Machine Learning Model on Stock Price Prediction

1st Vanshika Chanderiya 2nd Sachit Patel 3rd Nisarg Oza 4th Pranay Shah 5th Nitin Joshi 6th Sandesh Rao

***Abstract*—Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on a financial exchange. The successful prediction of a stock's future price will maximize investor's gains. This paper proposes a machine learning model to predict stock market price. It is complicated to estimate the stock market where relations between input and output are random in nature. Predicting the cost of the share market is the most complex job of the financial time series. Forecasting of the stock should be possible by utilizing the present and past information available on the market, growing trends and daily news updates. There are various algorithms and models in machine learning such as Regression, SVM, Dimensionality Reduction algorithms, Gradient reduction algorithms, fuzzy logic, etc. that are used for prediction. For predicting the price of the stock various parameters were taken into consideration such as date, open, high, low, last, close, total trade quantity, turnover, etc. The execution measurements that should be achieved if there should be an occurrence of the stock forecast are exactness, adaptability and less time utilization. There are numerous sorts of research done as such far with the end goal to predict the stock market to complete the characterized measurements. This paper uses the accessible advancements in the field alongwith extemporizing the algorithms and making it user-friendly and customisable. This will enable the client to set the prefrences, amount, timeframe and alert updates and will predict the outcome which otherwise was not possible in the market.**

***Index Terms*—Machine Learning, Stock Market, Data Mining, Long Short-Term Memory and Stock Price Prediction**

I. Introduction

The stock market refers to the collection of markets and exchanges where regular activities of buying, selling, and issuance of shares of publicly-held companies take place. A stock in, general represents ownership claims on business by a particular individual or a group of people. The attempt to determine the future value of the stock market is known as a stock market prediction. The successful prediction of a stock's future price could yield significant profit and is expected to be robust, accurate and efficient. The system must work according to the real-life scenarios and should be well suited to real-world settings. There are various methods and ways of implementing the prediction system like Fundamental Analysis, TechnicalAnalysis, Machine Learning, and Time-series aspect structuring. With the advancement of the digital era, the prediction has moved up into the technological realm. The most prominent and promising technique involves the use of Artificial Neural Networks, Recurrent Neural Networks, which is the implementation of machine learning. Machine learning involves artificial intelligence which empowers the system to learn and improve from past experiences without being programmed time and again. The methods used to predict the stock market includes a time series forecasting along with technical analysis, machine learning modeling and predicting the variable stock market.

The stock market is highly volatile. At the most fundamental level, it is said that supply and demand in the market determine stock price. But, it does not follow any fixed pattern and is also affected by a large number of highly varying factors.The investors on Wall Street are split in two largest factions of adherents; those who believe the market cannot be predicted and those who believe the market can be beaten.Recently, a lot of interesting work has been done in the area of applying Machine Learning Algorithms for analyzing price patterns and predicting stock price. Most stock traders nowadays depend on Intelligent Trading Systems which help them in predicting prices based on various situations and conditions. Recent researches uses input data from various sources and multiple forms. Some systems use historical stock data, some use financial news articles, some use expert reviews while some use a hybrid system that takes multiple inputs to predict the market.Also, a wide range of machine learning algorithms is available that can be used to design the system. These systems have different approaches to solve the problem. Some systems perform mathematical analysis on historic data for prediction while some perform sentiment analysis on financial news articles and expert reviews for prediction. However, because of the volatility of the stock market, no system has a perfect or accurate prediction.

The datasets of the stock market prediction model include details like the closing price opening price, the data and various other variables that are needed to predict the price of stock in a given day. Stock market prediction outperf when it is treated as a regression problem but performs well when treated as a classification. The aim is to design a model that gains from the market information utilizing machine learning strategies and gauge the future patterns in stock value development. The Support Vector Machine (SVM) can be used for both classification and regression. In this technique, we plot every single data component as a point in n dimensional space (where n is the number of features of the dataset available) with the value of feature being the value of a particular coordinate and, hence classification is performed by finding the hyperplane that differentiates the two classes explicitly.

II. Problem Definition

Stock market prediction is basically defined as trying to determine the stock value and offer a robust idea for the people

to know and predict the market and the stock prices. It is generally presented using the quarterly financial ratio using the dataset. Thus, relying on a single dataset may not be sufficient for the prediction and can give a result which is inaccurate. Hence, we are contemplating towards the study of machine learning with various datasets integration to predict the market and the stock trends. The problem with estimating the stock price will remain a problem if a better stock market prediction algorithm is not proposed. Predicting how the stock market will perform is quite difficult. The movement in the stock market is usually determined by the sentiments of thousands of investors. Stock market prediction, calls for an ability to predict the effect of recent events on the investors. These events can be political events like a statement by a political leader, a piece of news on scam etc. It can also be an international event like sharp movements in currencies and commodity etc. All these events affect the corporate earnings, which in turn affects the sentiment of investors. It is beyond the scope of almost all investors to correctly and consistently predict these hyperparameters. All these factors make stock price prediction very difficult. Once the right data is collected, it then can be used to train a machine and to generate a predictive result.

III. Literature Review

The performance of the stock market and it's prediciton is still a difficult job to do inspite of the various technologies available. Everything that can be used to predict a company's future has already been used by investors and is 'baked into' today's price. Only unforeseen events will cause stock prices to change. And, unfortunately, unforeseen events -good and bad- happen every day.There are various factors that affect the prediciton of the stock market. The analysis of the stock market is basically classified into two main parts:Fundamental Analysis and Technical Analysis. – Fundamental Analysis involves analyzing the company's future profitability on the basis of its current business environment and financial performance.(c) – Technical Analysis, on the other hand, includes reading the charts and using statistical figures to identify the trends in the stock market.(c) There are various algorithms available in the market for the prediction of the stock market.They are discussed below:

A. Average

Average refers to the arithmetic mean and is used in our day to day lives. For instance, calculating the average marks to determine overall performance, or finding the average temperature of the past few days to get an idea about today's temperature – these all are routine tasks we do on a . So this is a good starting point to use on our dataset for making predictions.

The predicted closing price for each day will be the average of a set of previously observed values. Instead of using the simple average, we will be using the moving average technique which uses the latest set of values for each prediction. In other words, for each subsequent step, the predicted values are taken into consideration while removing the oldest observed value from the set.

INFERENCE: The results obtained by this algorithm are not very promising.

B. Linear Regression

The most basic machine learning algorithm that can be implemented on this data is linear regression. The linear regression model returns an equation that determines the relationship between the independent variables and the dependent variable.

The equation for linear regression can be written as:

Y=1x1 + 2x2 + 3x3 +.....+ nxn

Here, x1, x2,....xn represent the independent variables while the coefficients 1, 2, ... n represent the weights.

INFERENCE: Linear regression is a simple technique and quite easy to interpret, but there are a few obvious disadvantages. One problem in using regression algorithms is that the model overfits. Instead of taking into account the previous values from the point of prediction, the model will consider the value from the same date a month ago, or the same date/month a year ago.

C. K-Nearest Neighbours

Another interesting ML algorithm that one can use here is kNN (k nearest neighbours). Based on the independent variables, kNN finds the similarity between new data points and old data points.

INFERENCE: Regression algorithms have not worked well on our dataset.

D. Auto Arima

ARIMA is a very popular statistical method for time series forecasting. ARIMA models take into account the past values to predict the future values. There are three important parameters in ARIMA:

p (past values used for forecasting the next value) q (past forecast errors used to predict the future values) d (order of differencing)

INFERENCE: As we saw earlier, an auto ARIMA model uses past data to understand the pattern in the time series. Using these values, our model captured an increasing trend in the series. Although the predictions using this technique are far better than that of the previously implemented machine learning models, these predictions are still not close to the real values.

E. Prophet

There are a number of time series techniques that can be implemented on the stock prediction dataset, but most of these techniques require a lot of data preprocessing before fitting the model. Prophet, designed and pioneered by Facebook, is a time series forecasting library that requires no data preprocessing and is extremely simple to implement. Prophet tries to capture the seasonality in the past data and works well when the dataset is large.

INFERENCE: Prophet (like most time series forecasting techniques) tries to capture the trend and seasonality from past data. This model usually performs well on time series datasets, but fails to live up to it's reputation in our case.

As it turns out, stock prices do not have a particular trend or seasonality. It highly depends on what is currently going on in the market and thus the prices rise and fall. Hence forecasting techniques like ARIMA, SARIMA and Prophet would not show good results for this particular problem.

F. LSTM

LSTMs are widely used for sequence prediction problems and have proven to be extremely effective. The reason they work so well is because LSTM is able to store past information that is important, and forget the information that is not. LSTM has three gates:

The input gate: The input gate adds information to the cell state The forget gate: It removes the information that is no longer required by the model The output gate: Output Gate at LSTM selects the information to be shown as output

INFERENCE:

The LSTM model can be tuned for various parameters such as changing the number of LSTM layers, adding dropout value or increasing the number of epochs. But are the predictions from LSTM enough to identify whether the stock price will increase or decrease? Certainly not!

IV. Proposed System

In this proposed system, we focus on predicting the stock values using machine learning algorithms like Long Short-Term Memory Networks. We proposed the system "Stock market price prediction" we have predicted the stock market price using the LSTM algorithm. In this proposed system, we were able to train the machine from the various data points from the past to make a future prediction. We took data from the previous year stocks to train the model. We majorly used two machine-learning libraries to solve the problem. The first one was LSTM, which is well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. The other was tensorflow, which is uesd to create dataflow graphs—structures that describe how data moves through a graph, or a series of processing nodes. The data set we used was from the previous years stock markets collected from the public database available online, 80

V. Mehtodologies

Various types of neural networks can be developed by the combination of different factors like network topology, training method etc. For this experiment, we have considered Recurrent Neural Network and Long Short-Term Memory. This section we will discuss the methodology of our system. Our system consists of several stages which are as follows: Stage 1: Raw Data: In this stage, the historical stock data is collected from https://www.quandl.com/data/NSE and this historical data is used for the prediction of future stock prices.

Stage 2: Data Preprocessing: The pre-processing stage involves a) Data discretization: Part of data reduction but with particular importance, especially for numerical data b) Data transformation: Normalization. c) Data cleaning: Fill in missing values. d) Data integration: Integration of data files. After the dataset is transformed into a clean dataset, the dataset is divided into training and testing sets so as to evaluate. Here, the training values are taken as the more recent values. Testing data is kept as 5-10 percent of the total dataset.

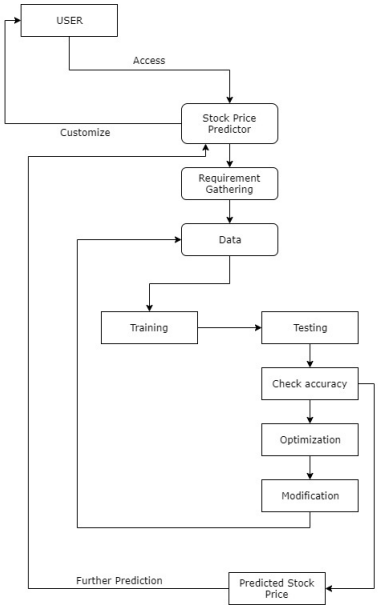
Stage 3: Feature Extraction: In this layer, only the features which are to be fed to the neural network are chosen. We will choose the feature from Date, open, high, low, close, and volume.

Stage 4: Training Neural Network: In this stage, the data is fed to the neural network and trained for prediction assigning random biases and weights. Our LSTM model is composed of a sequential input layer followed by 2 LSTM layers and dense layer with ReLU activation and then finally a dense output layer with linear activation function.

Stage 5: Output Generation: In this layer, the output value generated by the output layer of the RNN is compared with the target value. The error or the difference between the target and the obtained output value is minimized by using back propagation algorithm which adjusts the weights and the biases of the network.

VI. System Architecture

A. Architectural Design



1.User Visits the Website/WebApp

2.Previously saved model is loaded

3.User asks for a company's stock data

4.He requests for prediction to be made

5.The stock price prediction system trains the model using data from the database.

6.The model is saved for future use and closing price is predicted.

7.Result is displayed along-with the graph.

VII. Experimental Results

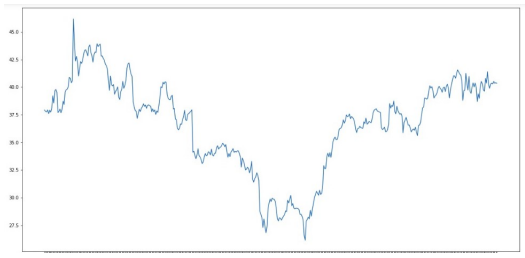
Dataset description: We acquired the data from https://www.quandl.com. We have collected the historical stock data of various companies like Google, Amazon from the National stock exchange.We have collected daily dataset and kept 80percent as training set and 20percent as testing set.

1 (2).png

	Date	Open	High	Low	Last	Close	Total Trade Quantity
0	2018-10-08	208.00	222.25	206.85	216.00	215.15	4642146.0
1	2018-10-05	217.00	218.60	205.90	210.25	209.20	3519515.0
2	2018-10-04	223.50	227.80	216.15	217.25	218.20	1728786.0
3	2018-10-03	230.00	237.50	225.75	226.45	227.60	1708590.0
4	2018-10-01	234.55	234.60	221.05	230.30	230.90	1534749.0
...	...	...	...	...	...	...	...
1230	2013-10-14	160.85	161.45	157.70	159.30	159.45	1281419.0
1231	2013-10-11	161.15	163.45	159.00	159.80	160.05	1880046.0
1232	2013-10-10	156.00	160.80	155.85	160.30	160.15	3124853.0

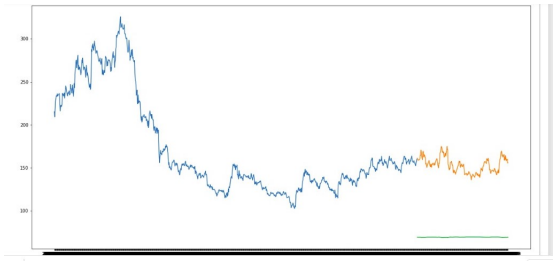
Fig. 1. 1235 X 8 columns

Training Detail: For training the model we used RMSprop as the optimizer and normalized each vector of the sequence. We used Google cloud engine as a training platform [Machine type: n1-standard-2 (2 vCPUs, 7.5 GB memory), CPU platform: Intel Ivy Bridge] and used Ubuntu 16.04, Keras (Frontend) and Tensorflow (Backend) as the learning environment.For our experiment, we have used a various set of parameters with a different number of epochs to measure the RMSE of Training and Testing dataset.



2.png

Fig. 2. Before Prediction



1.png

Fig. 3. After Prediction

In this work, we have used one of the most precise forecasting technology using Recurrent Neural Network and Long Short-Term Memory unit which helps investors, analysts or any person interested in investing in the stock market by providing them a good knowledge of the future situation of the stock market.

VIII. Conclusion

The popularity of stock market trading is growing rapidly, which is encouraging researchers to find out new methods for the prediction using new techniques. The forecasting technique is not only helping the researchers but it also helps investors and any person dealing with the stock market. In order to help predict the stock indices, a forecasting model with good accuracy is required.The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models.

References

[1] RakhiMahant, TrilokNathPandey, Alok Kumar Jagadev, and SatchidanandaDehuri —Optimized Radial Basis Functional Neural Network for Stock Index Prediction,‖ International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - 2016.

[2] Kai Chen, Yi Zhou and FangyanDai —A LSTM-based method for stock returns prediction: A case study of China stock market,‖ IEEE International Conference on Big Data,2015.

[3] A.U.S.S Pradeep, SorenGoyal, J. A. Bloom, I. J. Cox, and M. Miller, —Detection of statistical arbitrage using machine learning techniques in Indian Stock market,‖ IIT Kanpur, April 15, 2013.

[4] Ashish Sharma, Dinesh Bhuriya, Upendra Singh. ”Survey of Stock Market Prediction Using Machine Learning Approach”, ICECA 2017.

[5] Loke.K.S. “Impact Of Financial Ratios And Technical Analysis On Stock Price Prediction Using Random Forests”, IEEE, 2017.

[6] Xi Zhang1, Siyu Qu1, Jieyun Huang1, Binxing Fang1, Philip Yu2, “Stock Market Prediction via Multi-Source Multiple Instance Learning.” IEEE 2018.

[7] VivekKanade, BhausahebDevikar, SayaliPhadatare, PranaliMunde, ShubhangiSonone. “Stock Market Prediction: Using Historical Data Analysis”, IJARCSSE 2017.