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. 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.

II. Literature Review

This section consists of related works that contain information in regards to stock market predictions using either econometric models or generative adversarial networks. [10] Econometric models, in comparison to generative adversarial networks, have been proven to be the least accurate of the two. The following works emphasize how machine learning and deep learning have impacted the stock market and financial predictions throughout the world. Activation algorithms [12] provide predictions based on a non-linear progression. If applied to an artificial neural network, Universal approximations can be achieved. If it is not applied, the neural network can only determine a linear relationship between variables. The neural network is one of the foundations of AI. Machine Learning is a sub-field of AI. With machine learning, predictions can be made on rule-less programming. Predicting and forming patterns to inform or assist users. Providing valuable knowledge for finance industries to utilize.

Stock Market predictions using deep learning can provide serviceable data but lacks the information about future losses. [12] Even though deep learning can provide accurate data, it does, however, won't be able to provide all possible cases. For example, cases that are affected by personal sentiments, public opinion, natural disasters, etc. But even so, deep learning can still be expanded on to include these variables for a more accurate model. It merely just needs to have enough data on these variables to learn and recognize the patterns.

One of the most primitive approaches to stock market predictions was utilizing simple regression algorithms.[13] But with the introduction of the machine and deep learning, stock market predictions have become more accurate over the years. One of the best performing machine learning models is LSTM or Long Short-Term Memory. The only downfall from LSTM is the accuracy can only improve with bigger data sets. So a lack of information is a detriment to the training of LSTM. And may even provide LSTM to be unusable.

The pattern recognition was utilized on a single stock, Apple. With past data points, a machine learning algorithm is trained to predict the future.[14] All data points are normalized into one type of data. The algorithm works by predicting multiple patterns and testing these patterns against each other to determine the differences. Once the differences are determined, this further information is used by the machine learning algorithm to create more patterns. The utilized method was able to achieve an 80% accuracy of future stock predictions.

Deep learning provides the means to predict the dynamic movement of stock trading and prices. [15] The methods utilized come from a long short-memory network and temporal convolutions neural model. LSTM is the more obvious choice to go for learning a series of stock prices. But CNN does have the advantage of learning patterns and models more quickly due to its hierarchical creation of multiple layers to represent relationships between data. LSTM is able to learn long term data. While CNN benefits from sequential access to data over time.

The prediction of the stock market is highly sought after due to the high benefits someone can gain from the data. [16] Predictions, utilizing Generative Adversarial Nets looks at two frontiers. The first being forecast error. The second being future prediction accuracy. One of the basic frameworks of Generative Adversarial Nets is Long Short-Term Memory. Unlike the Moving Average 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. LSTM can further be improved to become a GAN-FD algorithm that also accurately predicts losses.

The data sets utilized are from the New-York Stock exchange and National Stock Exchange.[22] Four different deep learning algorithms were applied to the data sets; Multilayer Perception, Recurrent Neural Network, Long Short-Term Memory, and Convolutions Neural Network. The models utilized were able to identify a pattern between the different stock markets. Meaning there is an underlying similarity between both stock markets. Deep learning models are also shown to be outperforming ARIMA models or also known as Econometric models.

The Stock Market is growing and the access to it is easily made now with the utilization of the internet. [6] There are various techniques to predict outcomes. The best of the techniques being machine learning. There are two forms of machine learning, supervised learning and unsupervised learning. Supervised learning consists of Support Vector Machines

and Decision Trees. While unsupervised learning includes reverse K-means and LSTM. Even though all these algorithms may be accurate, they need to include sentiment analysis. One of the easier availability of sentiment data is social media outlets.

In recent research, it has been determined that not the majority of Americans own stocks. [21] Only a handful of modest-income Americans have a substantial stake in stocks and the majority of stocks are held by high-income families. Lack of college education, marital status, and ethnicity also plays a role in why the majority of Americans do not have stock stakes as well. Long Short Term Memory is considered a recurrent neural network. This neural network is pitted against an artificial neural network in terms of accuracy. [18] The weights for the data set are adjusted using stochastic gradient descent. LSTM is a Recurrent Neural Network that can capture dependencies for either a long or short term. LSTM maintains its own cell state while being able to store other memories. This allows the cells of LSTM to input it's old state to output a new cell state. In an ANN, the cells only store activation information. Thus LSTM performs better because it can store information for a long time while still making predictions.

LSTM is compared against a standard deep net, a random forest, and a simple logistic regression. [17] LSTM proved to be more accurate compared to the standard deep net and the simple logistic regression. But it did fall short from the accuracy of a random forest. The benchmark tests were performed on a large-scale financial market on the SP 500. LSTM did provide better predictions in comparison to random forest, but only due to the variable of an unforeseen financial crisis. Clustering is one of the main concepts to recognize patterns. Clustering is an unsupervised classification.[20] The stock market under speculation is the Tehran Stock Exchange. K-means is one of the more popular clustering methods. But before being able to utilize it, there need to be validity indexes. The results of clustering help industries make better decisions.

III. 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 highlighting 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, like 10 days, 20 days, 50 days to identify trend direction, and to get a basic idea of which way the price is moving. We observed the direction of the moving average 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.

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 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 SMA, the first M1 is not valid, as 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 have done this by choosing the K closest instances 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.

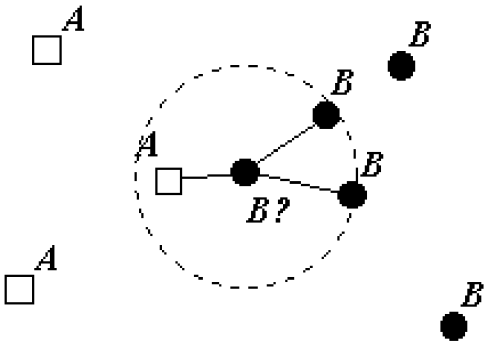
We divided the training data into vectors and then we calculated the distance from the test data to its neighbor using euclidean distance, which basically is the closest distance between the two components. Euclidean distance is the square root of the sum of the squared differences between a new point (x) and an existing point (xi) across all input attributes(j).

$$ED(x, xi) = sqrt(sum((xj–xij)^2)) \tag{1}$$

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

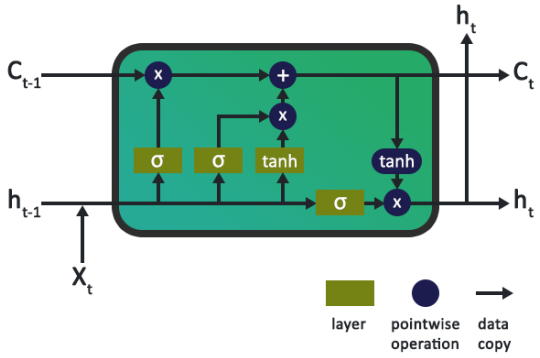
(Figure 1 : MA calculation for Close price)

The prediction we did 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 2 : KNN algorithm explained in a picture)

*c) **Univariate-LSTM**:* LSTM(Long Short Term Memory) is one of the most effective algorithms to predict time series data because of its capability to 'remember' important information while 'forgetting' unimportant information. The Univariate LSTM model is trained only on the stock close price data we obtained; that is why it is an univariate model. For our analysis we took the timesteps to be 2 Yrs(which means we are predicting the value of the future stock price using the past 2 Yrs of the stock price as an input). We got the data, rearranged, and scaled our 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 3 : Example of an unit of 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 important role in knowing the volatility, momentum of stocks and provide additional information if stock is overpriced or under-priced so the investors would take decisions wisely. There are various technical indicators available and for this model, the top ranked indicators which help invest better are used. The learning model takes five inputs, "close" prices of stocks, "confirmed" Covid-19 cases, MACD, RSI, Bbands where,

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

$$MACD = 12 - PeriodEMA \neg 26 - PeriodEMA \quad (2)$$

EMA is the type of moving average, which is also referred as exponentially weighted moving average which gives significance and greater weight to recent data.

2) **RSI** : The Relative Strength Index(RSI) is an a momentum indicator that charts the strength/weakness of stock based on stock closing prices for recent trading period. Computation follows the smoothing technique that is it considers prior as well as current values :

a) It is calculated using this simple formula:

$$RSI = 100 \neg [\frac{100}{1 + RS}] \quad (3)$$

where, RS=Average loss/Average gain

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

b) The very first average gain and average loss for 14 day period is calculated as:

$$1^{st} AvgGain = \frac{\Sigma Gains over the past 14 periods}{14} \quad (4)$$

$$1^{st} AvgLoss = \frac{\Sigma Losses over the past 14 periods}{14} \quad (5)$$

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

$$AvgGain = \frac{(prevAvgGain * 13) + currentGain}{14} \quad (6)$$

$$AvgLoss = \frac{(prevAvgLoss * 13) + currentLoss}{14} \quad (7)$$

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

a) Compute simple moving average(SMA) of the intended attribute,usually using 20 day SMA. This step averages out the first 20 days' closing prices, considering it as the first data point.

b) Drop the earliest close price and add the close price on day 21.Average it.This would be second data point.

c) Repeat step 2 for further days.

d) calculate standard deviation(SD) of the price. SD means how numbers are spread out from its average value.

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

$$BOLU = MA(TP, n) + m * \sigma[TP, n] \quad (8)$$

$$BOLD = MA(TP, n) \neg m * \sigma[TP, n] \quad (9)$$

Where,BOLU = Upper Bollinger Band, BOLD = Lower Bollinger Band, MA = Moving average, TP (typical price) = (High+Low+Close)÷3, n = Number of days in smoothing period (typically 20), m = Number of standard deviations(typically 2), σ[TP,n]=Standard Deviation over last n periods of TP.

Training procedure algorithm[9] for Multivariate LSTM is shown below:

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
```

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, uni-variate 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 measures. 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 equations for RMSE;

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\Big(\frac{d_i - f_i}{\sigma_i}\Big)^2} \quad (10)$$

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.

IV. Data Visualization

A. *Setup*

Everything what we implemented using code was done with Python. There are a lot of reasons to choose python like the big number of libraries and APIs which are capable to manipulate and visualize data. Once the data is displayed in the forms of charts, it becomes easy to analyze the trends and perform all kinds of analysis like exploratory, predictive and prescriptive. There are some environments which we used for the development:

1) **Jupyter Notebook :** All of these documents are produced by an app named Jupyter Notebook. This app contains computer programming code (like Python) and also text elements (like paragraphs, hyperlinks figures etc.) as well. All the documents produced by notebook are human-readable which contains the complete description of analysis and results in the forms of charts, tables etc. The installation of the Jupyter notebook is through conda by running a command conda: pip3 intall Jupyter

2) **Anaconda:** It is a distribution for programming languages like python and R which is free and open-source. It is mainly used for computations which are scientific (Data Scince, data processing, predictive analysis and machine learning). The aim is to simplify the management of packages and deployment. The command which we used to install the updated version of Anaconda manager is bash Anaconda-latest-Linux-x86_64.sh There are a lot of packages which are used to perform the analysis of the above mentioned data. Some of the packages are:

- **Scikit-learn** also known as sklearn is a software library which is used for machine learning in python. The main features of this library is to implement various machine leanring algorithms like support vector machines, random forests, gradient boosting k-means etc. This library is mainly used with other python numerical and scientific libraries like NumPy and SciPy.
- **Pandas** is another software library which is written with Python for manipulation and analysis of data. Particularly, it offers how to use different data structures

to store the data from csv files and also how to perform operations with the data stored. This library has a lot of features. Some of them which are used for the above analysis are:

  - Dataframe for manipulation of data with inbuilt indexing.
  - Data alignment and cleaning of the data (remove values which are missing or incorrect)
  - Data structure which is used to create additional columns.
  - Group by which was used to perform split-apply-combine operations.

- **NumPy** is also a library for Python which adds support for large and high-dimensional matrices and arrays. It also has a large group of high-level mathematical functions which can be used to perform multiple operations on these arrays.
- **Matplotlib** is a python library mainly used for plotting graphs and charts. and its numerical mathematics extension NumPy. Pyplot is a module of Matplotlib which is used to provide an interface that resembles MATLAB. The design of Matplotlib is in such a way that it can be used as MATLAB, having an advantage of open-source and free and consisting of all the abilities to use Python, and the advantage of being free and open-source.

B. 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 daily basis in a text file for all companies. There were approximately 1300 text files for last 5 years of data which we categorized based on the year and merged it to get six datasets for each year 2015 - 2020. All these datasets have same columns namely Name, Stock_Exchange, Year, Month, Date, Open, High, Low, Close, Volume, Diff. All the columns are used to store the following information about the stock of a company:

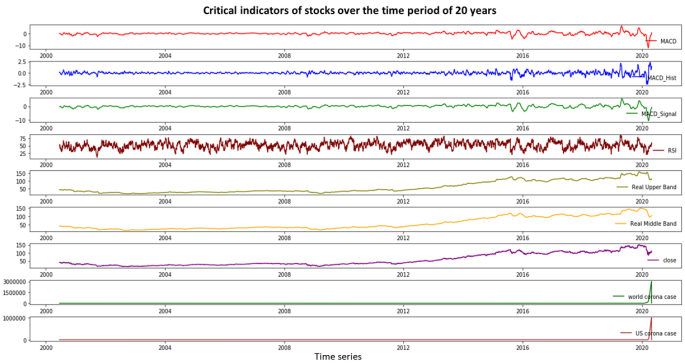
1) **Name:** This contains the abbreviations of the companies which are listed in the stock market
2) **Stock_Excahnge:** This contains the name of the stock exhange the company is listed in. This could be NASDAQ or NYSE.
3) **Year:** Represent the year of the stock.
4) **Month:** Represent the month of the stock.
5) **Date:** Represent the date of the stock.
6) **Open:** Represent the opening price of the stock on a particular day.
7) **High:** Represent the highest price of the stock on a particular day.
8) **Low:** Represent the lowest price of the stock on a particular day.
9) **Close:** Represent the closing price of the stock on a particular day.
10) **Volume:** Number of shares traded(Volume column is used to register the number of shares that got traded during a single day)
11) **Diff:** Represent the difference between the highest and the lowest price of the stock on a particular day.

Because the name column contains the abbreviations of the companies therefore, to identify which abbreviation is representing which company we have a 7th data sets named Symbol which consists of three columns namely Stock_Exchange, Symbol, Description. It contains the master data for each unique symbol with their corresponding company.

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

V. Preliminary Results

Our raw data set was transformed into something that we can use for time series forecasting.We prepared the data and fit an LSTM for a multivariate time series forecasting problem. We made a forecast and re-scaled the results back into the original units.We created a quick plot of each series that we have in Figure 4 as below:



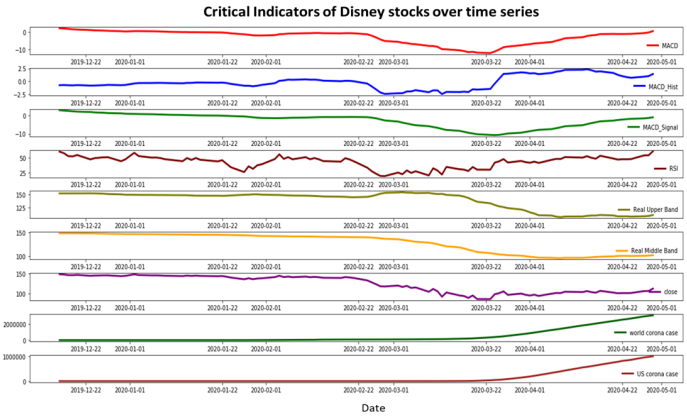
(Figure 4 : Critical Indicators of Stocks over the time period of 20 yrs)

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? Our answer is no, as stock price are affected by the news about the company and other factors like demonetization or merger/de-merger of the companies. There are certain intangible factors as well which can often be impossible to predict like the Covid situation at hand. Existing predictions models are done using only one dimension that is the Close Price. we used Close price to predict the future close price of the company. But the reality is that there are many dimensions other than Time, historical close price which govern the stock market behaviour. 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. Each of these are known to be indicators in stock market predictions except that we have the corona effect on the close price attached to them. BBands are not a standalone trading system,but simply an indicator designed to provide traders with information regarding price volatility. Many traders believe the closer the prices move to the upper band, the more overbought(overvalued owing to excessive buying at unjustifiably high prices.) the market, and the closer the prices move to the lower band, the more oversold (sold to a price below its true value.) the market. BBands are to discover opportunities that give investors a higher probability of success. The MACD has a positive value whenever the 12-period EMA (blue) is above the 26-period EMA and a negative value when the 12-period EMA is below the 26-period EMA. The more distant the MACD is above or below its baseline indicates that the distance between the two EMAs is growing.MACD measures the relationship between two EMAs, while the RSI measures price change in relation to recent price highs and lows. These two indicators are often used together to provide analysts a more complete technical picture of a market.

The relative strength indicator (RSI) aims to signal whether a market is considered to be overbought or oversold in relation to recent price levels. The RSI is an oscillator that calculates average price gains and losses over a given period of time; An asset is usually considered overbought when the RSI is above 70% and oversold when it is below 30The RSI was designed to indicate whether a security is overbought or oversold in relation to recent price levels. The RSI is calculated using average price gains and losses over a given period of time. RSI Like most technical indicators, its signals are most reliable when they conform to the long-term trend.

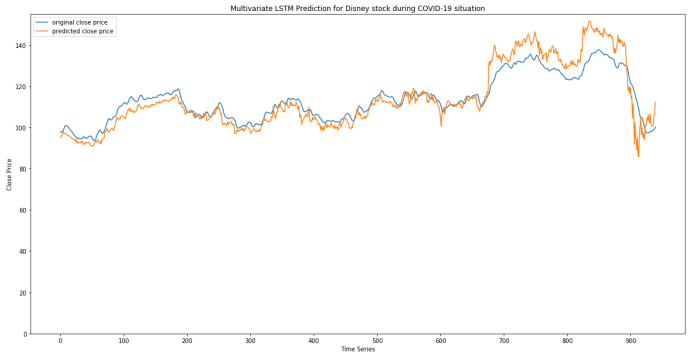
Investors have been contending with the potential for the Covid-19 outbreak to disrupt global trade and slow growth, and stock prices around the world have reflected concerns about the virus's spread outside China. Concerned investors are dumping stocks, seeking safer investments like government bonds, pushing prices up and yields down.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 analysed the data for Disney stocks for the past year. Below figure shows the trends for our varieties.



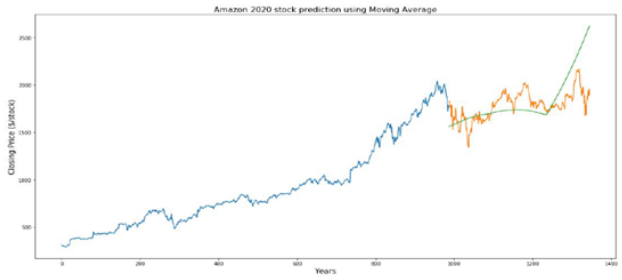
(Figure 5 : Critical Indicators of Disney Stocks over the Time Series)

Predictably, the accuracy of the multivariate model is much better than the univariate model. By using the same hyper parameters, the model achieved a validation loss of 6.2046e-04 after 50 epochs. For Multi step forecast, we just need to rearrange the data to use 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.

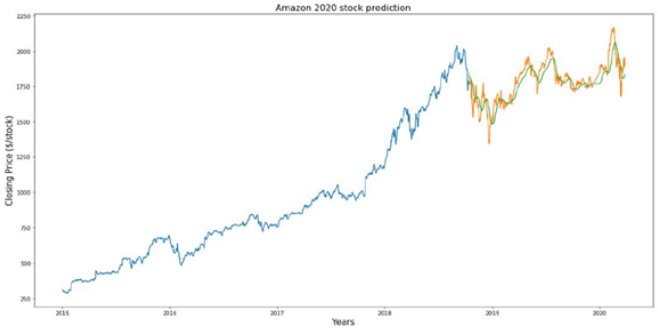


(Figure 6 : Multi variate LSTM Prediction for Disney Stocks during Covid-19 situation)

Using our proposed approach we were able to achieve the lowest root mean square error value so far 5.828.



(Figure 7 : Amazon 2020 Stock Prediction using Moving Average)



(Figure 8 : Amazon 2020 Stock Prediction using LSTM)

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.

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