Assessing The Impact Of COVID-19 on Stock Price Of A Few Large Companies

### Project Members

Donna Benitez  
Krishna Piratla  
Valentina Gonzalez

# Abstract

Stock market prediction has always been a lucrative research area given its potential to reap high rewards. Stock value prediction models are highly valued by data-driven stock traders since such models help make an informed decision as opposed to resorting to speculation. Typically, ARIMA models are known for making short-term predictions. Similarly, neural networks can also be used for the same purpose. In this project, we tried to assess the type of impact the COVID pandemic had on a company’s stock data: First we zeroed in on the best prediction model among ARIMA, KNN, and Neural Network – which turned out to be a single-layer neural network; then we compared the forecasted values with the actual values to get an idea of the kind of impact COVID had on the stock at hand. We also validated our results by building an LSTM model and comparing its predictions with the actual stock data. When the training data is large enough, a single layer neural network seems to provide a good idea of the qualitative impact of COVID on stock prices. The LSTM model gave a good qualitative idea in every case.

# Introduction

The Covid pandemic has shook the world and changed the landscape of businesses across the globe. Some companies saw an exponential increase in their growth (E.g.: Zoom) while others saw their business impacted negatively. In this project we tried to come up with an idea to understand the qualitative impact of Covid on various companies/industries.

# Problem Statement

It would be interesting to predict how the companies’ stocks would have performed if the covid pandemic hadn’t hit the world. We will compare these predictions to the actual performance of the stock post-covid.

# Dataset Description

The stock price related data has been collected using the *quantmod* package in R.

Since the major market crash due to the pandemic occurred towards the end of February 2020, we segregated the timeline as follows for the purpose of our analysis:

Pre-Covid Data Range: 1 Jan 2015 to 28 Feb 2020

Post-Covid Data Range: 1 Mar 2020 to 13 Apr 2020

NOTE: The pre-covid data for ZM (Zoom) was available only from 18 Apr 2019.

### Predictor and Response Variables

We chose the Closing Price of a stock for each day in the time range for the purpose of our analysis. The closing price acted as both the predictor and the response variable.

Wherever necessary for the model, the closing price has been lagged appropriately in order to act as a predictor and generate the required response predictions.

# Models used:

We started with the ARIMA model since it is like a benchmark for timeseries related data. Then we used the KNN Model and a single-layered Neural Network to generate the predictions. Finally, we also used an LSTM model’s predictions and compared them visually with the actual stock data.

# Methodology

We built ARIMA, KNN, and Neural Network models for the stock data until 28/Feb/2020 – the date around which the COVID situation has crashed the stock market for the first time. We call these models as ‘before covid models’. We compared the accuracy of these models through parameters such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Once we finalized the best model out of the three, we considered its predictions for the next 30 market days, i.e., from 02/Mar/2020 until 13/Apr/2020. We compared these predictions with the actual stock data to gauge whether the company has performed better or worse than expectation post-covid. We also used predictions from the LSTM model to validate these predictions.

We repeated this analysis for the stock data of three companies – Microsoft, Ferrari, and Zoom.

## ARIMA Model

We used the auto.arima() function in R to determine the best arima fit for the given stock data. Following are the ARIMA models determined as best fit for each company’s stock data.

**MSFT (*Microsoft*)**: ARIMA (3, 1, 3)

**RACE (*Ferrari*)**: ARIMA (1, 1, 1)

**ZM (*Zoom*)**: ARIMA (2, 1, 1)

## KNN Model

We chose the K value by taking the square root of the number of observations in the training set. The k value is 36, 33, 14 respectively for MSFT, RACE, and ZM. The h value was 30 for each of these models to generate the next 30 days predictions.

## Single Layer Neural Network:

The number of hidden nodes in the network were calculated using the relation:

Ns = Number of training samples; Ni = Number of input neurons; No = Number of output neurons;

a = 1.5-10

The lambda value has been estimated using the BoxCox function

## LSTM Model:

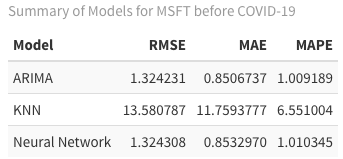
Data was input in 60 timesteps. The output space dimensionality was set to 50.

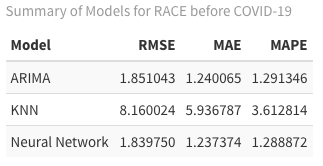
A 4-layered LSTM model was built with 4 dropout layers of dropout rate 0.2.

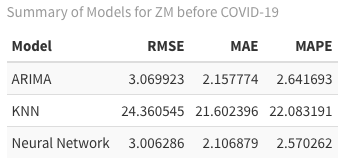
Adam optimizer was used and the loss was set as mean squared error. The model was fit on 100 epochs with a batch size of 32.

# Results:

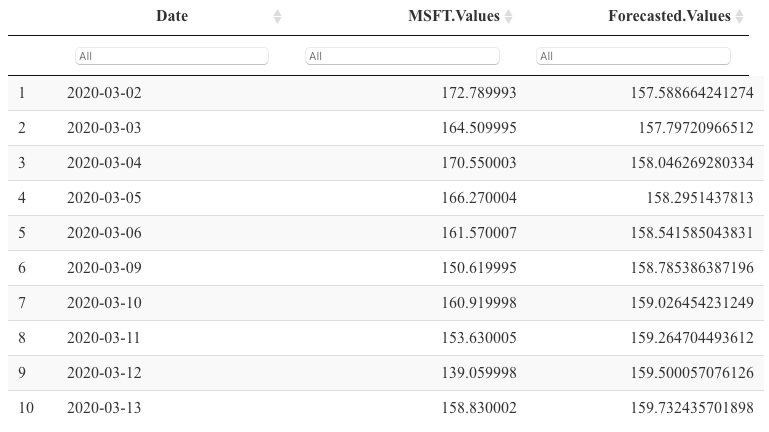
In each of the three cases, Neural Network model performed the best (along with ARIMA).

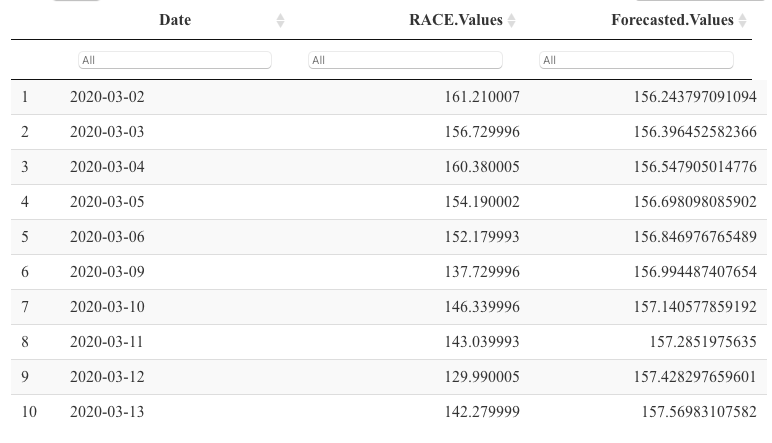


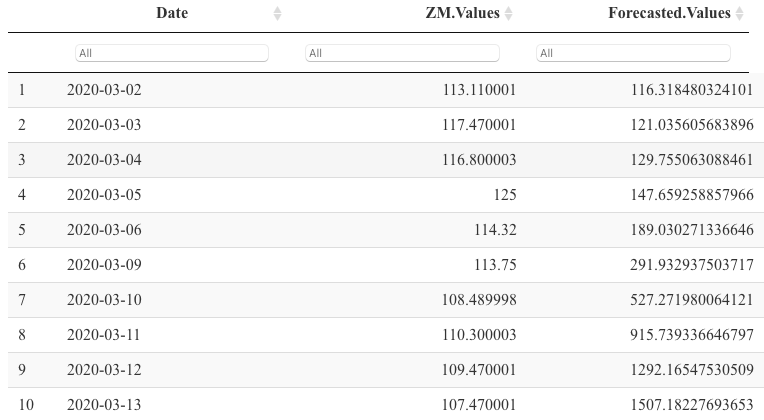




So, we chose Neural network model to compare the predictions with actual results.

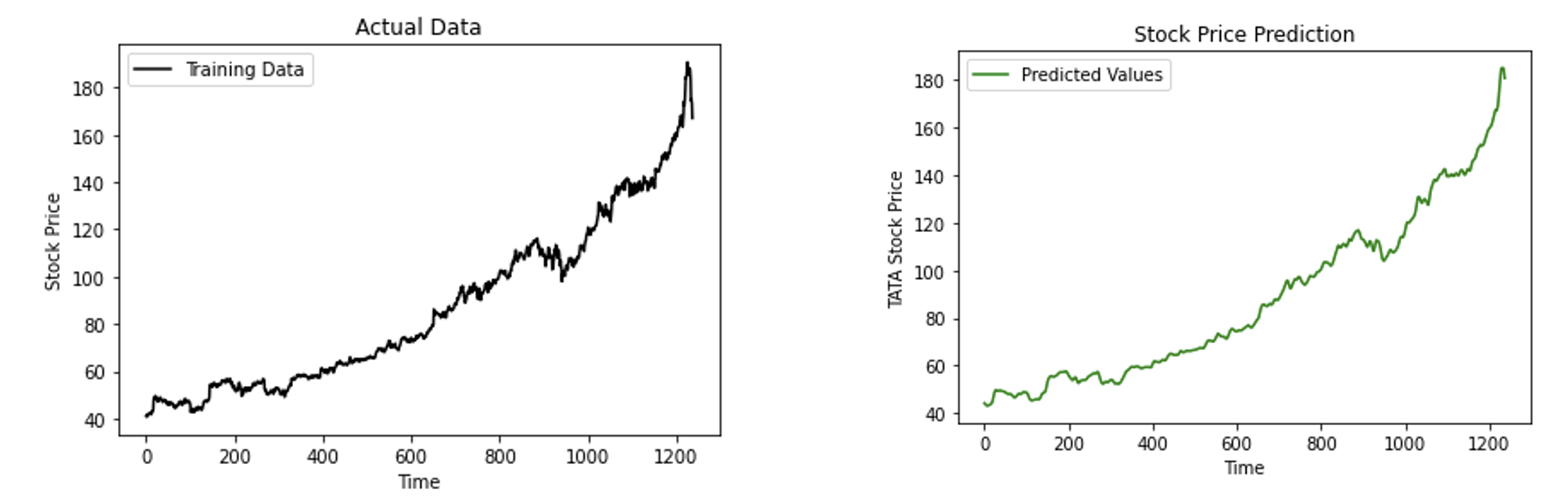




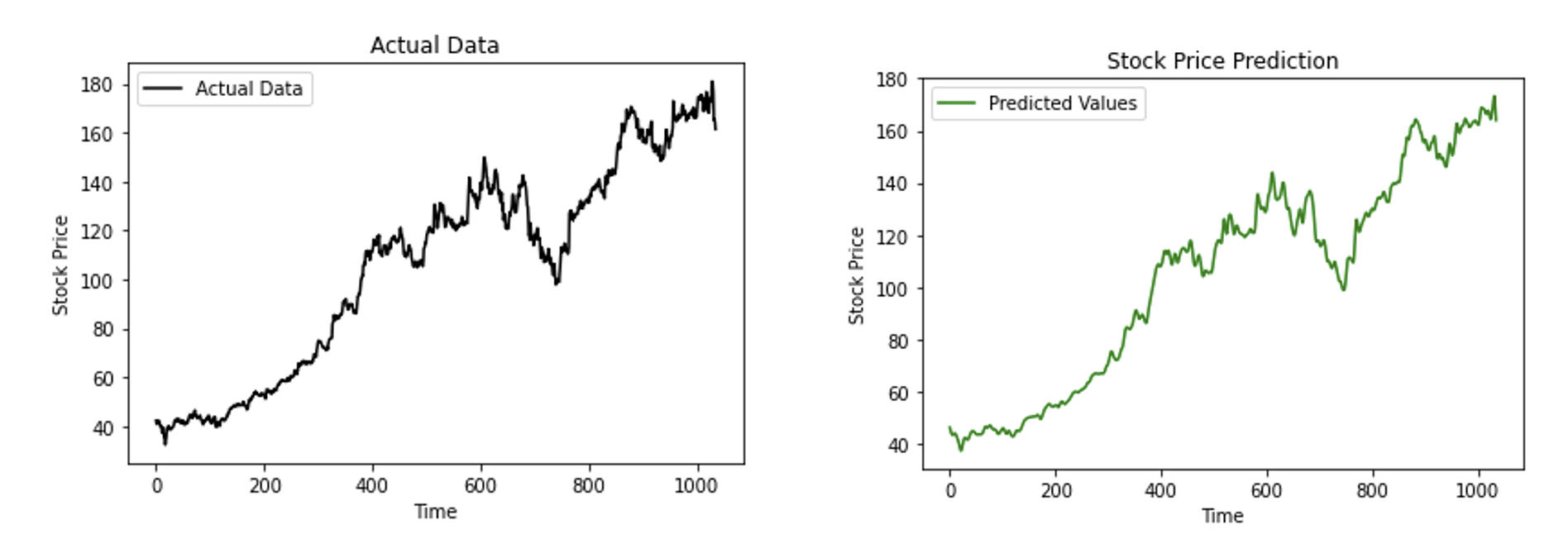


We then used the LSTM model predictions to validate our observations.

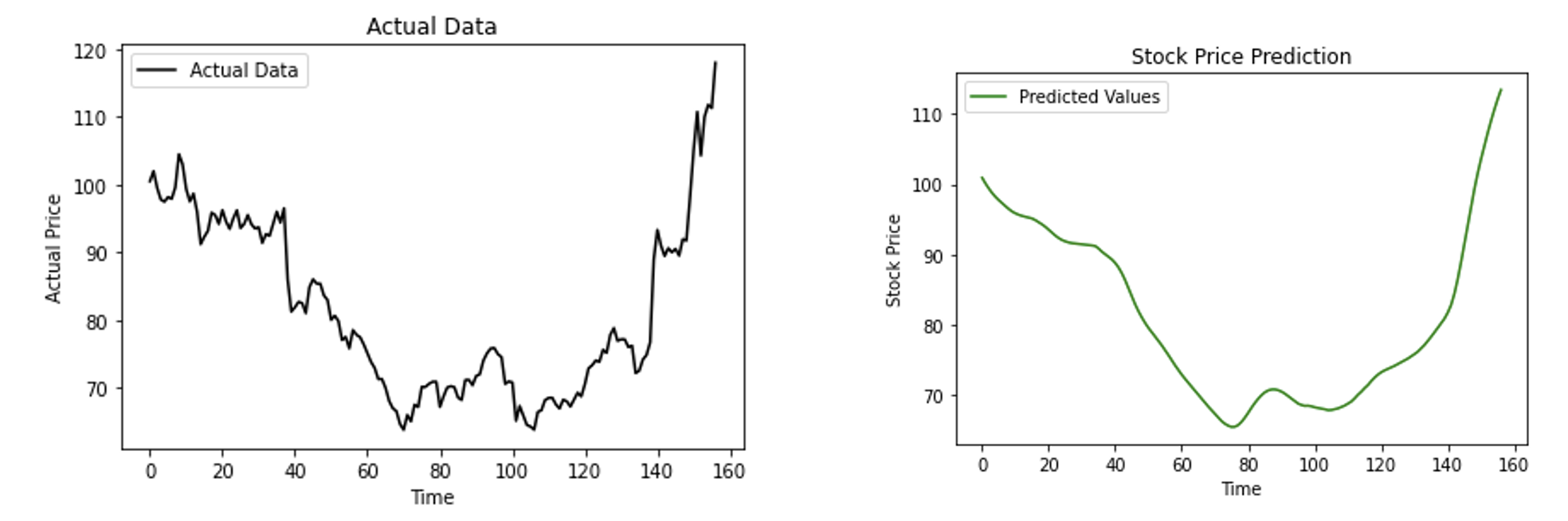
**MSFT**:



**RACE**:



**ZM**:



# Observations

**MSFT**: MSFT has performed better than predictions for the initial few days but later performed much worse than the predictions than the remaining time. Since the model is built on pre-covid data, the predictions do not include the impact due to Covid situation. LSTM model too predicts higher stock price post covid. Therefore, a lower actual stock price indicates that covid impacted MSFT negatively.

**RACE**: RACE has performed better than predictions initially but eventually became worse than predictions. Following similar reasoning, these predictions do not include Covid impact. LSTM model too predicts higher stock price post covid. Therefore, a lower actual stock price indicates that covid impacted RACE negatively.

**ZM**: We expected ZM to perform much better than predictions as it is a known fact that Zoom company has grown quite rapidly post the covid impact and the worldwide lockdown situation. However, the comparison between the actual stock values and the predictions from the neural network did not support this expectation. Moreover, the forecasted values grew much more rapidly than usual indicating an anomaly.

A re-look at the data pointed that there are much fewer data points (about 85% fewer than MSFT and 82% fewer than RACE) to train the neural network model for Zoom.

But the LSTM model confirmed the actual values of ZM and indicated an increase in the stock values of zoom further in the future – which we know is true from actual data.

# Conclusion

Since predicting much further into future requires quite high computational power, we had to limit to forecasting for 30 days. However, this was sufficient to draw some qualitative conclusions about covid impact on stock prices. The LSTM model seems to give a good qualitative idea in every case, whereas a single layer neural network worked well when the training data was large enough.

This approach seems worth investigating further. With enough data and computational power, the impact of Covid (or similar such worldwide phenomena) on different industries could be assessed by extending this approach appropriately.