4799-4/14, pp 89–92
- Sun T, Wang J, Zhang P, Cao Y, Liu B, Wang D (2017) Predicting stock price returns using microblog sentiment for Chinese stock market. In: 2017 3rd international conference on big data computing and communications (BIGCOM). <https://doi.org/10.1109/BIGCOM.2017.59>
- Usmani M, Adil SH, Raja K, Ali SSA (2016) Stock market prediction using machine learning techniques. In: 2016 3rd international conference on computer and information sciences (ICCOINS), pp 322–327
- Vijayakumar K, Arun C (2017) Automated risk identification using NLP in cloud based development environments. *J Ambient Intell Humaniz Comput*, ISSN 1868-5137. <https://doi.org/10.1007/s12652-017-0503-7>
- Vijayakumar K, Arun C (2019) Continuous security assessment of cloud based applications using distributed hashing algorithm in SDLC. *Cluster Comput*. <https://doi.org/10.1007/s10586-017-1176-x>
- Wen M, Li P et al (2019) Stock market trend prediction using high-order information of time series. *IEEE Trans Big Data Learn Discov* 7(2019):28299–28308
- Zhao J, Sun N, Cheng W (2019) Logistics forum based prediction on stock index using intelligent data analysis and processing of online web posts. ISSN: 1868-5145. <https://doi.org/10.1007/s12652-019-01520-x>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.