

---

# CS559: TECH STOCK MARKET ANALYSIS AND PREDICTION

---

**David Fu**

Stevens Institute of Technology  
dfu6@stevens.edu

## ABSTRACT

There are many machine learning algorithms that tries to predict the stock market and make trades in order to turn profit via capital gain. There are always risk and unilateral trust on these system to make micro second decision to move millions of dollars. Trying to fully understand what the margin of error is and what pattern arises when looking into Tech Stock. Creating a simple risk and prediction model between a randomly selected stock and their relationship. Then see how accurately a simple model could predict stock trend.

## 1 Introduction

The project will be tasked with analyzing tech stock base on few simple metrics such as their gains, returns, and overall risk. Then with such information see if its valid assessment for the publicly known that many tech stock have positive correlation with each other especially FAANG stock that are the most popular. For instance, it is extremely useful even for personal investors to look at these trends given the risk tolerance of individual to make decisions base on years of data. Also as a product robo advisory has been a trend without the expense or the risk of human error but risk the bias of market data.

Focusing on three distinct dataset of CRM, AAPL, and MSFT as baseline we are able to measure the market pattern for major fortune 500 companies and their behavior. Being able to make perfectly accurate prediction would not be the purpose but to see if these algorithm is able to more properly predict market trend and changes such as increases and decreases of stock value. If the trend is relatively accurate although the net gain would not be high but it has the ability to have higher risk tolerance to preserve capital. It is a much simpler goal for average individuals who just want to be safe from major market flux.

## 2 Background

Stock movement prediction has long attracted both investors and researchers in all industry as it denotes the impact of various factors outside of the companies purview. We present a model to predict stock price movement from simply its' historical stock prices without other factors so its accuracy will not be great due to features that are not considered. Classic research relies heavily on feature engineering but thats is not the focus of the assignment but to apply Machine Learning model and how it makes an simple impact in real where where it is applied. Stock movement prediction is widely considered problematic due to the high stochasticity of the market: stock prices are largely driven by new information and emotions that is influenced by other factors. Using simple historical data we are able to just gage some trends but not necessary actual prediction. Stock movement prediction is a time series problem without considering future data without outlier events. For instance, if a company suffers from a major scandal its stock will plummet regardless of its historical success thus drastically skew stock prediction.

To fully exploit market information, simple historical data have been extracted and because of the pademic being an unquie situation its impact on the market is global so not a company specific event. With the focus on tech stock we anticipate high correlation and relation amoungs them to be able to off-set the impact of external factor not consider in this assignment.

### 3 Data

The data is acquired from Kaggle in regarding to tech stock analysis, using various reference a model is tweaked and design modified slightly to account for pandemic era data. Let dig deeper into the data set, as it focus on a few key features

Date - The full trading date the data came from show us the time table High - Highest value of the trading day Low - Lowest value of the trading day Open - Opening Price Close - Closing Price Adjusted Close - Includes after hour trading

These are pretty standard historical stock features, using these basic data, correlation, trends, and other simple statistical metric shall be derived. The mean, rolling average, and other factor to deduct certain patterns. The data is then split into 80:20 for training and testing to see how likely it is to predict the right value, then finally future data since newest data set was taken in the past additional live data set can be used for further validation and test.

### 4 Methods

In order to achieve the goal of prediction and analysis. First we have to use the existing feature and look at some patterns. Referencing to the code first we just graph the historical price data. The general shape should denote any simple pattern or market wide events that impact all stocks. Then we referencing that with volume of stock traded to see if this impacts the price in any meaningful way. Also the daily returns and delta between these historical stock then eventually these data will be fed into Risk model base on these information to see how risky each other is base on mean and standard deviation of changes. To make the final prediction it will be fed into LSTM model train and evaluated for accuracy.

### 5 Evaluation

The primary evaluation tool of the model will be accuracy but as thats not the primary goal the trend will also be measure meaning is the closing price positive or negative in relation to the daily gain trend. As long as it is correct on the trend the simple model is deem successful to preserve individual capital as lost avoidance. Also looking at future prediction beyond the original data set. We are able to justify that the model has not be overfitted and can generalize even none training data.

### 6 Conclusion

Although there is low accuracy in this model it is to be expected as its predictive nature is following historical trends only. However the positive or negative end of day pattern is pretty accurate determining addition tweak to the model to increase overall accuracy.

The low accuracy can be increase by modifying learning rate and using other technique such as relu or modified LSTM to avoid bias from outliers trading days. Base on the data we do see that the correlation between all tech stock is quite high and these changes could be added as features when predicting to increase accuracy. The common knowledge of stocks in similar industry impacting total market confidence lead to sentiment analysis of the market as a whole as oppose to just individual stock.

### References

1. <https://www.kaggle.com/tomasmantero/top-tech-companies-stock-price>