

Prediction of stock market using Artificial Intelligence

Akash Patel

K.J. Somaiya College of Engineering
& Information Technology,
Sion, Mumbai
University of Mumbai, India
akash.rp@somaiya.edu

Devang Patel

K.J. Somaiya College of Engineering
& Information Technology, Sion
Mumbai
University of Mumbai, India
devang.sp@somaiya.edu

Seema Yadav

K.J. Somaiya College of Engineering
& Information Technology, Sion
Mumbai,
University of Mumbai, India
syadav@somaiya.edu

Abstract—Stock market is place where people buy and sell shares of publicly listed companies. Every buyer and seller try to predict the stock market price movements to get maximum profits and minimum losses. Using cutting edge technology such as AI can improve prediction stock price. In the procedure of considering strategies and variables to be considered, we found ML algorithmics such as Random forest, LSTM, SVM, ANN was not fully utilized. In this model we will introduce and review more a possible way to predict stock movements with high accuracy. The first thing we considered is data of previous year's share market prices, historical prices of currency and commodity market and the historical news headlines. The datasets were pre-processed and prepared for actual analysis. Therefore, our model will also focus on preprocessing of datasets. Second, after processing the datasets earlier, we will review the use of major AI technique for that data and productive results. In addition, the proposed system evaluates the application of the forecast system to the real-world scenario and the problems associated with the accuracy of the total values provided. The high accuracy and profitability was achieved when results of all algorithms are combined and considered all factors affecting the stock prices. Successful valuation prediction of share price can become a big asset for stock market firms and provide real life solutions to the difficulties faced by stock market individual investors have.

Index Terms—Stocks, AI, data

I. INTRODUCTION

Predicting the Stock Market has been the goal of investors since its existence. Everyday billions of dollars are traded on the exchange, and behind each dollar is an investor hoping to profit in one way or another. Should an investor be able to accurately predict market movements, it offers tantalizing promises of wealth and influence. It is no wonder then that the Stock Market and its associated challenges find their way into the public imagination every time it misbehaves. The share market is a compilation of different people buying and selling the shares. Mostly known as stock (stake) which generally refers to claims of ownerships over a business by an individual or group of individuals. The way of finding the future valuation of the stock market prices is called the stock market estimate. Expected to be Strong, accurate and effective. The system should work in line with real life scenarios and be well connected to that. The movement in the stock market is usually determined by the sentiments of thousands of investors [3]. These events are political events such as the statements of a ministers or government officials, statement of government bodies such as RBI, SEBI, scandal news, etc. It can also be the global happening such as rapid movements in currencies prices and commodities prices [8]. All this thing affects the earning of companies, which ultimately affects the sentiment of stock market investors. This is beyond the reach of almost all individual to assess these accurately and consistently. This method usually requires collection various social media data, news that affect stock market investors sentiment and the feelings expressed by individuals. Other data such as last year's stock prices are

also considered. The relationship between different data points. Is considered and an estimation is done using these variety of data points.

A. Aim

To make the stock market investment process simple, less time consuming and less tedious, and make process of investing in stock market This system will help news investors understand the stock market very quickly and easily. To make profitable trade in the stock market, investors need to predict the trend of stock price, future range of stock price, which requires daily updating of market movements, and keeping eye on daily market data. This whole process makes the small investors or people with constraints due to their routine struggle in the stock market keeping a regular update of such information becomes difficult for them.

B. Objectives

- Increase the accuracy of price prediction
- Make stock market prediction easy and simple
- Provide useful insights to new investors to understand market quickly.
- Reduce the time required to make prediction by providing different data analysis at one point.
- To make the stock market investment process simple.

C. Scope

Predicting stock price range, volatility, risk, reward, trend of stocks, comparison of stock with other peer stocks, comparison with global market will be achieved using stock's open, high, low, close, volume data, price of other commodities related to stock, forex(currency) price. Sentiment analysis of news is performed to find overall sentiment of stock market investors. Combining all these data and different algorithm's result to obtain better accuracy of prediction.

II. LITERATURE SURVEY

A. Related Work

The paper "Machine Learning Approach in Stock Market Prediction" was written by authors Raut Sushrut Deepak, Shinde Isha Uday, Dr D. Malathi. The paper states that a high level of accuracy and precision are the key parameters to be considered while predicting the share prices. Time series analysis, fundamental analysis of companies, Technical analysis are used by many share market individual investors and institutions during the prediction. However, these methods are not completely reliable, so there is a need to provide supporting way to predict the stock market. In this paper, a Machine Learning (ML) method that will be trained in available stock data, acquire intelligence and use the information obtained to accurately predict. After extensive research of various algorithms and their robustness in the various problem domains, the Artificial Neural Network (ANN) was more appropriate than any other algorithm. The Customized neural network model which has wide number of features and parameters makes it easier to apply. An important method used in this paper is to obtain the results of predicting machine learning concepts and the results tested on the Bombay Stock Exchange index data set [7].

The paper "Stock Price Prediction Using Technology Analysis and Mechanical Learning" was written by author Jan Ivar Larsen. The paper states that historical stock prices are used to predict future stock price movements. The used stock price model uses a two-layer consultation approach that uses background information from technical analysis to the first layer of thinking to direct the second layer of thinking based on machine learning. The model is supplemented by a financial management strategy that uses the historical success of the predictions made by the model to determine the amount of money that will invest in future estimates. Based on several portfolio simulations and trading models generated by the model, they conclude that the forecast model successfully exceeds the Oslo Benchmark Index (OSEBX) [9].

The paper "A Machine Learning Model for Stock Market Prediction" was written by authors Omar S. Soliman, Osman Hegazy and Mustafa Abdul Salam. The paper states that the Share market price prediction is an attempt of estimating the future values of a stock prices and other financial instruments that are traded in different exchanges. Successful forecasting of future share prices can increase individual investor's profits. This paper proposes a machine learning model for this assessment of the share market price. The proposed algorithm combines particle swarm optimization and random forest algorithm. The Particle swarm optimization algorithm is used to grow a random forest to estimate prices everyday. This given model works on the study of past dataset of share market and technical indicators. The Particle swarm optimization algorithm selects a combination of the most parameters for random forest integration which avoids further congestion and problem of local minima which can improve prediction accuracy. These presented model tested on multiple financial sets of databases and was compared with the Leuralberg-Marquardt neural network algorithms. The results attained indicated the model proposed here can have better accuracy of prediction and robust Particle swarm optimization algorithm [10].

The paper "Prediction Models for Indian Stock Market" was

written by authors Aparna Nayak, M. M. Manohara Pai and Radhika M. Pai. The paper states that share market price data is produced in large amount and changes on every moment. The share market is a perplex and demanding system where investors can earn more money or lose all money. In this, an attempt of predicting share price trend is done. Two types are made up of one for predicting next day price prediction and one is for prediction of prices after one month. The learning algorithms of the monitored machines are used to build the models. As part of the daily predictive model, historical values are associated with emotions. Up to 70% of perceived accuracy using standardized machine learning techniques in daily forecast models. The monthly forecasting model attempts to assess whether there are any similarities between any two-month course. Tests prove that the trend for at least one month is related to the trend of other month. [11]

The paper "Study on the prediction of stock price based on associated network model of LSTM" was written by authors Guangyu Ding and Liangxi Qin. This paper says The share market has attracted a wide range of individuals and institutions. It has always been a great junction for individuals and Investment firms, constantly adopting changes in the stock market and its trend. Right now there multiple ways to estimate the future share prices. future share price prediction methods can be divided into at least two categories: Mathematical Methods and techniques of AI. Mathematical methods include the systematic model, the ARCH model, and so on. Among the methods of AI are the MLP, the convolutional Neural Network, the Bayes Network, Backward Propagation Network, SVM, Single-Layer Long Short-Term Memory, etc. In order to estimate multiple values as output in a model, it is necessary to design a model that can various values as output parallelly and take various inputs for this purpose, an deep neural network model is proposed based on a short-term memory network with multiple inputs and outputs. This network can simultaneously estimate price, open, high and low price of shares. Convenient network model Compared to the Long Short-Term Memory model and the Deep Neural Network model. Results show that accuracy in estimating multiple values parallelly the associated model has greater accuracy than the other models and Its estimated accuracy is more than 95%. [12]

The paper "Stock Market Forecasts Using Machine Learning" was written by authors Prof. S.P. Pimpalkar, Jenish Karia, Muskaan Khan, Satyamandand, Tushar Mukherjee. The paper states that the construction of the guesswork will use a variety of attributes such as input and will predict whether the market value will be positive or negative. The various attributes used in the model include oil, Foreign Exchange Rate, interest, gold and silver prices, NEWS, twitter news feeds and pattern matching. A variety of ML methods are used including Regression Support Vector Machine and Recurrent Neural Network [5].

The paper "Automated Stock Price Prediction Using Machine Learning" was written by authors Mariam Moukalled, Wassim El-Hajj, Mohamad Jaber. The paper says that traditionally, investors analyze share values and stock indices to predict market movements along with event news related to these shares. Therefore the importance of news in share market. The majority of the models created earlier

focuses on categorizing news as good, bad, neutral and finding their impact on share prices or considering past price changes and estimating their upcoming price changes. In this, they presented an automated trading system that combines statistical method, Artificial Intelligence and other visible factors such as media sentiment with the aim of gaining better stock forecasting and profitable trading. In particular, we aim to determine the valuation or trend of a particular share price for the future date by considering the initial market time. To obtain this objective, they trained conventional machine learning models and developed / trained multiple in-depth reading models taking into account the importance of relevant issues. Various tests were performed, with the highest accuracy (82.91%) obtained using Apple Inc.'s Support Vector Machine stock. (AAPL) stock [14].

The paper "Stock Market Forecasts Using Machine Learning" was written by author V.V Kranthi Sai Reddy. The paper states that stock trading globally is one of the most important. Share market prediction is an attempt to estimate the future prices of other financial instrument traded on a currency exchange. This paper describes stock forecasts using Learning Machines. Technical and basic or time analysis is used by most stockbrokers when making stock forecasts. The language of the system used to predict the stock market using Python learning tools. In this paper we propose a Machine Learning (ML) method that will be trained in available stock data and gain intelligence and use the information obtained to accurately predict. In this context the study uses a machine learning system called the Support Vector Machine to predict the prices of major and minor stocks and three different markets, using daily and minute wave prices [15].

B. Existing System

In the present system the various algorithms used for forecasting can be divided into queues (AR, MA, ARIMA, ARMA) and incompatible models (ARCH, GARCH, Neural Network) and AI framework such as Naive Bayes, the closest neighbours k (k-NN), Support Vector Machine (SVM), Linear Regression, Artificial Neural Network (ANN) and Random Forest were used to advance the gauge model. Points are settled on stock values [].

Current models predict that the stock market uses only one algorithm to predict different conditions and variables and also does not combine multiple algorithm results or consider multiple algorithms to accurately predict. Previous results indicate that the stock price cannot be calculated using traditional divisions. The current system does perform optimally if there is a change in the operating surrounding because it does not focus on external events occurring in the surrounding such as news events and other factors affecting prices such as the Forex and Commodity market. It uses only one data source, thus being extremely biased. The existing system requires some kind of input translation, so it needs to be measured. The existing system uses only historical data or media analysis simultaneously, not used together

The current system does not take into account certain important data such as trade volume and transaction value in the trading volume and the percentage of the amount that are be delivered and the percentage of delivery that predicts an investment or investment that occurs in a particular stock by a major fund manager or large investors. This feature is often overlooked by new investors and existing algorithms, that does not take these factors into account for better market movement analysis. The combination of these factors and volume data can be an important predictor parameter that is not considered much in existing

systems.

Most of the existing systems uses only one algorithm and one data at a time. The existing system also does try to predict share prices in all conditions and on all days but in real world share market can not be predictable every time so certain conditions need to be checked before predicting share market.

III. PROPOSED METHODOLOGY

The system is divided into the following modules:

- Data Collection

Data is collected from various sources, such as yahoo finance and google finance, which are in the form of .csv format. For news data gathering system uses Google News api.

- Analysis Manipulation and Visualization of data

Data collected is cleaned and pre-processed to make it ready to use for mal algorithms and models. and along with this data visualization is also performed [2].

- Build a Model

The cleaned and pre-processed data is used to create, build and train various machine learning algorithms which can be used in predictions [1].

- Predict Outcomes

After the model has been built successfully, the next thing to do is predict an outcome pattern for a particular stock and check the accuracy of the predictions

- Predict combining results of all algorithms

After models are built system combines their results for better prediction and high accuracy and predicts the output based on real time data provided to it

In this proposed system, we aim to predict the future share prices using various machine learning methods. In this proposed system, we were able to train and test the the algorithms of ML from the different data points from the past for making the future share price estimation. We considered EOD (End OF the Day) data of the past years to test and train the ML models [15]. We used various ML libraries and frameworks to obtain the goal system uses Numpy, Pandas, for data manipulation and visualization, Sklearn, Kears, Tensorflow, for machine learning models and math for mathematical operations on data. The other dataset was Historical news headlines used for sentiment analysis. The main library was NumPy library, which is commonly used to manipulate and clean and datasets, to convert it into form which can be directly used for ML models [6]. The other was sklearn library, which is used for actual calculation, estimation and prediction. The data we used was of historical share markets, and was gathered from the various public and open sources available online, and 85 or more percentage of dataset was used to training of the ML models depending upon the need and the rest was used for validation and testing purpose [4]. The common way of the supervised ML model is to analyse and remember the pattern and correlations in the dataset from the train and validation dataset and then produce result for the test dataset [13].

System does feature extraction, various pre-processing of data to make it ready for training of model [13]. The python library named pandas was used for pre-processing data such as it was used to combine various datasets into one dataframe. This prepared data frame data was ready for feature extraction. The features of these dataframe was date, close, open, high,

low, volume, delivery percentage, number of trades, turnover and other derived features. All features was used by us to train machine of a random forest model and predicted the object variable, and we used some selected features to predict using the LSTM, SVM algorithm also which was the values of share of upcoming day [14]. We also measured the accuracy of prediction on the test dataset and on real values. The presented method includes different aspects of research including data manipulation and pre-processing.

System uses ANN (Artificial Neural Network) for sentiment analysis of news data, and Google News API for fetching real time data of news related to stock market

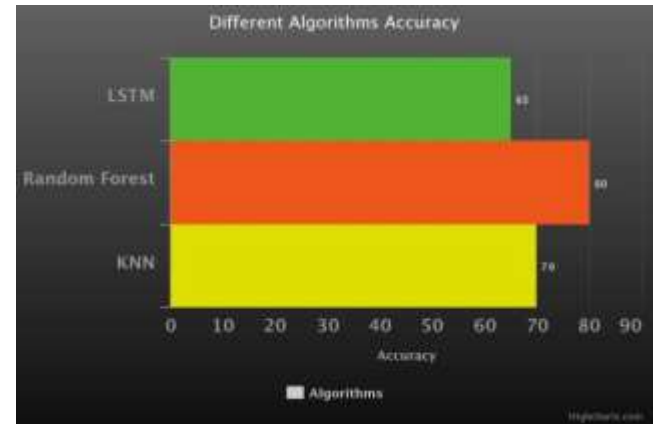


Fig.3. Accuracy for Different Algorithms

Date	Open	High	Low	Close	Volume	Turnover	Trades	Volume/Shareholder
10-08-2021	913.81	919	891	898.8	888.88	2.5450E+14	40822	1221800
11-08-2021	899.9	875.09	872	875.75	827.7	857.58	109154	2.8000E+14
12-08-2021	825.7	793	821	784.85	784.05	888.46	4982007	5.9900E+14
13-08-2021	827.7	793	801	784.85	784.05	888.46	4982007	5.9900E+14
14-08-2021	784.85	824.8	814.8	814.8	822.2	819.22	1091229	2.8000E+14
15-08-2021	884.2	829	882.85	827	849	884.77	2023235	5.2000E+14
16-08-2021	849	838.88	858.81	785.1	800	888.85	528147	4.2000E+14
17-08-2021	800	812.1	829.9	780.25	803.35	829.78	1.2000E+14	4.0712
18-08-2021	809.35	799	800.25	772.1	778.35	783.28	2381275	1.8000E+14
19-08-2021	778.45	786.7	787.9	787.9	782.05	788.12	2128993	5.9900E+14
20-08-2021	787.9	793.28	789.8	782.25	782.85	783.59	1091896	9.2100E+14
21-08-2021	711.55	719.45	742.1	789	737	728.1	4471003	5.2000E+14
22-08-2021	787	788	784.8	793	727.8	788.83	3453076	2.8000E+14
23-08-2021	727.8	727	728.8	885.85	885.15	728.3	3298702	2.1000E+14
24-08-2021	705.75	711.05	711.8	705.2	714.25	714.05	3462963	5.4700E+14
25-08-2021	714.25	712.1	712.45	889.8	886.2	714.38	3085183	2.1000E+14
26-08-2021	889.8	712	784.8	789	788.85	782.19	1021896	2.1000E+14
27-08-2021	789.85	781.5	784.8	785.85	782.5	788.53	3211896	2.1000E+14
28-08-2021	782.5	788	783.8	787	784.75	783.34	3021188	2.8000E+14
29-08-2021	784.75	788	789	781.2	781.2	788.8	3088008	1.1000E+14
30-08-2021	781.2	788.2	771.8	788.2	788.8	784.44	1774445	5.4000E+14
31-08-2021	788.2	775.28	780.45	784.15	788.2	788.83	3088008	2.2000E+14
01-09-2021	784.15	781	788	788.85	788.85	788.85	3088008	2.2000E+14
02-09-2021	788.85	781.4	781.4	788.8	788.8	788.8	3088008	2.2000E+14
03-09-2021	788.8	781.4	781.4	788.8	788.8	788.8	3088008	2.2000E+14
04-09-2021	788.8	781.4	781.4	788.8	788.8	788.8	3088008	2.2000E+14
05-09-2021	788.8	781.4	781.4	788.8	788.8	788.8	3088008	2.2000E+14
06-09-2021	788.8	781.4	781.4	788.8	788.8	788.8	3088008	2.2000E+14
07-09-2021	788.8	781.4	781.4	788.8	788.8	788.8	3088008	2.2000E+14
08-09-2021	788.8	781.4	781.4	788.8	788.8	788.8	3088008	2.2000E+14
09-09-2021	788.8	781.4	781.4	788.8	788.8	788.8	3088008	2.2000E+14
10-09-2021	788.8	781.4	781.4	788.8	788.8	788.8	3088008	2.2000E+14

Fig. 4. Stock market historical Data of stocks



Fig. 5. Prediction results of Random forest algorithm

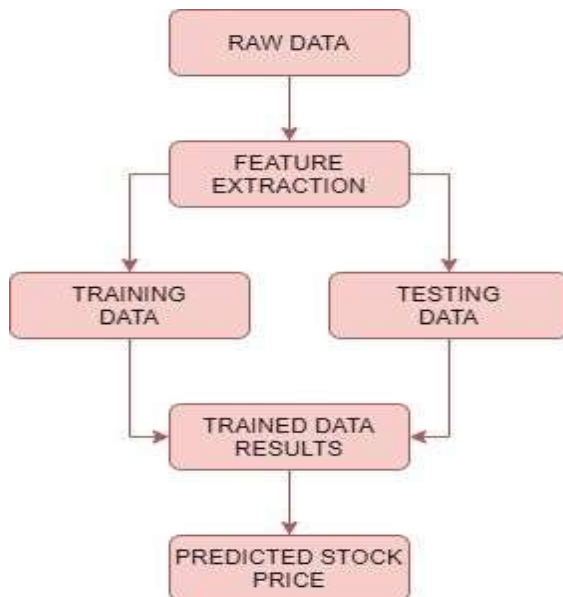


Fig. 1. System Architecture

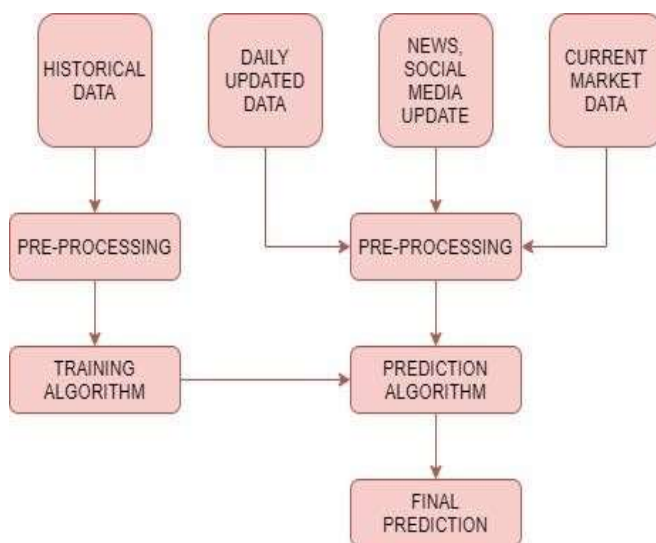


Fig. 2. Stock Price Prediction

IV. CONCLUSION AND FUTURE PLAN OF WORK

A. Conclusion

By estimating the parameters of accuracy, it was found that the most appropriate method for stock market prediction is to use multiple algorithms such as ANN, Random Forest, SVM, LSTM and combine their results for predictions, and also considering all the factors affecting the stock market price such as news sentiment, currency, commodity prices and other international stock exchange market data. High accuracy was achieved by doing all this. Considering other factors affecting the stock price improves accuracy of prediction. Trying to predict the stock market only when its more predictable and avoiding when market is more uncertain gives greater profits,

such as prediction only when standard deviation was within the threshold values improved accuracy of our algorithm by 10 or more percentage. The algorithms are an extraordinary resource for investors and financial institutions or investment in share market as they are trained on a huge collection of past data and have been picked subsequently to be tested on sample data.

B. Future scope

Using more granular data such as data of every minute, every second or tick by tick data can improve accuracy of prediction as it contains more insightful information than one day data. Future work contains providing predictions and recommendations for individual according to its risk and reward appetite, and also providing a portfolio management to mitigate the risk associated with investing in stock market. Process of placing orders and making trades can be automated which work according to suggestions given by AI and ML algorithms, which can remove human errors and factors such as fear and greed of human being which can result in faster and more profitable trading.

REFERENCES

- [1] Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017.
- [2] Loke.K.S. "Impact Of Financial Ratios And Technical Analysis On Stock Price Prediction Using Random Forests", IEEE, 2017.
- [3] Xi Zhang¹, Siyu Qu¹, Jieyun Huang¹, Binxing Fang¹, Philip Yu², "Stock Market Prediction via Multi-Source Multiple Instance Learning." IEEE 2018.
- [4] Vivek Kanade, Bhausaheb Devikar, Sayali Phadtare, Pranali Mande, Shubhangi Sonone. "Stock Market Prediction: Using Historical Data Analysis", IJARCSSE 2017.
- [5] Sachin Sampat Patil, Prof. Kailash Patidar, Asst. Prof. Megha Jain, "A Survey on Stock Market Prediction Using SVM ", CTET 2016.
- [6] Hakob GRIGORYAN, "A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA)", DSJ 2016.
- [7] Raut Sushrut Deepak, Shinde Ishu Uday, Dr. D. Malathi, "Machine Learning Approach In Stock Market. Prediction", IJPAM 2017.
- [8] Pei-Yuan Zhou, Keith C.C. Chan, Member, IEEE, and Carol Xiaojuan Ou, "Corporate Communication Network and Stock Price Movements: Insights From Data Mining", IEEE 2018.
- [9] Jan Ivar Larsen, Norwegian University of Science and Technology "Predicting Stock Prices Using Technical Analysis and Machine Learning"
- [10] Osman Hegazy, Omar S. Soliman and Mustafa Abdul Salam A Machine Learning Model for Stock Market Prediction", Researchgate 2013
- [11] Prediction Models for Indian Stock Market" is written by the authors Aparna Nayak, and Radhika M. Pai. M. M. Manohara Pai., Researchgate 2016
- [12] "Study on the prediction of stock price based on the associated network model of LSTM Liangxi Qin, and ", Guangyu Ding Springer 2020
- [13] " Stock Market Prediction using Machine Learning" , authors Prof. S .P. Pimpalkar, Jenish Karia, Muskaan Khan, Satyam Anand, Tushar Mukherjee, IJETCS
- [14] Automated Stock Price Prediction Using Machine Learning", Mariam Moukalled, Wassim El-Hajj, Mohamad Jaber, EPLIU
- [15] Stock Market Prediction Using Machine Learning" V Kranthi Sai Reddy, Researchgate

