Financial Series Prediction During COVID-19 using Multivariate LSTM

Monalisa Sha
Computer Science Dept
California State University
Fresno, CA
monsha@mail.fresnostate.edu

Prachiti Kulkarni
Computer Science Dept
California State University)
Fresno, CA
prachitik@mail.fresnostate.edu

Mattina Mano Manoharan
Computer Science Dept
California State University
Fresno, CA
mattina_mano@mail.fresnostate.edu

Jatin Taneja
Computer Science Dept
California State University
Fresno, CA
jatin1609@mail.fresnostate.edu

Albert Tjong
Computer Science Dept
California State University
Fresno, CA
rafael_albert27@mail.fresnostate.edu

John Her
Computer Science Dept
California State University
Fresno, CA
johndh@mail.fresnostate.edu

I. Abstract

Stocks keep changing all the time and stock market generates large amount of data. Not only knowledgeable investors but also beginners are interested in the study of prediction of stock prices. For wise and successful investment, all are curious to know the future of stock market. Currently, investors are skeptical to invest in stock market due to global pandemic situation. In such case, optimal prediction of stocks guides investors and traders by showing them future trends of stocks they are interested in, by considering all the possible factors that may affect stocks.

In this paper, we propose a multivariate LSTM approach to predict stock prices considering COVID-19 situation along with multiple factors for stock volatility. By taking the historical activities of stock market transactions of all the companies listed under SP500 index ,we compared the results of existing methods like k-nearest, moving average, univariate LSTM with our proposed method to analyze future of stocks . The proposed model outperforms other methods and stands the best fit for predicting future stock trends during any pandemic.

Keywords– Univariant, Multivariant, Moving Average(MA), K-Nearest Neighbor(KNN), Autoregressive integrated moving average(ARIMA), RSI, MACD, BBands, Root Means Square Error,Long Short Term Memory(LSTM),Coronavirus disease 2019(COVID-19),United Airlines(UAL)

II. Introduction

Investing in a stock or an act of buying stock to gain benefit or profit is a common thing nowadays. Based on PolitiFact, a lot of people are invested in the stock market, but most of them are of higher income because they have little risk to their economy when they invest in a large sum. The normal tendency is to buy a stock based on the prices and sell them when the prices are quite high. However, people still lack a lot of knowledge that is based around the stock market if they are not well experienced in the field or never have touched the stock market at all. This project ensures to help people understand how stock market is behaving considering all the possible events so that they can decide whether to invest in the stock market or not. So, even though they have low knowledge or even no knowledge at all in the stock market, this project would guide to look over the trends and learn the behaviors. There already exist the studies on stock market data,predicting the stock prices using various methods or factors and many predictions were close or similar to the right answer.

To help people to invest in the stock market, this project will predict what will mostly happen in the stock market based on the historical data we have for the years.Furthermore, the project is aimed to guide people who have high experience in the field or people who have little to no experience at all in the field to invest in the stock market and help them see the upcoming trends for the stocks that they bought.

There are a lot of questions that we are trying to answer regarding this topic of our project. One of the main questions or the biggest reason for this project is, “What will happen to the company where I have invested in?” Concerning this question, there are other questions that are similar, “Are the prices of my stocks going to go up or down?” , “What is the status or update of a certain stock?” and “Will the company that I am investing in will produce profit?”. These are the questions that are usually asked by the usual investors of the stocks. At present, many people would have questions like, ”What will happen to my stocks if this COVID-19 situation gets worse or better?". Many of these questions can be answered from this project using the most optimized way.

It is assumed that this project will predict the behavior of the stocks whether the prices will go up or down and predict the outcome or the future of the stock prices based on the past data that we used. The data used is collected from the New York Stock Exchange(NYSE) and the National Association of Securities Dealers Automated Quotations(NASDAQ). Yahoo Finance and 'Alpha Vantage Api' is also used to have industry-wise stocks data. The data used in this project, is the complete data from the past 5 years for all the companies under NYSE and NASDAQ, containing all the information needed from everyday activities of stocks, excluding weekends as market is closed on weekends.

The research is conducted after cleaning the data and creating multiple analyses of the data, for example explanatory analysis, descriptive analysis, and prescriptive analysis. Many methods like Moving average, k-nearest neighbour, Univariate LSTM were used on the data to perform the analysis. Comparing results received from these various methods, the most optimal prediction was achieved using the LSTM or Long Short-Term Memory method, since it analyzed the data and predicted it similarly to the factual data based on the past 5 years of the companies. The method results the trends or behavior of the stocks regarding the company that was selected. The project used the top 5 companies that are popular nowadays and made it as an example for the LSTM.However, the global pandemic COVID-19 has affected the stock market and predicting stocks during such situation is quite challenging, as such none of these existing methods consider data of pandemic to predict the stocks. This project proposes the multivariate LSTM learning model which considers the COVID-19 data as one of the variants along with technical indicators like Bolling bands, the Relative Strength Index and Moving average convergence and divergence to predict the trends of stock. This method meets the gap that is present in earlier mentioned methods in which how the pandemic is growing and affecting stocks data is not considered.

As a result, the method predicted a trend or behavior of the stock price of the company that was selected by the user. This method fits the best for such data-sets and gives optimal results. After the results or the trend has been achieved, we also prescribe or give solutions or five tips regarding the trends or how you would invest in the stock market. The solution was achieved by using the Stochastic Optimization method. The method was selected because the data set has many instabilities. As a final result, we can state that the stock prices are going to drop if there is less activity on the stock market or any pandemic occurs like COVID-19, also news and statuses on the company are beneficial regarding the stock prices, as an example, if the company is having a scandal or bad reputation, then the stock prices will automatically drop for that company. This is the most optimal time to invest in a good company and earn maximum profit.

III. Literature Review

For our experiment, we decided to utilize four different methods to achieve our goal. The models we decided to adapt and pit against each other are the Moving Average (MA), K-Nearest Neighbor (KNN), Univariate Long Short-Term Memory (ULSTM), and the Multivariate Long Short-Term Memory (MLSTM).

Before we begin, what is the difference between the four? ULSTM and MLSTM are not much different as seen in their names, but KNN and MA don't share any similar items within the context of their names. Each model is a type of Artificial Intelligence. The reason why we decide to utilize AI techniques is because it has been proven that it's effective against non-linear relationships between variables[13]. And the stock market encompasses a lot of non-linear variables. Machine learning and deep learning are a sub-field of AI. All four methods are either a Machine learning or Deep learning algorithm. With both types of learning techniques, predictions can be made on rule-less programming[10]. Initially, simple regression models and algorithms were utilized to predict these outcomes. But with the introduction of machine and deep learning, these predictions became even more powerful[32]. Some predictions even reached an 80% accuracy on certain stocks like Apple[14]! However MA and KNN are different from ULTSM and MLSTM.

Moving average and K-Nearest Neighbor are models from machine learning. While the latter two, ULSTM and MLSTM are both deep learning models. Due to both types of learning algorithms being a sub-field of AI, people often confuse the two[29]. Machine learning is an unsupervised or supervised algorithm that recursively learns from presented data to make decisions and predictions. It involves the training of the algorithm to learn and see patterns that were not apparent. The amount of data exposed to the machine learning algorithm is in proportion to how much the algorithm will learn[27]. Deep learning, on the other hand, is modeled after neural networks. Similar to the brain. The algorithms are implemented in layers to form a hierarchy[28].

Initially before learning algorithms were introduced to the stock market world, participants were using a "buy-and-hold" method to determine market benchmarks of if a stock is worth profits. The buy-and-hold method was pitted against five different versions of the Moving Average algorithm [22] . And the conclusion was that MA was far superior. Due to MA being able to factor in a time-varying volatility factor and adjusting to it. Though MA may be better than the buy-and-hold method it still has an abundance of issues. One of which is that when it was applied to the India Stock Exchange, there were varying results and predictions[24]. The reason was because most MA did not factor in the values of trading costs. This flaw was considered and further implemented into the MA algorithm so it would consider breaking even when predictions of stocks were to decline. But even with this, most predictions only yielded a 1% profit. Though this is a good safety net to have, the results were sometimes skewed. The results using certain methodological assumptions, like you may break even by jumping ships now, would make the learning algorithm superior but also biased towards desired outcomes[25].

K-Nearest Neighbor is also classified as a machine learning algorithm. In comparison to AM, KNN has a prediction accuracy of 14.7% above AM[23]. K-Nearest Neighbor involves the usage of regression. We initially begin by choosing a point to query. After determining that point, the algorithm will determine the value of the point by utilizing the averages of surrounding neighbor points. The closer the neighbors are, the more weight they'll have in contributing to the average value. There's a plethora of calculations to measure the distance, but the most suitable method is the Euclidean distance[23]. A study was done by applying the KNN to the LQ45 Stock index (an index with stocks under good financial status). The conclusion was that the results were simply not good or accurate. The main reasons were because the stocks weren't scaled and it was hard to choose the appropriate K value[23]. Even though KNN does perform better than AM it still gets outperformed by deep learning algorithms. Another study was done to compare KNN with an ANN(Artificial Neural Network)[26]. The results showed that a neural network, also known as deep learning, had superior performances. Though it should be noted that KNN did have a shorter training time in comparison to the ANN.

Both ULSTM and MLSTM are a deep learning algorithm incorporating recurrent neural networks[18]. It is different from an ANN type deep learning algorithm. Even though both are a form of deep learning, LSTM performs better. The reason as to why LSTM generally performs better than an ANN is because LSTM maintains its own cell state while concurrently storing other new memories. The stored memory can last for a long time while the program still makes predictions. For an ANN, only activation information is stored[18]. Another comparison can be made with LSTM and MA. Unlike the MA Model, LSTM does not require specific expert knowledge to impose restrictions to provide accurate predictions. LSTM can utilize simple gating functions to determine long-term dependencies[16]. LSTM has proven to be accurate at predicting opening prices of stocks[31]. Deep learning algorithms have been one of the leading AI in recent stock market predictions.

One big reason as to why we decided to utilize the NYSE and NASDAQ as our main source of stock market data is because both have underlying similar dynamic properties. They both are also two of the top leading stock markets as well[22]. But even by utilizing AIs, it's still difficult to predict future losses or profits[12]. Though our data set contains an abundant source of data, it does not, however, include data that may be affected by social sentiments or natural crisis. Another study applied multiple learning algorithms to the stock market. And what was able to be concluded was that no matter the accuracy the predictors may become from training over numerical data, they won't achieve absolute accuracy without factoring in social sentiments [6]. As of currently, one of the biggest factors as to how the stock market is behaving right now is because of COVID-19. Within our data, noticeable declines in certain stocks can be seen. Recent research has also supported the conclusion that COVID-19 has drastically affected the Stock Market[30]. Showing that equity in Markets within America and the EU has dropped down up to 30%! A reason as to why we're utilizing LSTM over the other AI algorithms is because LSTM performs more accurately, even in a financial crisis. Another type of accurate deep neural network is a random forest, a deep neural network algorithm that focuses on classification. A random forest utilizes multiple learning trees to make decisions. LSTM was able to outperform the random forest. Giving more accurate results. LSTM was able to also extract useful information from noisy information[17].

We will be utilizing two forms of LSTM. Univariate and multivariate. We propose that multivariate will outperform the univariate because a univariate will only consider one variable that changes while a multivariate considers multiple changing variables.

IV. Methods

A. *Background*

For stock prediction, Machine Learning approaches specifically supervised learning have shown great promises. Here we have used a machine learning method to evaluate and forecast stocks based on the COVID-19 situation. Time series of stock market forecasting is a sequence of events such as daily sales volumes and stock closing prices. Here we have discussed the methods and some interesting results we found and highlighted their strengths or limitations for the stock market.

*a) **Moving Average(MA):*** One subset of mathematical methods falling within the field of univariate analysis is the Moving Average. The moving average(MA)[3] is a simple technical analysis tool, which creates a constantly updated average price to smooth out the different price data. We calculated MA for amazon over a specific time period using 10 days, 20 days, and 50 days to identify trend direction, and to get a basic idea of which way the price is moving. An angled upward MA curve indicates the price is moving up. An angled downward curve indicates the price is moving down overall. And if it is moving sideways, then the price is likely in a range.

$$SMA = \frac{\sum_{i-1}^{N} CLOSE(i)}{N} \tag{1}$$

(MA calculation for Close price)

The MA acts as a floor, so the price bounces up off of it. In a downtrend, a MA may act as resistance. The price hits the MA level and then starts to drop again. One strategy we applied while calculating MA is we took two moving averages to a chart. One longer and one shorter. We calculated the rolling simple moving averages (SMA) of the three-time series and when calculating the M days of SMA, the first M1 is not valid. Because we require the first moving average data point for M prices. When the[8] shorter-term MA crosses the longer-term MA, it's a signal to buy as it indicates that the trend is shifting up. This is known as a "golden cross." Meanwhile, when the shorter-term MA crosses below the longer-term MA, it's a signal to sell, as it indicates that the trend is shifting down. This is known as a "dead/death cross."

*b) **k-nearest neighbors (KNN)**:* KNN algorithm is a supervised machine learning algorithm that we used to solve both classification and regression problems.The KNN algorithm is used to measure the distance between the given test instance and all the instances in the data set. We first start by choosing the K closest instance and then predicting the class value based on these nearest neighbors. We divided the data into two parts - the training data, which the algorithm bases its predictions on, and the test data, which the algorithm makes the predictions about. The test data consists of the values that are being predicted with the algorithm. We divided the training data into vectors and then we calculated the distance from the test data to its neighbor using euclidean distance. Euclidean distance is the square root of the sum of the squared differences between a new point (x) and an existing point (xi) across all input attributes(j).

$$ED(x, xi) = sqrt(sum((xj–xij)^2)) \quad (2)$$

The prediction we made for the regression problem using KNN is based on the mean of the K-most similar instances. This can be seen in Figure 1, where the nearest neighbors to B are selected using the euclidean distance.

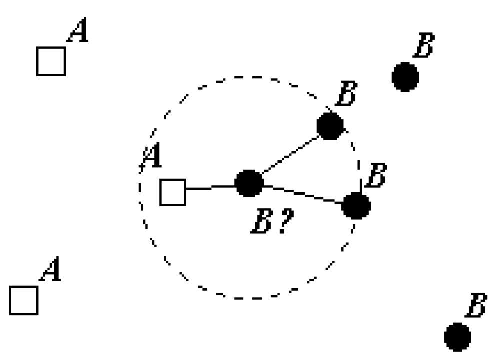


Figure 1 : KNN algorithm explained in a picture

*c) **Univariate-LSTM**:* LSTM is one of the most effective algorithms to predict time series data because of its capability to remember important information while forgetting unimportant information. For our Univariate LSTM model, we trained it only on the stock closing price data we obtained. For our analysis we took the time steps of 2 years (which means we are predicting the value of the future stock price using the past 2 years of the stock price as an input). We got the data, rearranged, and scaled the data to make it univariate input data. We performed univariate LSTM on amazon stocks to see the trends. Figure 3 [1]shows an example of a unit of univariate LSTM.

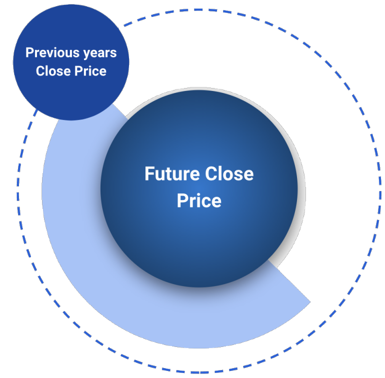
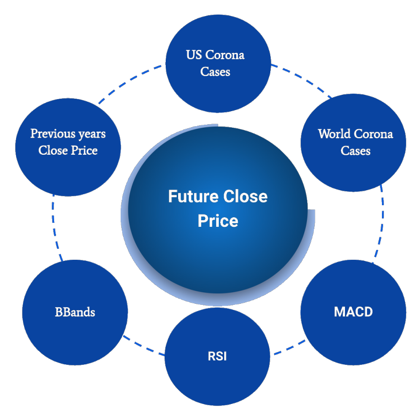


Figure 2 : Univariate LSTM

*B. **Proposed Method(Multivariate LSTM considering COVID-19 situation)***

As said above LSTM is one of the most effective algorithms to predict time series data. Considering the current COVID-19 global pandemic situation, we propose a method here to have a multivariate LSTM learning model to predict future behaviours of various stocks during such situations. Including its technical analysis. Technical indicators play an important role in knowing the volatility and momentum of stocks. They also provide additional information if stocks are overpriced or under-priced so the investors can make wiser decisions. There are various technical indicators available, but for this model the top ranked indicators, which help invest better, are used. The learning model takes five inputs, closing prices of stocks, confirmed COVID-19 cases, MACD, RSI, and Bbands. Below are a summarizing of the latter three inputs.



(***Figure 3 :*** Multivariate LSTM)

1) **MACD :** Moving Average Convergence Divergence, or also MACD, indicates the changes in the direction, strength, duration and momentum of stock price trends. It is calculated by subtracting the long term Exponential moving average(EMA) over 26 periods from short term EMA over 12 periods.

$$MACD = 12 - PeriodEMA \neg 26 - PeriodEMA \quad (3)$$

EMA, or exponentially weighted moving average, is the type of moving average. The EMA gives greater weight and significance to recent data.

2) **RSI** : The Relative Strength Index is a momentum indicator that charts the strengths and weaknesses of stocks based on stock closing prices for the most recent trading period. RSI computation follows the smoothing technique. The smoothing technique considers prior as well as current values. The formula is written as below.

   a) It is calculated using this simple formula:

   $$RSI = 100\neg[\frac{100}{1+RS}] \tag{4}$$

   where, RS=Average loss/Average gain

   The standard is to compute first initial RSI value based on 14 periods.

   b) To calculate the first average gain and the average lost for a 14 day period we used these formulas.

   $$1^{\text{st}}AvgGain = \frac{\Sigma Gainsoverthepast14periods}{14} \tag{5}$$

   $$1^{\text{st}}AvgLoss = \frac{\Sigma Lossesoverthepast14periods}{14} \tag{6}$$

   c) Next, the second or successive average loss and gain is calculated based on previous average and current loss and gain.

   $$AvgGain = \frac{(prevAvgGain * 13) + currentGain}{14} \tag{7}$$

   $$AvgLoss = \frac{(prevAvgLoss * 13) + currentLoss}{14} \tag{8}$$

3) **Bbands:** The Bollinger Band(BB) is the volatility indicator that plots two standard deviations, positive and negative, from a simple moving average of stocks. In simple words, it shows the band for the movements of stocks prices for a period of time. Calculating the Bbands follows the below steps:

   a) Compute the simple moving average (SMA) of the intended Attribute by using a 20 day SMA. This step averages out the first 20 days' closing prices and considers it as the first data point.

   b) Dropping the earliest close price and also adding the close price of day 21. Averaging it again and the result will become the second data point.

   c) Repeat step b for the rest of the remaining days.

   d) Calculate standard deviation (SD) of the price.

   e) Multiply the SD by two and then add and subtract this value from each data point from SMA. This will form the upper and lower bands respectively.

   $$BOLU = MA(TP, n) + m * \sigma[TP, n] \tag{9}$$

   $$BOLD = MA(TP, n)\neg m * \sigma[TP, n] \tag{10}$$

   Where BOLU is the Upper Bollinger Band, BOLD is the Lower Bollinger Band, MA is the Moving average, and TP is the typical price. Tp also equals High plus Low plus Close divided by 3, with n equaling the Number of days in a smoothing period and m equaling the Number of standard deviations. The last part σ[TP,n] equals the Standard Deviation divided by the last n periods of TP.

Algorithm 1 Training Procedure for MLSTM model

```
min_lr = 1e-4; epoch = 50; initial _lr = initial _lr
factor
for n <epoch do
    wait += 1
    if best_score >RMSE then
        best_score = RMSE
        save model
        if wait ≥ 10 then
            if initial_lr >min_lr then
                min_lr = initial_lr × factor
                new= max(new_lr, min_lr)
                wait = 0
            end if
        end if
    end if
end for
```

(***The training procedure algorithm[9] for Multivariate LSTM***)

This model is trained with stocks and COVID-19 data. The accuracy of the performance of this model gave much better accuracy than moving average, k-nearest, and the univariate LSTM learning models. Proving this model is the best fit for stocks data-set. The proposed model performs best based on Root Mean Square Error (RMSE) as an error measure. RMSE is calculated directly from predictions. Considering yi and y i respectively as the true close price and predicted price. For RMSE, the lower value is better. Shown in Eq 10 is the respective equation for RMSE.

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\Big(\frac{d_i - f_i}{\sigma_i}\Big)^2} \tag{11}$$

We trained our model to predict[9] for a period of time, and then the actual data of the current period is provided to the model so that we could use it for the prediction of subsequent periods. This is not only applicable to the way the model is used in practice but also opportune to the model using the best available data. The Multivariate LSTM Model[9] is trained and evaluated as follows.

Algorithm 2 Training Procedure for MLSTM model

Step 1. At the beginning of the test set, the last set of observations in the training set is used as input of the model to predict the next set of data.
Step 2: The model makes a prediction for the next period of time data set.
Step 3: Gets real-time observation and add it to history for predicting the next time.
Step 4: The prediction is stored and evaluated against the real observation.
Step 5: Go to step 1.

V. Data Visualization

A. *Data-Sets*

For the above-mentioned analysis, all the datasets were taken from a website named EOD data. The data which was downloaded from the website was the data on a daily basis in a text file for all companies. There were approximately 1300 text files for the last 5 years of data which we categorized based on the year. We further merged it to get six datasets for each year all years of 2015 through 2020.

Dataset List			
Datasets	Functions	Dimensions	Websites used
2015.csv	Companies' Stock Data for 2015	495504 x 11	EOD Data
2016.csv	Companies' Stock Data for 2016	1048575 x 11	EOD Data
2017.csv	Companies' Stock Data for 2017	1048575 x 11	EOD Data
2018.csv	Companies' Stock Data for 2018	1426829 x 11	EOD Data
2019.csv	Companies Stock Data for 2019	1517813 x 11	EOD Data
2020.csv	Companies' Stock Data for 2020	433965 x 11	EOD Data
Symbols.csv	Companies' Ticker Description	6652 x 3	EOD Data
COVID-19.csv	No. of cases in USA By Date	25609 x 5	Kaggle

All these datasets share the same column names of Name, Stock Exchange, Year, Month, Date, Open, High, Low, Close, Volume, and Diff. Each column stored the following information.

1) **Name:** This contained a Company's abreviated name.
2) **Stock_Excahnge:** This contains the stock exchange the company is listed in. (Either NASDAQ or NYSE)
3) **Year:** The year of the recorded stock information.
4) **Month:** The month of the stock information.
5) **Date:** The date of the stock information.
6) **Open:** Represents the opening price of the stock on the recorded date.
7) **High:** Represent the highest price of the stock on the recorded date.
8) **Low:** Represent the lowest price of the stock on the recorded date.
9) **Close:** Represent the closing price of the stock on the recorded date.
10) **Volume:** Number of transactions for the stock on the recorded date.
11) **Diff:** Represent the difference between the Open and the Close price of the stock of the recorded date.

We had a separate data set named Symbols that contained Stock Name abbreviations to identify which abbreviation is representing which company. The data set consists of three columns. Stock Exchange, Symbol, and Description. It contains the master data for each unique symbol with their corresponding company.

B. *Data Cleaning And Processing*

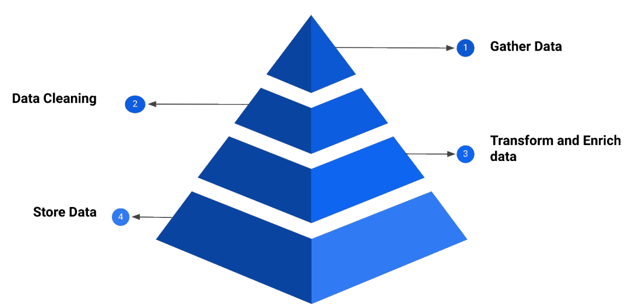


Figure 5 : Data set preparation steps

1) **Gather Data :** The method of measuring information and gathering on interest variables is called as data collection. It also has most important role in the big data cycle. For any particular topic, the best source which provides unlimited data is the internet. Most of the companies has the ability to get different sauces of external data and merge them with other transnational data.
2) **Data Cleaning :** As the data which we have from diverse data sources may have different characteristics , the first step right after gathering of data is to make the sources of data homogeneous and continue to develop our data product. This is called as the cleaning of data.
3) **Transform Data :** The process of changing the structure of data and converting it from one format into other is called as Data Transformation. This includes a wide range

of activities like converting the types, removing the null values or data which is duplicate, junk data like alphabets in age column and enrich the data.

4) **Store Data :** After transforming the data, data is output into a resultant file which can then be loaded into a data frame or database, and the data is now ready to be analyzed.

VI. Results And Analysis

The correlation matrix in Figure 6, shows us the correlation each data point has with each other within our data set.

It can be seen that the open and close data points have a high correlation with each other, as significant changes do not happen in the stock market in a short span of time.

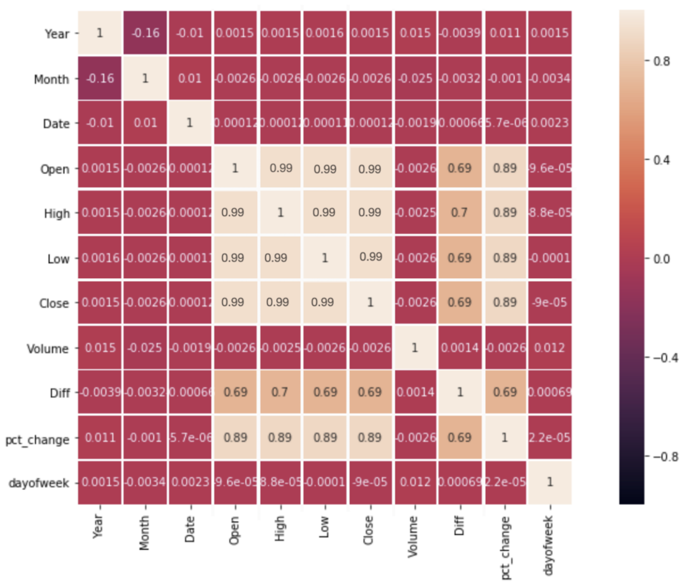


Figure 6 : Correlation Matrix

A. *Exploratory Analysis*

Initially the goal was to do an investigation on the stock data to discover unique patterns or to spot anomalies to test our hypothesis using different visualizations.

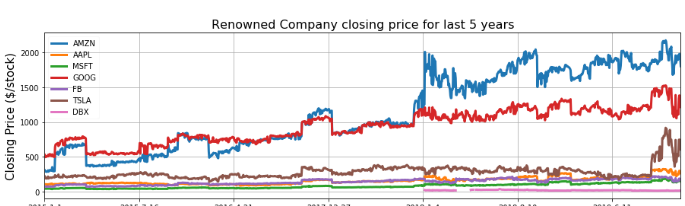


Figure 7 : Companies trend for Last five Years

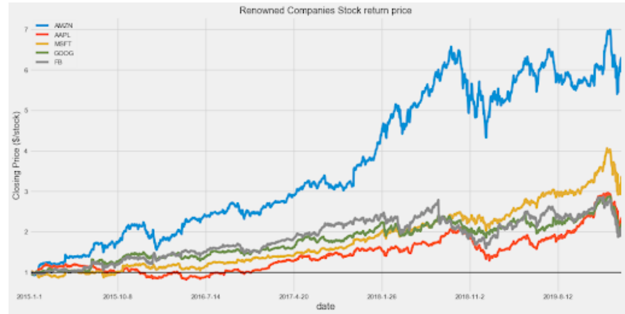


Figure 8 : Companies return for last one year

The stocks prices for the listed companies in Figure 7 and Figure 8 had increase in close price and returns of stocks respectively in the subsequent years since 2015. This helps to understand that the stock market has been having a steady rate of increase from 2015 to early 2020.

As COVID-19 was unforeseen, we decided to look more closely at the most recent stock information of few renowned companies for the last one year. It can be seen clearly in Figure 9, towards March 2020 (When COVID-19 cases were growing globally), stock prices started decreasing substantially.But the stock prices for companies like Apple and Amazon are doing good comparatively as the pandemic thwarts a variety of activities, people are heavily leaning on their tech gadgets, especially smartphones made by companies such as Apple.

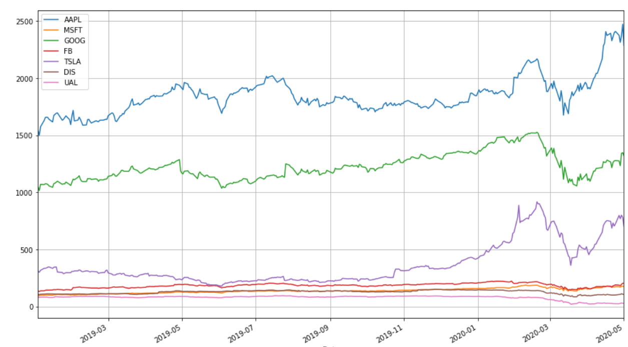


Figure 9 : Companies trend for 2019

We further investigate how COVID-19 is affecting stock market, we narrowed down our focus to individual companies.

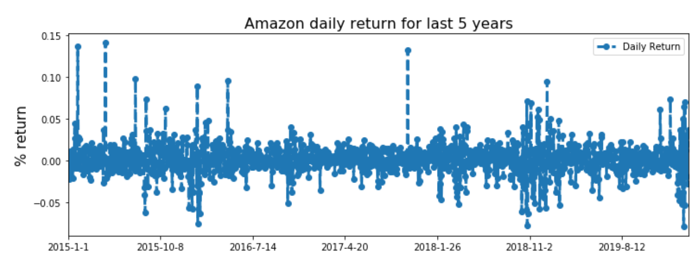


Figure 10 : Amazon % return for Last 5 Years

Figure 10 illustrates the return percentage of Amazon in the last 5 years. As seen, within 2016 and 2018, the return percentage was relatively the same. Amazon has garnered increased activity due to quarantine orders.

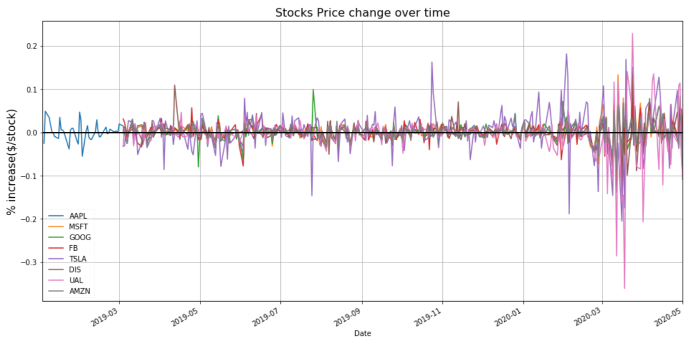


Figure 11 : % Change in companies stock price for 2019

The NASDAQ, where most of these new companies were listed, lost more than 80 percent of its value from it's peak due to the COVID-19 regulation orders.As the COVID-19 devastates travel demands, the airline industries stocks are going down as shown in the Figure 11(UAL Stock price decreased after March 2020).

Figure 12 : Amazon last one year closing price

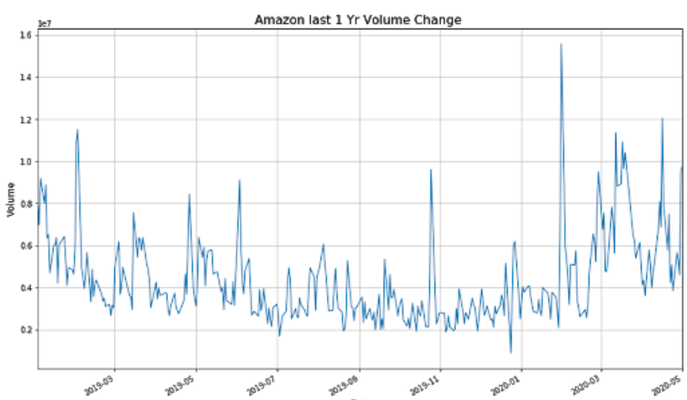


Figure 13 : Amazon last one year volume change

From Figure 12 and Figure 13, it's evident that the stock closing prices and volumes for Amazon drastically decreased once after March 2020 due to COVID-19 outbreak but it again increased due to the stay-at-home orders.

Figure 14 : Apple closing price

Figure 14 and Figure 15 represents how Apple experienced a slight decrease as COVID-19 was developing. But after March 2020, Apple's stock began to increase again.

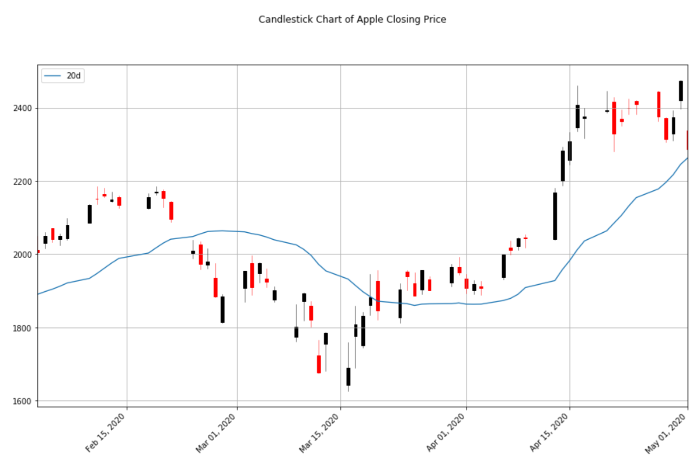


Figure 15 : Apple closing price candlestick

Taking COVID-19 as one of the variant, we plotted the closing price for different industries. Figure 17 shows that the stock prices are decreasing as the confirmed rate of COVID-19 is increasing. COVID-19, is in no doubt, affecting different markets in different ways.

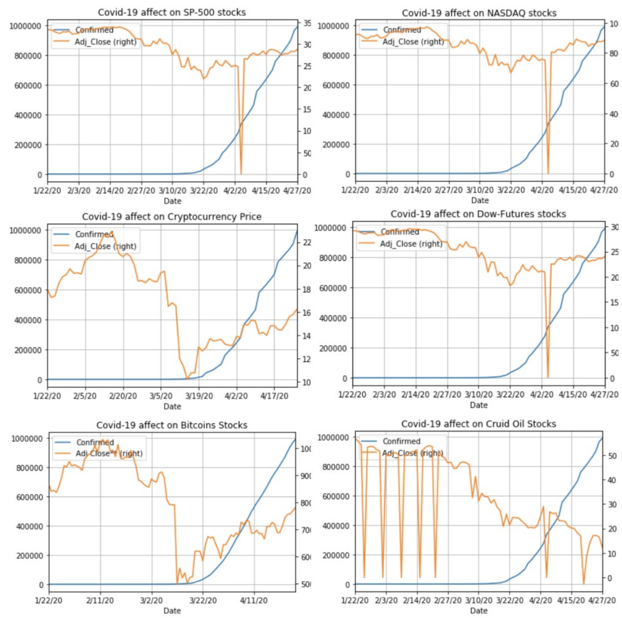


Figure 17 : COVID-19 effects on monetary markets

B. *Descriptive Analysis*

As a part of our descriptive analysis, we quantitatively described and summarized certain aspects of our data sets. Averaged out the Difference values(difference of opening and closing prices) from each year shown the Figure 18 to show whether the majority of stocks had any change at all during the year.The data showcases that from 2015 to 2017, there was a huge decrease in changes.The overall value of the stock market did not change drastically, even though few stocks may have high changes individually.

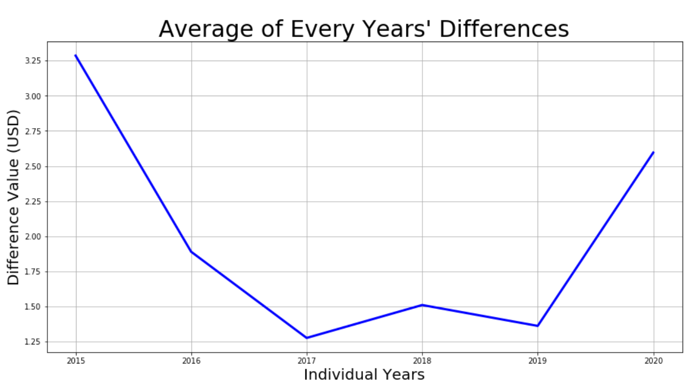


Figure 18: Line Graph of Difference Average

The Figure 19 shows the daily trading activity i.e volume of stocks for last five years to understand the reason for rise/fall of the profits.This showed that, the more the trading happened in year,lesser the difference for that year, also lesser the general profit for that year.

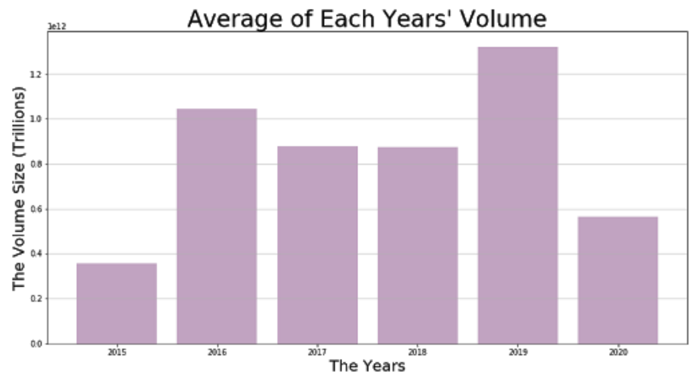


Figure 19 : Line Graph of Difference Average)

To make sure our both findings have correlation to each other, we looked at the percentages of each year where the stocks either closed positively/negatively, or no changes at all to their daily opening prices, which is being represented with six separate pie charts for each year in the Figure 20.

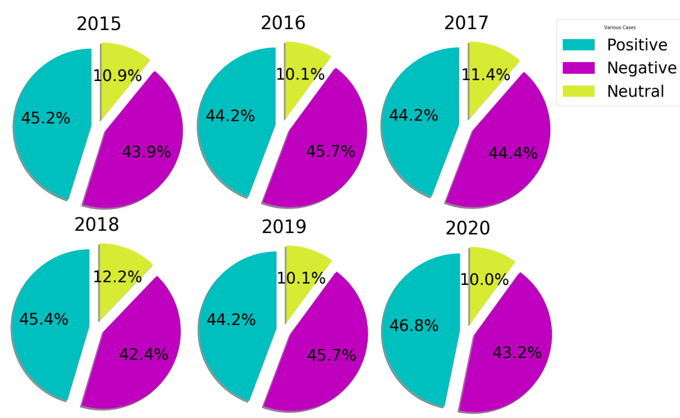


Figure 20 : Pie Charts of Opening/Closing Percentages

So, it can be concluded that the more transactions/volumes the stocks are receiving for each year, the more likely stocks in general will close on a negative value. As the Figure 18 shows the same years that had a steep fall in the line graph also had a high volume with a greater negative closing prices.

C. *Predictive Analysis*

Our raw dataset was transformed into a usable form that we can use for time series forecasting.

Figure 21 : Disney Closing Price over the years

We tried using the moving average to predict the stock price, the predicted closing price for each day will be the average of a set of previously observed values.

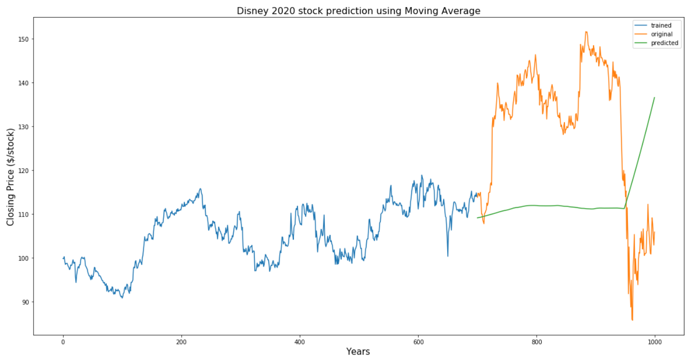


Figure 22 : Disney 2020 Stock Prediction using Moving Average

We prepared the data to fit an LSTM for a multivariate time series forecasting problem. 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 and moving average enough to identify whether the stock price will increase or decrease? Our answer is no, as stock prices are affected by social sentiments and uncontrollable natural disasters. The existing predictions model is done using only one variate, the Close Price. We also used Close price to predict the future close price of the company.

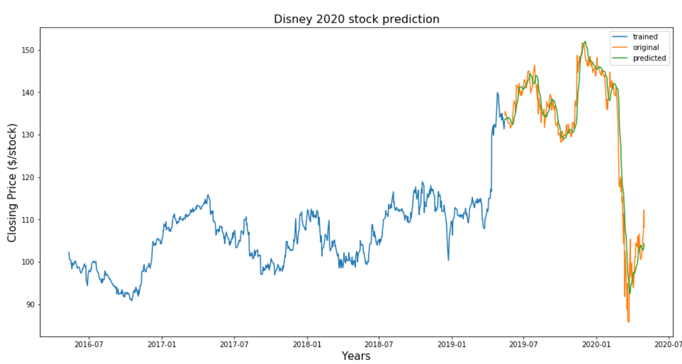


Figure 23 : Disney 2020 Stock Prediction using LSTM

But the reality is that there are many dimensions other than Time and historical close price that governs the stock market behavior. The proposed approach is the multivariate LSTM which takes into account BBands, MACD, RSI, historical close price, and the number of corona cases across the world and particularly in the United States.

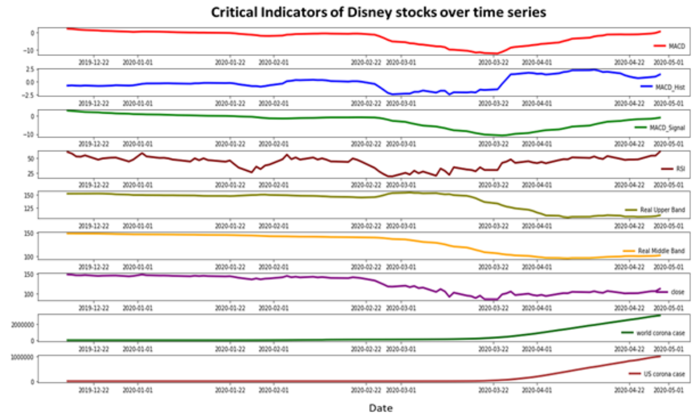


Figure 24 : Critical Indicators of Disney Stocks over the Time Series after 2019

Each of these is known to be indicators in stock market predictions except that we have the corona effect on the close price attached to them.

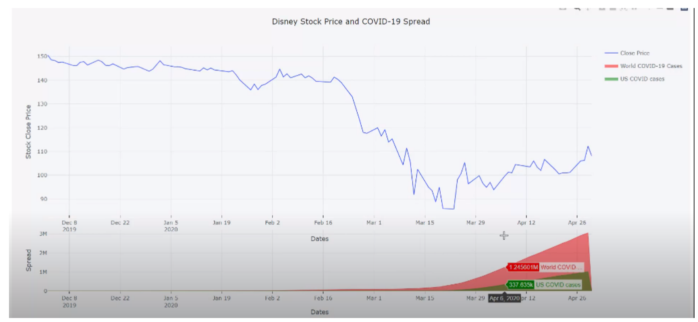


Figure 25 : Disney Stocks price and COVID spread

It is now safe to say that the MACD, BBands, RSI, close price, world corona cases, and the US corona cases in the past 60 days can be used to predict the close price of the following day. Using this as our hypothesis, we analyzed the data for Disney stocks for the past year. The below figure shows the trends for our variates.

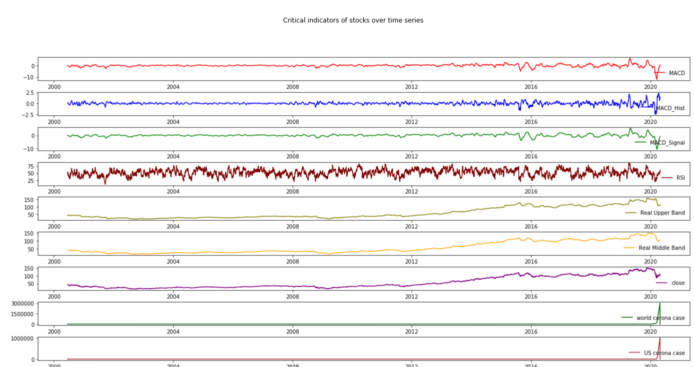
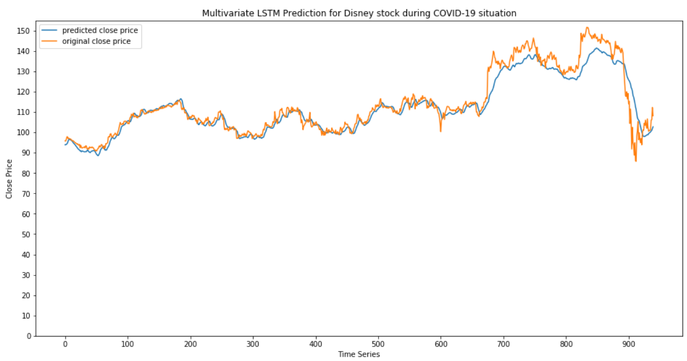


Figure 26 : Critical Indicators of Disney Stocks over the Time Series

As predicted, the accuracy of the multivariate model is much better than the univariate model.



(***Figure 27 :*** Multi variate LSTM Prediction for Disney Stocks during COVID-19 situation)

By using the same hyperparameters, the model achieved a validation loss of 6.2046e-04 after 50 epochs. For a Multistep forecast, we just need to rearrange the data to use the past 60 days of data to predict N steps into the future and change the units of the Dense layer to N. The mean squared error showed linear growth against the value of N. Using our proposed approach we were able to achieve the lowest root mean square error value so far 5.828. In comparison with other models like moving average which gave us Root Mean Square error of 220.55 and for the LSTM model the Root Mean Square Error value of stock prediction was 71.75. Our model has outperformed all other models visibly and we were successfully able to predict the future stock price during this COVID-19 situation.

D. Prescriptive Analysis

After knowing the results from descriptive and predictive analysis, prescriptive analysis gathers these results to apply decision making. Figure 28 shows the measures of the stock market to be considered to have the best outcome and take the best possible decisions of buying or selling stocks during uncertain situations like COVID-19.

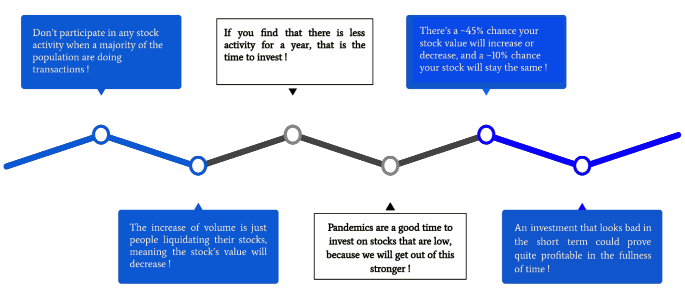


Figure 28 : Prescriptions Before Taking Actions

Since there are uncertainties in the data, we used stochastic optimization as a base prediction for prescriptive analysis. It says, if someone would like to have profits, they should not participate in any stock activity when a majority of the population are doing transactions. The volume increase resulting in lower profits means just that people are liquidating their stocks, meaning the stock's value will decrease. When there is less activity for a year, that is the time to invest because, previously stocks were low and it can only go up from there.

VII. Conclusion And Future Work

In this paper, we have examined the methods to predict stock prices using k-nearest, moving average and uni-variate LSTM with historical closing prices of stocks data. Moreover, to consider multiple factors to predict future of stocks during COVID-19 situation, this paper proposed the Multi-variate LSTM method with previous year closing price of stock, Bollinger Band, the relative strength index of stock, MACD, world corona cases and US corona confirmed cases as input variables. The experimental results showed that the proposed method with these input variables can successfully predict the future closing prices of stock market. These results after comparing with the results of different methods like KNN, Moving average, uni-variate LSTM, showed the stock market has long term existence on these input variables and this model is capable of fitting to any pandemic situations those can occur in future too. It is possible that there are other factors like the latest news/updates of a particular company that may affect public sentiments and affect stocks of that company. In future, we will take into account these problems along with already existing input variables to make our prediction stronger. This feature will make sure that the model is ready to predict future stocks based on pandemic situation as well as human sentiments.The main focus will be to improve accuracy of stock predictions for this change, using our multi-variate LSTM model.

References

[1] Sangyeon Kim, Myungjoo Kang, "Financial series prediction using Attention LSTM," arXiv:1902.10877 [cs.LG].
[2] Julius Tanuwijaya, Seng Hansun,"LQ45 Stock Index Prediction using k-Nearest Neighbors Regression,"International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3, September 2019
[3] Ida Vainionpää and Sophie Davidsson,"Stock market prediction using the K Nearest Neighbours algorithm and a comparison with the moving average formula,"CSC KTH 29 April 2014
[4] Thushan Ganegedara, "Discover Long Short-Term Memory (LSTM) networks in Python and how you can use them to make stock market predictions!," Data-Camp, January 1st, 2020
[5] Osman Hegazy, Omar S. Soliman and Mustafa Abdul Salam, "A Machine Learning Model for Stock Market Prediction."International Journal of Computer Science and Telecommunications [Volume 4, Issue 12, December 2013] 17
[6] Dev Shah; Haruna Isah and Farhana Zulkernine, "Stock Market Analysis: A Review and Taxonomy of Prediction Techniques," International Journal of Financial Studies, 7, (2), 1-22, 2019
[7] Thirunavukarasu Anbalagan, S. Uma Maheswari, "Classification and Prediction of Stock Market Index Based on Fuzzy Metagraph,"Procedia Computer Science, ISSN: 1877-0509, Vol: 47, Issue: C, Page: 214-221, 2015
[8] CORY MITCHELL,"How to Use a Moving Average to Buy Stocks",website: "https://www.investopedia.com/articles/active-trading/052014/how-use-moving-average-buy-stocks.asp"
[9] Wan, R., Mei, S., Wang, J., Liu, M., Yang F.," Multivariate Temporal Convolutional Network: A Deep Neural Networks Approach for Multivariate Time Series Forecasting." Electronics 2019, 8, 876.

[10] Jamilu A Adamu,“Advanced Stochastic Optimization Algorithm for Deep Learning Artificial Neural Networks in Banking and Finance Industries,”Risk and Financial Management, Vol. 1, No. 1, December 2019

[11] Alaa Sagheer, Mostafa Kotb, “Unsupervised Pre-training of a Deep LSTM-based Stacked Autoencoder for Multivariate Time Series Forecasting Problem,”Scientific Reports volume 9, Article number: 19038 (2019)

[12] Kokila,“Implementation of Extended Deep Neural Networks for Stock Market Prediction.”International Journal for Research in Applied Science and Engineering Technology. 8. 29-34. 10.22214/ijraset.2020.2006.

[13] Abhishek Dutta,Saikat Mondal, “Application of Deep Learning Techniques for Precise Market Prediction,”2020, DCAL National Conference on Machine Learning and Artificial Intelligence

[14] T. Arthi,RA. Chooriya Prabha,“Pattern recognition in stock market ,”IJCSMC, 2020

[15] Catalin Stoean ,Wiesław Paja,Ruxandra Stoean,Adrian Sandita,“Deep architectures for long-term stock price prediction with a heuristic-based strategy for trading simulations,”doi.org/10.1371/journal.pone.0223593,2019

[16] Xingyu Zhou ,Zhisong Pan,Guyu Hu ,Siqi Tang, and Cheng Zhao1,“Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets,”Volume 2018 —Article ID 4907423 ,15 Apr 2018

[17] Fischer, Thomas, Krauss, Christopher,“Deep learning with long short-term memory networks for financial market predictions,”FAU Discussion Papers in Economics, No. 11/2017

[18] Raghav Nandakumar, Uttamraj K R, Vishal R, Y V Lokeswari,“Stock Price Prediction Using Long Short Term Memory,”Volume: 05 Issue: 03 — Mar-2018

[19] Mansoor Momeni,Maryam Mohseni,Mansour Soofi,“CLUSTERING STOCK MARKET COMPANIES VIA K- MEANS ALGORITHM,”Kuwait Chapter of Arabian Journal of Business and Management Review Vol. 4, No.5; January. 2015

[20] Ro Khanna,“Most Americans don’t own stocks.” website: https://www.politifact.com/factchecks/2018/sep/18/ro-khanna/what-percentage-americans-own-stocks/

[21] Hiransha M., Gopalakrishnan E.A., Vijay KrishnaMenon, Soman K.P.,“NSE Stock Market Prediction Using Deep-Learning Models,”Volume 132, 2018, Pages 1351-1362,https://doi.org/10.1016/j.procs.2018.05.050

[22] Jacinta Chan Phooi M’ng ,Rozaimah Zainudin,“Assessing the Efficacy of Adjustable Moving Averages Using ASEAN-5 Currencies,”https://doi.org/10.1371/journal.pone.0160931

[23] Julius Tanuwijaya,Seng Hansun,“LQ45 Stock Index Prediction using k-Nearest Neighbors Regression,”September 2019 DOI: 10.35940/ijrte.C4663.098319

[24] S K Mitra,“Usefulness of Moving Average Based Trading Rules in India ,”Accepted: February 11, 2011 doi:10.5539/ijbm.v6n7p199

[25] Author links open overlay panelZoicaş-IenciuAdrian,“The Sensitivity of Moving Average Trading Rules Performance with Respect to Methodological Assumptions,”Volume 32, 2015, Pages 1353-1361,https://doi.org/10.1016/S2212-5671(15)01512-9

[26] A. Moosavian ,H. Ahmadi,A. Tabatabaeefar and M. Khazaee1,“Comparison of Two Classifiers; K-Nearest Neighbor and Artificial Neural Network, for Fault Diagnosis on a Main Engine Journal-Bearing,”Volume 20 ,Article ID 360236,10 pages,https://doi.org/10.3233/SAV-2012-00742

[27] Jenna Burrell,“How the machine ‘thinks’: Understanding opacity in machine learning algorithms,”First Published January 6, 2016 Research Article,https://doi.org/10.1177/2053951715622512

[28] Vargas, Rocio, Mosavi, Amir, Ruiz, Ramon,“Deep learning: A review.[Working Paper].”https://eprints.qut.edu.au/127354/

[29] Zachary C. Lipton Jacob Steinhardt,“Troubling Trends in Machine Learning Scholarship,”Presented at ICML 2018, arXiv:1807.03341v2

[30] Niels Joachim Gormsen,Ralph S. J. Koijen,“Coronavirus: Impact on Stock Prices and Growth Expectations,”(May 13, 2020). University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2020-22, http://dx.doi.org/10.2139/ssrn.3555917

[31] Adil Moghara, Mhamed Hamicheb,“Stock Market Prediction Using LSTM Recurrent Neural Network,”https://doi.org/10.1016/j.procs.2020.03.049

[32] L. Chen, Z. Qiao, M. Wang, C. Wang, R. Du and H. E. Stanley, ”Which Artificial Intelligence Algorithm Better Predicts the Chinese Stock Market?,” in IEEE Access, vol. 6, pp. 48625-48633, 2018, doi: 10.1109/ACCESS.2018.2859809.