**Mini Project CS6140: Machine Learning**

a) **Stock Market Prediction Using Deep Learning**

b) **Name**: Vishal Shah  
 **Email**: [shah.vishal@northeastern.edu](mailto:shah.vishal@northeastern.edu)  
  
c) **Goal**: The objective of this project is to predict the future stock prices with the help of LSTM (Long Short Term Memory cells) model. I have in particular decided to go with the FAANG stocks or stocks of tech biggies: Facebook, Amazon, Apple, Netflix and Google. The goal is to predict the closing price of each of these stocks as accurately as possible which can help in maximizing an investor’s gain.

**Importance**: Predicting stock prices is an important objective in the world of finance; this is because a really good prediction has the potential to produce high financial benefits and hedge against market risks. But the catch is, predicting stock prices is a herculean task, due to the complexity and volatility of the markets and the many unpredictable, dynamic variables involved. Many scholars from different fields have researched the pattern recognition of financial time series and have suggested various models for to predict stock prices. In order to establish substantial results, most of these ways require careful attention to deciding the input variables.  
  
**Motivation**: Algorithmic trading has changed the dynamics of stock market and its relevant industries. More than 70% of all trades happening in the US at the moment are being handled by artificial intelligence or machines. This inspired me of how I could develop my own machine learning algorithm for trading stocks.

d) **Background**a) After going through multiple papers and work on this topic, I learned that LSTMs work best for stock market prediction. “Long Short Term Memory cells are like mini neural networks designed to allow for memory in a larger neural network. This is achieved through the use of a recurrent node inside the LSTM cell. This node has an edge looping back on itself with a weight of one, meaning at every feed-forward iteration the cell can hold onto information from the previous step, as well as all previous steps. Since the looping connection’s weight is one, old memories won’t fade over time like they would in traditional RNNs.”

LTSMs and recurrent neural networks and as a result are good at working with time series data thanks to their ability to remember the past. By storing some of the old state in these recurrent nodes, RNNs and LSTMs can reason about current information as well as information the network had seen one, ten or a thousand steps ago.

b)In the past a number of mathematicians and economists have tried and applied multiple algorithms to predict the stock market. One person who will go down as a legend is probably Jim Simons, a billionaire mathematician. He is perhaps the most successful hedge fund manager with his Renaissance technologies and also the costliest. He has averaged more than anyone and consistently year after year in the stock market. His methods are very secretive, but it does make me wonder, no matter how volatile and difficult they may say it is to predict the market, with the help of Mathematics and Machine Learning we will predict the market much better than what we are aware right now.

c) I have taken some inspiration from homework 5 and have decided to go along with similar evaluation strategies and find out which model with what parameter combinations on 5 different stocks works the best.

e) **Methods**

a) I had many options as to where to obtain my data from. Out of all I found AplhaVantage and finance.Yahoo to be the most convenient and up to date sources to retrieve a stock data from. The data as well is very well organized displaying all the technical indicators.

The dataset looks something like this:



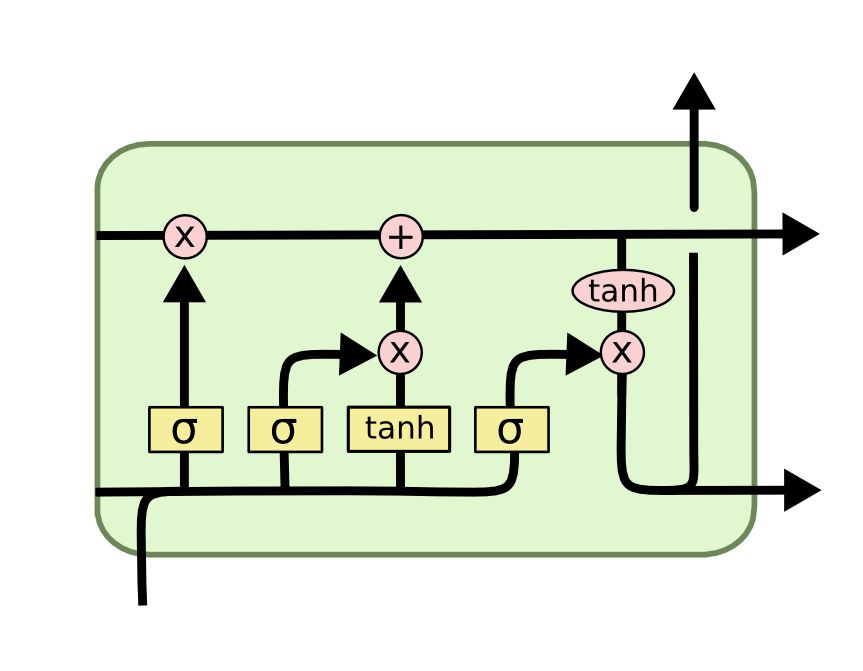
Between the two, I decided to go with finance.yahoo as my data source.

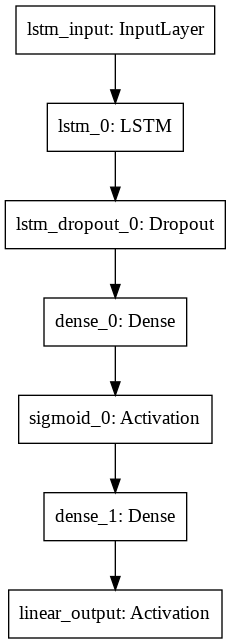
b) Methodology:   
First I will retrieve the required dataset of the particular stock from yahoo’s website. I will be taking data for the last 8 years to upto recent ones. Then I will first visualize how the graph for the stock looks like and how it has performed over the years.

After this I split the data into 10 cross folds, and also take the “close” column of dataset out, as the objective is to predict the closing price of the stock on a particular day.

After this I build my LSTM deep learning model by declaring the number of layers and neurons within each layer along with the optimization technique.  
  
Post this I fit the training data into the model I created and test out its accuracy against the test set.  
  
Once this is done, I will use RSME as the regression metric to test the accuracy of the model. I will compare this RSME with RSME I calculate from other models on the same stock.  
  
I plot the predicted data against the test data to see how well the model worked.  
  
I repeat this for the other 4 stocks.

A visual representation of LSTM:

   
The model looks like this:



c) Evaluation Strategy

I have compared different machine learning algorithms against LSTM and check if any other comes even close. Additionally, I have tried different combinations of layers and neurons to find a better result.

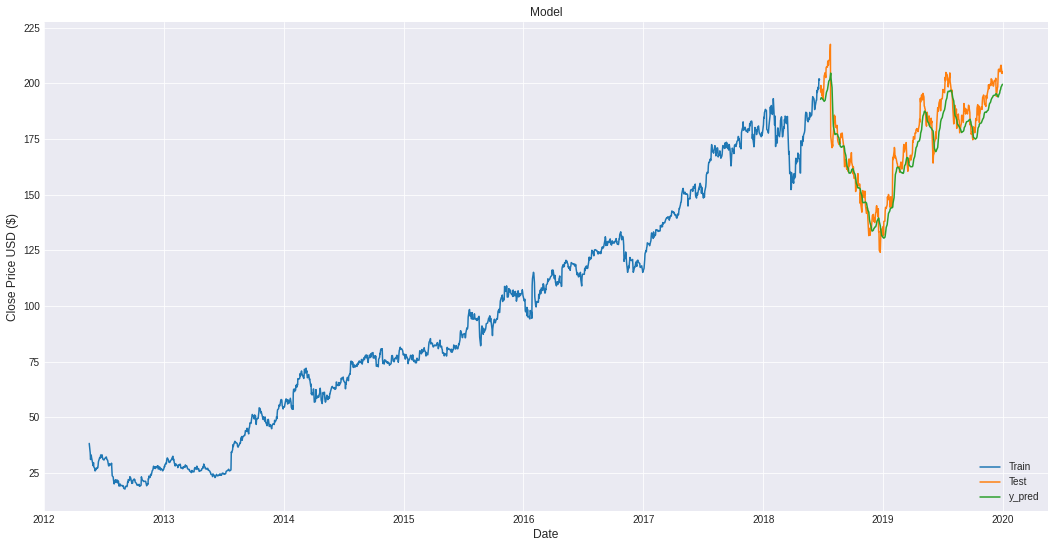
I have used RSME to compare and measure the accuracy of the model.

I visualized the performance of predictions against the test set.

F) **RESULTS**

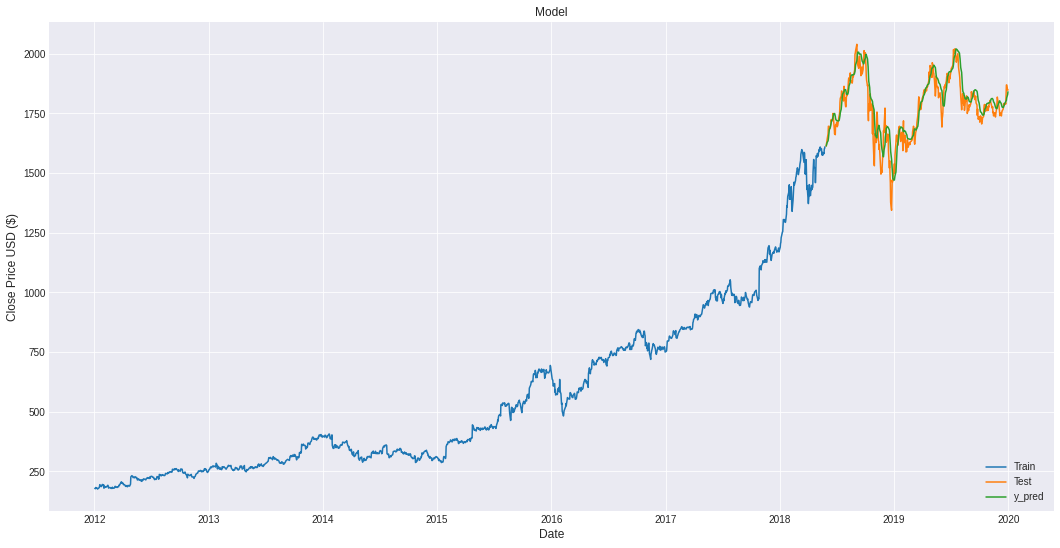
I will one by one post results of the visualizations (how predictions worked against test set) and the RSME with 68% confidence interval for my best algorithm for all the five FAANG stocks.

**Facebook (FB):**



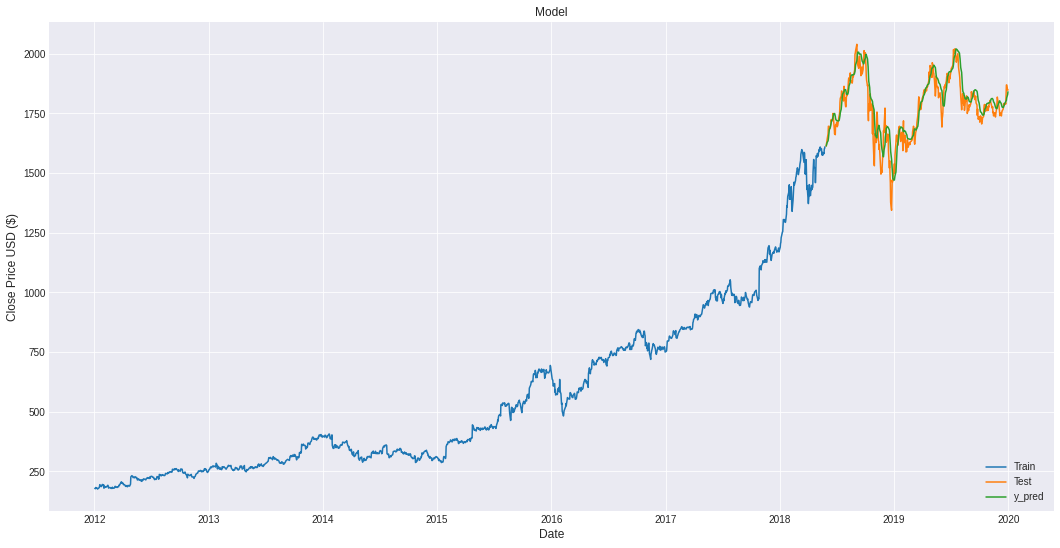
RSME: 7.53 + 1.3

**Amazon**



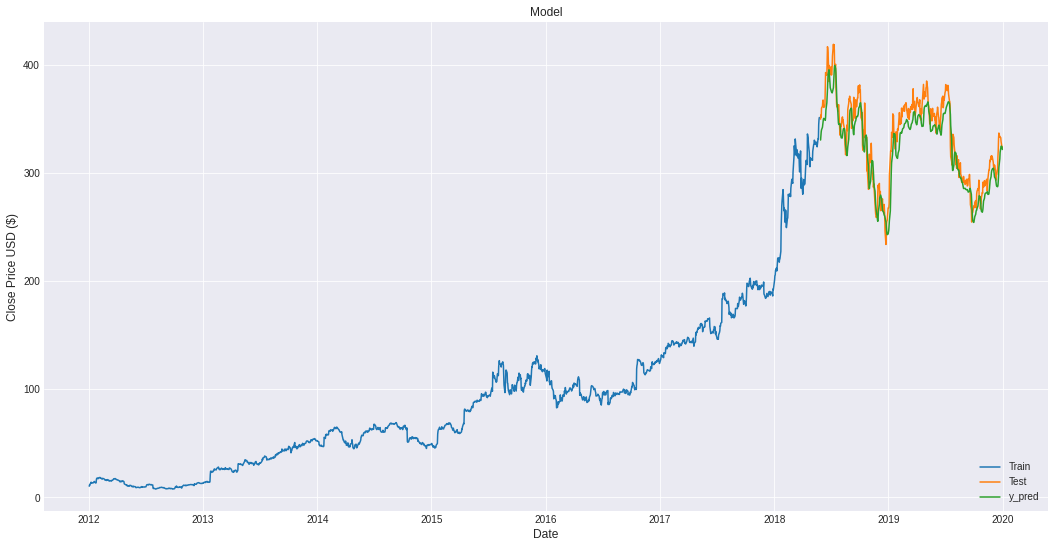
56.71 + 7.3

**Apple**



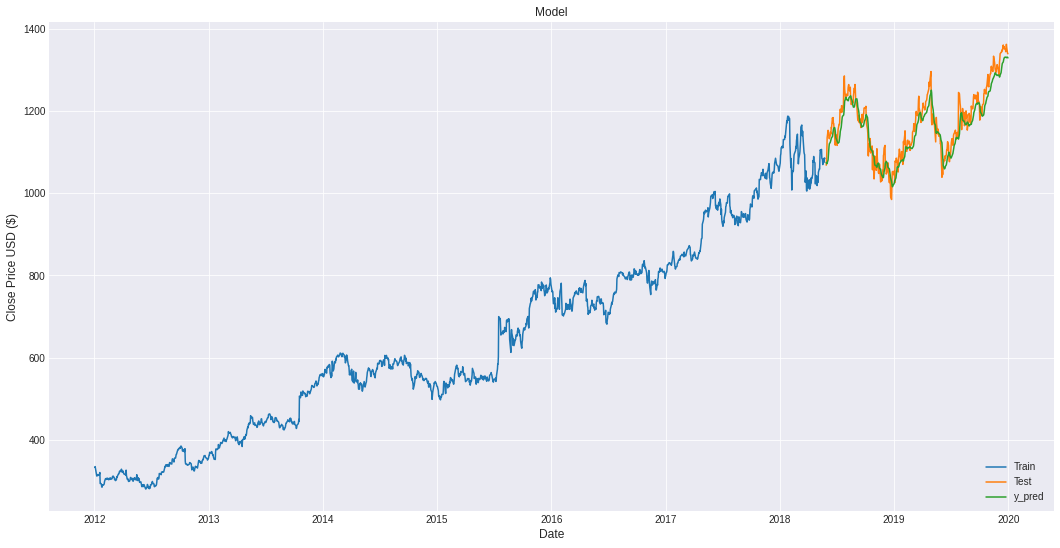
RSME: 56.71 + 5.35

**Netflix**



RSME: 16.65 + 2.87

**Google**



RSME: 30.05 + 3.14

g) **Conclusions**

I feel LSTM does a really good job on predicting the stock market price and the predictions speak for themselves.

However I would like in future to improve the model by including the intangible factors the impact stock market. For example the current news, emotions of the traders or like the most recent, impact of covid 19 situation on the stock market

h) **References**

<https://www.hindawi.com/journals/mpe/2018/4907423/>

<https://www.researchgate.net/publication/259240183_A_Machine_Learning_Model_for_Stock_Market_Prediction>  
<https://towardsdatascience.com/getting-rich-quick-with-machine-learning-and-stock-market-predictions-696802da94fe>

<https://medium.com/@randerson112358/stock-price-prediction-using-python-machine-learning-e82a039ac2bb>