# Stock Market Analysis

Abhishek Shakwala, Pratik Bongale, Viral Parmar

## 1. INTRODUCTION

Stock market plays multiple crucial roles in the economic growth of the nation. Equity markets are primarily, financial institutions established to help businesses, organizations, and entrepreneurs come together to trade in the form of buying and selling of shares for capitalizing the enterprises in need of cash infusions. However, it is essential to observe the nature of stocks which are usually highly volatile and complex. Several factors affect such highly unpredictable nature of stocks such as political perturbation, company's future of growth and expansion, any overseas affairs, influenced by the companies belonging to the same sector, news articles, analyst calls and much more. Today any investor would be interested in understanding the variable nature of the stock market to calculate the risk of their investment in a stock and gain maximum benefits from the investments. The investors highly demand the prediction of the stock price, and hence there is a need for stock market analysis.

The fundamental idea of this project is to build a model to predict the movement of the stocks of different companies. To carry out the prediction, we propose to use the real-time data made available by quandl.com and yahoo finance. The dataset provided by these sources include information regarding the date, price of the stock at the opening of the trading day, price of the stock at the closing of the trading day, price of the stock at the closing of the trading adjusted with the dividends, highest price of the stock during the trading day, lowest price of the stock during the trading day, volume of shares traded and turnover of the stock. We will use this data in conjunction with some machine learning or data mining techniques to guess the most accurate stock prediction for the investors.

In the second phase of our project, we dive deep into developing the business understanding, data understanding, and data preparation. The primary goal of this phase is to pre-process our data by following the CRISP for Data Mining process and get clean and transformed data set which can be used for our mining goals.

Our third phase will contribute towards applying various machine learning or data mining techniques on the gathered information, building mining models, validating and verifying the models and finally produce the knowledge that will interest the investors or end users.

## 2. CURRENT WORK AND CHALLENGES

### 2.1 BUSINESS UNDERSTANDING

The business understanding phase of our project includes three tasks namely problem identification, resource identification, and problem specification. It is one of the critical steps which represents about 20% of the effort but contribute to 80% of the success in achieving the mining goals.

- Problem Identification:

  The question that we ask in this sub-task of business understanding is, what is the main issue driving the data mining effort? To answer this question, the first thing that we must do in any project is to find out exactly what are we trying to accomplish from a business perspective. The primary motivation of the stock market analysis is to understand the variable nature of the stocks market, to calculate the risk of investment in a stock and gain maximum benefits from the investments.

- Resource Identification:

  This task involves detailed fact-finding about the available resources, assumptions, constraints, and various other factors that should be considered in determining the data analysis goal. The question that we try to answer here is, what people/ data/ hardware/ software are available for our project?

- Problem Specification:

  Finally, the question we answer in this sub-task of business understanding is, what are the goals of this project? Moreover, what would success look like? The purpose of a data mining project is of two kinds: one is the business goal which states objectives in business terminology and another we have is data mining goals which aim at technical terms.

For example, for our stock market analysis, the business goal that we determined is “Whether investors should invest in particular stock or not.” A data mining goal determined for our project is, “We will attempt to predicting the value of the stock, how much value does the investor put at risk by investing in particular stock and correlation between different stock prices.” Initially, we will be focusing on the classification problem (for example: Whether the price of the stock will increase or decrease) and then we will focus on the prediction problem (for example: How much we put the user at risk by investing in a particular stock).

## 2.2 DATA UNDERSTANDING

### *2.2.1 Data Gathering*

In the first phase of the project, our major focus has been on gathering the data. We have gathered most of the stock market data sets from quandl.com and yahoo finance. We found two approaches to get our data:

- Fetching data from QUANDL and YAHOO API

  ```
  starttime = dt.datetime(2002, 1, 1)
  endtime = dt.datetime(2017, 10, 28)
  dataSheet = webData.DataReader(‘MSFT’, ‘quandl’,
  starttime, endtime)
  dataSheet.to_excel(’MSFT.xls’)
  ```

  For fetching data through API and then storing it into excelsheet required us to use pandas_datareader.data package for our project.

- Downloading Excelsheet from the QUANDL and YAHOO and parsing it with python pandas package

  ```
  dataSheet = p.read_excel(“EOD-MSFT.xls”,
  parse_dates=True, index_col=0)
  ```

  Here we choose to go with the second approach as it is easy to get the data right from the source for the specified time.

The datasets that were available to us were in .xls format (excelsheets) which then was parsed using python script. For parsing the excel data we used Pandas, an open source data analysis library which provides easy-to-use data structures and data analysis tools for the python programming language.

### *2.2.2 Understanding the data at hand*

Initially, the data that we plan to use for mining is in excel sheet, so we started understanding the scope of our data by targeting one company’s data. The consideration of targeting one company’s data that we made here is because the stock data (i.e., attributes present) for various other companies remains the same. Thus, we first got the insight of our mining goal from business understanding, and then we decided which different attribute will help us to achieve our mining goal. The intuition regarding the feature selection and what discoveries can be made from the available datasets are based on this understanding regarding the scope of data. We have revised the list of attributes that will be used to perform analysis that helps us in getting insight and how these insights help us in better prediction and decision-making.

## 2.3 DATA PREPARATION

### *2.3.1 Data Cleaning*

**Observations from the dataset:** Following is the list of problems/ observations and action taken:

1. **Records with “null” value for all the important fields such as (Open, Close, High, Low, Volume):** Remove the record altogether
2. **Missing values identified for attributes Open, Close, High, Low:** Wrote a python script to identify the missing fields and insert an estimate of missing value calculated based on the remaining available values for that record.
3. **Redundant data:** Remove redundant records using python script.
4. **Outliers:** Did not find any outliers
5. **Invalid characters:** Did not find any invalid characters

**How we handled missing values:**

- Dataset comprises of a few important fields (Open, High, low, close) which are necessary for any kind of analysis.
- If any of the above attributes have missing values we find an estimate for the value using the non-missing values in that particular record.
- Example: Suppose, “Open” price of stock is missing in a record, we follow the below steps to make a close estimate for Open price for that day:

**Table 1: Cleaning example**

| Date | Open | Highest | Lowest | Close |
|---|---|---|---|---|
| 10/27/2017 |  | 79.2 | 78.46 | 78.86 |

1. Extract all records with Close, Highest, and Lowest values that are close (+/- 5) to the Highest, Lowest, Close of the record above.
2. Filter extracted records to ensure they have “Open” values between the high, low values of current record.
3. Calculate the mean of filtered records obtained from step 2.
4. Replace the Open value with estimated open value.

### *2.3.2 Data Transforming and Formatting*

Data transformation is the process of converting the data from one format to another which is usually the conversion of format from source system into destination system. So, after understanding our mining goals we decided to create two attributes (coded attributes), from the data sets available to us. The attributes which we created are:

- **% Gain/ Loss**

  We calculated the values for this attribute by taking the difference of Close price and Open price for the present day and that difference we divided by Close price of previous day. To get the percentage value we then multiplied the division value by 100 which returned us true % Gain / Loss for that particular stock given the present day.

  ```
  previousClose = dataSheet[“Close”].shift(1)
  dataSheet[“%Gain/ Loss”] = ((dataSheet[“Close”] -
  dataSheet[“Open”])/previousClose)*100
  ```

  The challenge that we faced here was after we created the attribute, for the first record in our dataset there was no previous close present, so our dataset recorded NaN (Not a Number) which we then replaced it with 0.

- **90 Days moving average**

  We calculated the values for this attribute by taking the average of 90 days Adjusted Closing price.

  ```
  dataSheet[“90DayMA”] = dataSheet[“Adj_Close”]
  .rolling(window=90).mean()
  ```

  The challenge that we faced here was after we created the attribute for the first 90 records the value stored in the data set was NaN since we had no relative 90 days adjusted closing price for the stock. We resolved this issue by taking the 90DayMA for the first day same as Adj_Close and then we kept on taking the average based on available 90DayMA for previous days until 89th day and after that we considered the actual 90 Days moving average.

  ```
  dataSheet[“90DayMA”] = dataSheet[“Adj_Close”]
  .rolling(window=90, min_periods=0).mean()
  ```

  Similarly, we can calculate the moving averages of 30 days and 60 days which will help us in visualizing the up trend or down trend of a particular company's stock.

### *2.3.3 Data Reduction*

Data reduction is the process of minimizing the amount of data that are required to store in the data set. It can increase the efficiency and reduces the cost to parse the data as well reducing the complexity to consider them for mining. The data reduction technique that we used for our project is data deduplication, which eliminates redundant data from the data set.

```
dataSheet = p.read_excel(“EOD-MSFT.xls”,
parse_dates=True, index_col=0)

dataSheet = dataSheet.drop_duplicates()
```

We also removed one attribute from our original dataset named “Adj_Volume” since we already had Volume attribute present for each record and keeping in mind our mining goals it was intuitive to remove Adj_Volume from the mining data set as it might not contribute well towards our mining goal.

## 3. DATA ANALYSIS AND MINING

In CRISP methodology, the mining phase basically involves two steps: identifying the data mining method and discovering the pattern within the dataset. There are different methods available for mining the dataset which are used currently and are dependent on the data mining tasks. Also, the data mining method we choose is highly dependent on the goals and the user of the results.

## 3.1 CORRELATION BETWEEN STOCKS

Correlation is used to define the strength of two variables. Here we find the correlation between various stocks from Apple, Microsoft, Amazon, Google stocks. We calculate the correlation coefficient between stocks to analyze the relation between them.

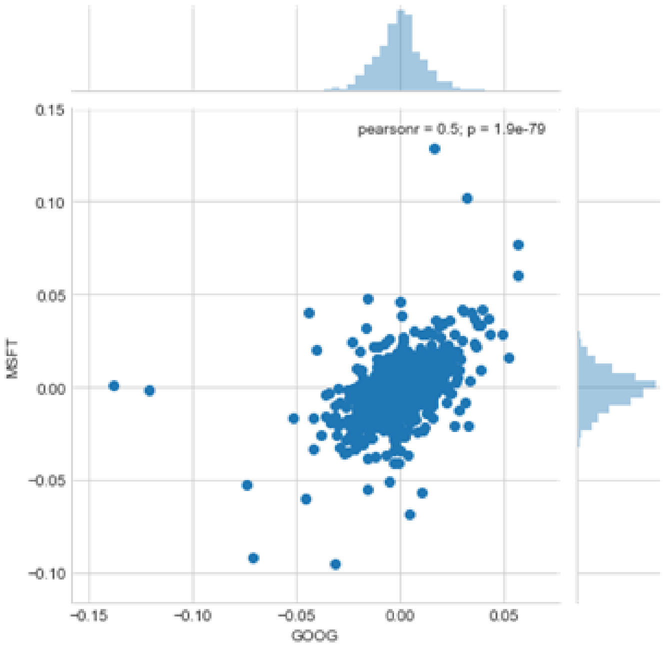


**Figure 1: Correlation graph between Microsoft and Google**

Here we take data from Yahoo finance API making use of pandas datareader to read and store the data. The dataset contains records of past five years the present day. We make use of various external libraries such as numpy to make the computation easy, pandas for storing data frame and reading CSV file, for visualization we use seaborn and matplotlib for plotting graph. In figure 1, we have calculated the pearson correlation coefficient. We see the there is a positive correlation between Microsoft and Google.

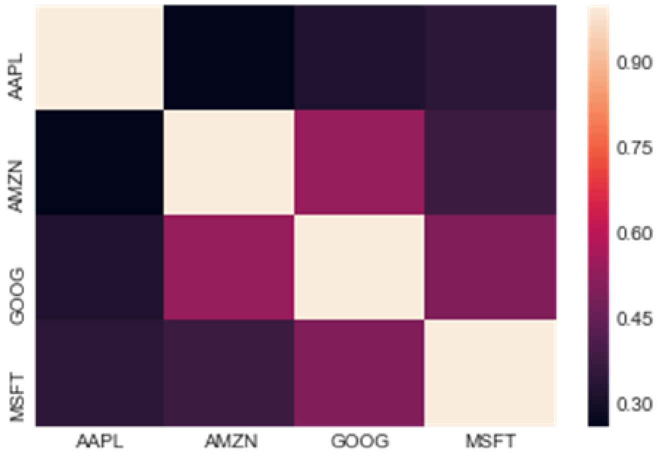


**Figure 2: Heat Map**

In figure 2, we can see the correlation between various stocks. Apparently, there is a perfect correlation when stock compared with itself. We also see there is a high correlation between Amazon and Google. After comparing the various stocks from above mentioned, we can say that there is a positive correlation of the stocks in the same sector. For example, Microsoft, Amazon, Google, Apple shows a positive relation.

## 3.2 S&P 500 RETURNS PREDICTION

The US stock market is best represented by Standards & Poor's 500 indices. S&P 500 is an American Stock market index which is based on the market capitalizations of 500 companies having common stock listed on the NYSE or NASDAQ.

S&P 500 is the best indicator of trends in the US economy, so, we decided to do our best to analyze the historical trends and make a prediction (better than chance) about the change in the market in the next few minutes. That means whether the index will rise or fall based on the change in stock prices of top 5 companies (Microsoft, Amazon, Facebook, Google, Apple) as per the S&P 500 index. We chose only the top 5 because we observed that these companies are major contributors to fluctuations in the S&P index. We used the deep learning technique for prediction. The model used is a Multi-Layer Perceptron Neural Network.

The total S&P 500 index price with stock prices of 500 companies from April 2017 to August 2017 in minutes (to keep our analysis details and very low level).
Number of records: 41266
Number of attributes: 501 (500 companies + 1 S&P aggregated index)

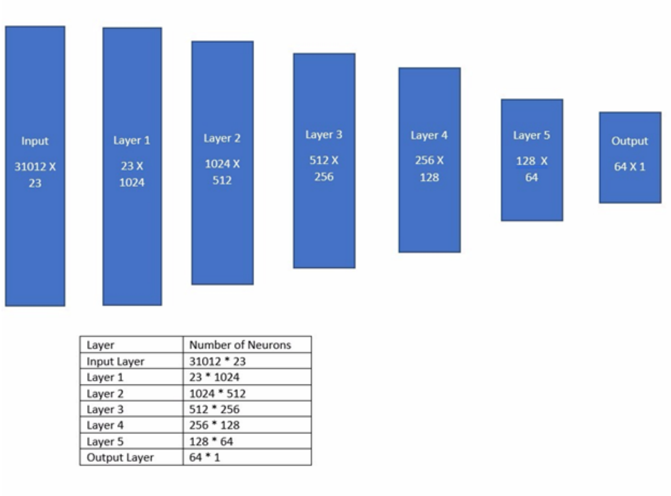
To obtain better features suitable for our analysis(and the Neural Network), we decided to derive a set of attributes which can better predict the returns from S&P index.

1. Percent change: change in stock prices every minute.
2. Average change: on an average how much do the returns from each stock change over past 10 minutes.
3. Delta change: how much did the stock price as compared to price 10 minutes earlier.
4. Delta Lag(makes prediction possible): shift the obtained returns 1 minute in the future so that when we find the S&P index returns for 09:05 AM we are looking at the returns from S&P index at 09:06 AM.

Steps for data preparation is as follows:

1. We dropped the 495 columns of companies which we were not considering for the analysis.
2. Compute every minute returns from stocks of all 5 companies and from S&P index to know the percent change in prices as compared to last minute.
   Pandas function used:
   pct_change(1) # change in last one minute
   pct_change(10) # change in last 5 minutes
3. To see the average fluctuations occurred in past 10 minutes, we compute rolling mean of returns over last 10 minutes.
   Pandas function used:
   rolling(window=10).mean()
4. Compute delta lag using function dataframe[col_name].shift(1) shifts the entire dataframe column by one row.

### *3.2.1 Multi-Layer Perceptron Neural Network*



| Layer | Number of Neurons |
|---|---|
| Input Layer | 31012 * 23 |
| Layer 1 | 23 * 1024 |
| Layer 2 | 1024 * 512 |
| Layer 3 | 512 * 256 |
| Layer 4 | 256 * 128 |
| Layer 5 | 128 * 64 |
| Output Layer | 64 * 1 |

**Figure 3: Neural Network Architecture**

The model is trained on 80% of training data(31012 records) and tested using the remaining 20% data(10610). We used Google's tensorflow library and referred the Neural net model created by STATWORX[2].

We performed transform learning on the model by adding a new hidden layer, and tuning several parameters to suite our prediction goals. The neural network is built to reduce the Sum of Squared Errors between the predicted returns of S&P index and true returns known to us from the data.

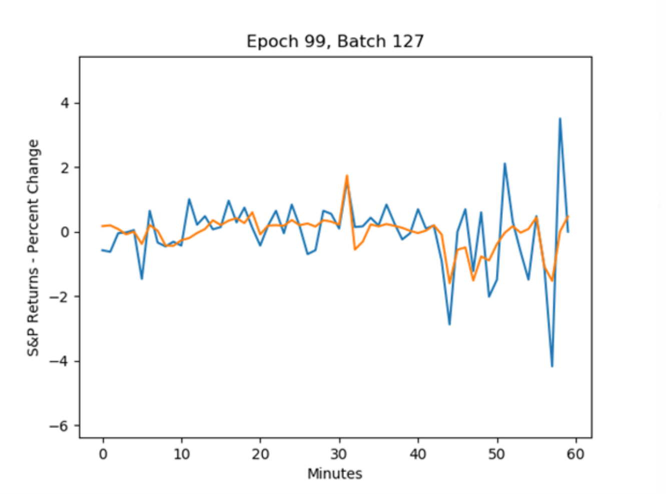


**Figure 4: Neural Network1**

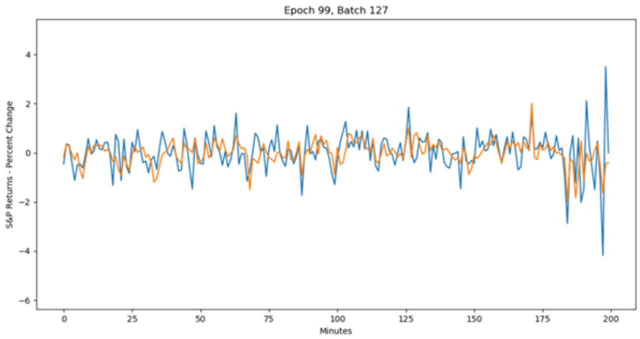


**Figure 5: Neural Network2**

The figures 4, 5, 6 and 7 the result shows the plot of S&P Returns-Percent Change VS Minutes. The blue line shows the Actual value of S&P500 value and orange line shows the predicted value of S&P500. Model is trained on training data which is divided into batches of size 256. The following factors were turned for obtaining the better accuracy and efficiency; batch_size, number of neurons in each layer, number of epochs, number of hidden layers. As observed from the above results that the model predicts the peaks and troughs of the actual value accurately.

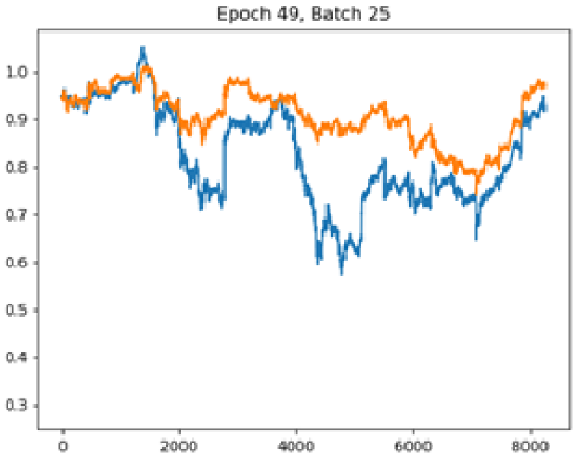


**Figure 6: Neural Network3**

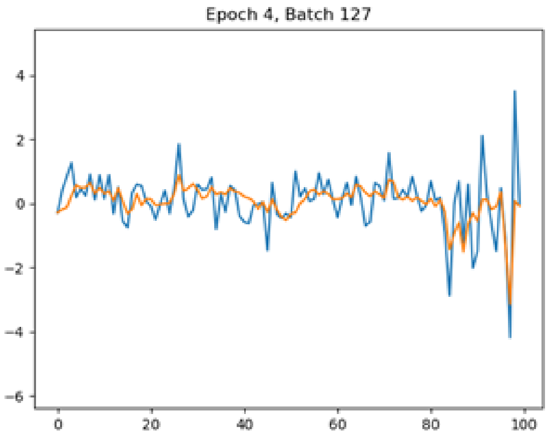


**Figure 7: Neural Network4**

Artificial neural network(ANN) plays a crucial role in predicting stock prices because ANN has the capability of learning and correcting errors by itself with the help of gradient descent and back propagation algorithm. Using back propagation ANN adjusts the weights and bias to minimize the error in turn increasing the accuracy. If the results are accurate it will be beneficial for the bankers and investors predict and analyze the next min's price of the stock. It will help an individual to reduce the loss.

## 3.3 INVESTMENT RISK ANALYSIS

Investment risk analysis focuses on the underlying uncertainty for the action made by the investors and refers to the uncertainty of forecasted cash flow, variance of stock returns, the probability of making profits or loss and possible future economic states. Here, we are particularly looking at the technology stocks like Microsoft, Google, Facebook, IBM, Apple and Amazon for performing the value at risk for the given stocks. We treat the value at risk as the amount of money we could expect to lose or putting at risk of loss for a given confidence interval.

There are several methods through which we can estimate the value at risk but here we will look into two methods which we have used in our prediction: bootstrap and monte carlo method.

### 3.3.1 *Bootstrap Method*

In data mining, bootstrapping is the test or metric that relies on random sampling with replacement. Bootstrapping can be done by assigning the measures of different accuracy parameters such as bias, variance, confidence intervals and prediction error, to sample estimates. It is the method which allows estimation of the sampling distribution of different statistic using random sampling methods. Using bootstrapping for stock market prediction is of great advantage because of its simplicity. It is a straightforward method for deriving the estimates of standard errors and confidence intervals for complex estimators such as proportions, percentile points, and correlation coefficients.

For predicting the risk of investing into given stocks we are considering 0.1 empirical quantile of daily returns i.e., the percentage change adjusted closing price. That means our prediction estimates the risk with 90% confidence level.

```
closingDataFrame = webData.DataReader([’MSFT’, ’AAPL’, ’IBM’, ’GOOG’, ’AMZN’], ’yahoo’, starttime, endtime)[’Adj Close’]
pctChange = closingDataFrame.pct_change()

cleanPctChange = pctChange.dropna()

ibmQuantile = cleanPctChange[’IBM’].quantile(0.1)

ibm = investmentAmount * abs(ibmQuantile)
```

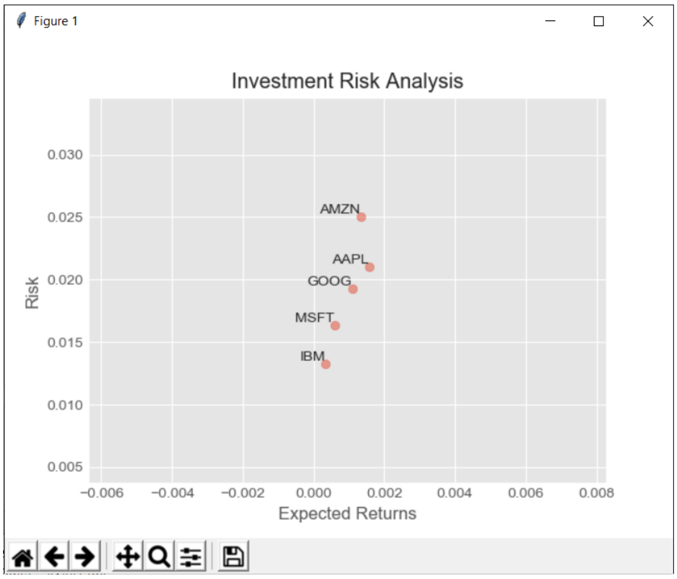
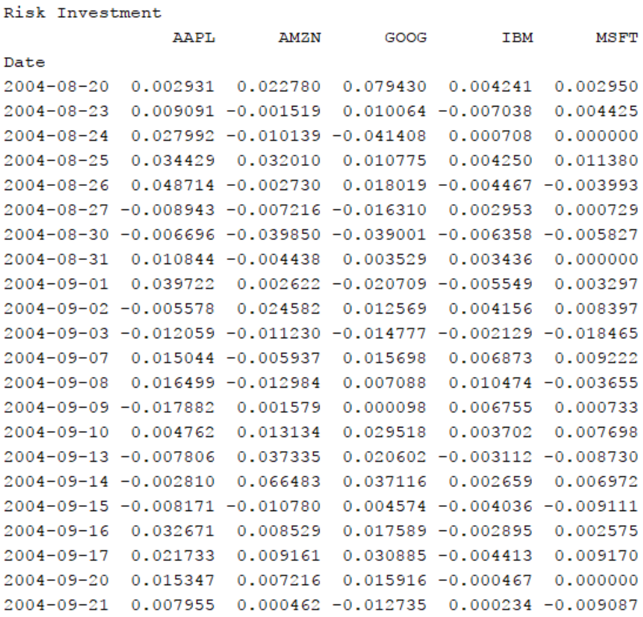


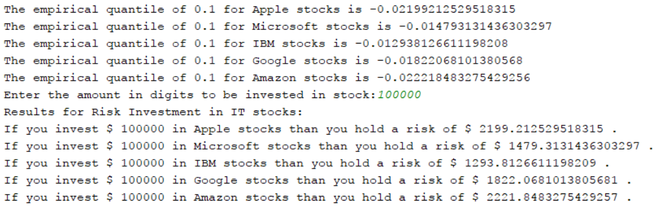
**Figure 8: Risk Analysis**

In figure 8, we have made the comparison between the risk and the expected returns using the technology stock data as shown in figure 9. Here, we have considered the adjusted closing price for the technology stocks and we have find the percentage change for each stock which is same as the daily returns. The scattered plot in figure 8 shows the mean change on x-axis (i.e., Expected Returns) and standard deviation on the y-axis (i.e., Risk).



```
Risk Investment
                AAPL      AMZN      GOOG       IBM      MSFT
Date
2004-08-20  0.002931  0.022780  0.079430  0.004241  0.002950
2004-08-23  0.009091 -0.001519  0.010064 -0.007038  0.004425
2004-08-24  0.027992 -0.010139 -0.041408  0.000708  0.000000
2004-08-25  0.034429  0.032010  0.010775  0.004250  0.011380
2004-08-26  0.048714 -0.002730  0.018019 -0.004467 -0.003993
2004-08-27 -0.008943 -0.007216 -0.016310  0.002953  0.000729
2004-08-30 -0.006696 -0.039850 -0.039001 -0.006358 -0.005827
2004-08-31  0.010844 -0.004438  0.003529  0.003436  0.000000
2004-09-01  0.039722  0.002622 -0.020709 -0.005549  0.003297
2004-09-02 -0.005578  0.024582  0.012569  0.004156  0.008397
2004-09-03 -0.012059 -0.011230 -0.014777 -0.002129 -0.018465
2004-09-07  0.015044 -0.005937  0.015698  0.006873  0.009222
2004-09-08  0.016499 -0.012984  0.007088  0.010474 -0.003655
2004-09-09 -0.017882  0.001579  0.000098  0.006755  0.000733
2004-09-10  0.004762  0.013134  0.029518  0.003702  0.007698
2004-09-13 -0.007806  0.037335  0.020602 -0.003112 -0.008730
2004-09-14 -0.002810  0.066483  0.037116  0.002659  0.006972
2004-09-15 -0.008171 -0.010780  0.004574 -0.004036 -0.009111
2004-09-16  0.032671  0.008529  0.017589 -0.002895  0.002575
2004-09-17  0.021733  0.009161  0.030885 -0.004413  0.009170
2004-09-20  0.015347  0.007216  0.015916 -0.000467  0.000000
2004-09-21  0.007955  0.000462 -0.012735  0.000234 -0.009087
```

**Figure 9: Technology Stock Data**



```
The empirical quantile of 0.1 for Apple stocks is -0.02199212529518315
The empirical quantile of 0.1 for Microsoft stocks is -0.014793131436303297
The empirical quantile of 0.1 for IBM stocks is -0.012938126611198208
The empirical quantile of 0.1 for Google stocks is -0.01822068101380568
The empirical quantile of 0.1 for Amazon stocks is -0.02221848327542925б
Enter the amount in digits to be invested in stock:100000
Results for Risk Investment in IT stocks:
If you invest $ 100000 in Apple stocks than you hold a risk of $ 2199.212529518315 .
If you invest $ 100000 in Microsoft stocks than you hold a risk of $ 1479.3131436303297 .
If you invest $ 100000 in IBM stocks than you hold a risk of $ 1293.8126611198209 .
If you invest $ 100000 in Google stocks than you hold a risk of $ 1822.0681013805681 .
If you invest $ 100000 in Amazon stocks than you hold a risk of $ 2221.848327542925 7 .
```

**Figure 10: Risk Amount**

In figure 10, the quantile value that we have considered here is 0.1 for each of the stock. That means our prediction answers the risk of investing the money into particular stock with 90% confidence interval.

### 3.3.2 *Monte Carlo Method*

The Monte Carlo method is used in stock price risk prediction to simulate the uncertainty that affects the value of stocks and then calculate a representative value given the possible values of the underlying inputs. We have used Monte Carlo to run many iterations with random stock conditions, and then we calculated the stock losses for each iteration. Once this is done, we have aggregated all these simulations to predict how risky the stock is.

Here we have determined the change in stock price as the current stock price multiplied by average daily return which is multiplied by the change of time and shock which will randomly push stock price. The figure 11 discusses the results for facebook’s stock prediction over the period of 6 months which is 180 days approximately and shows the variation of

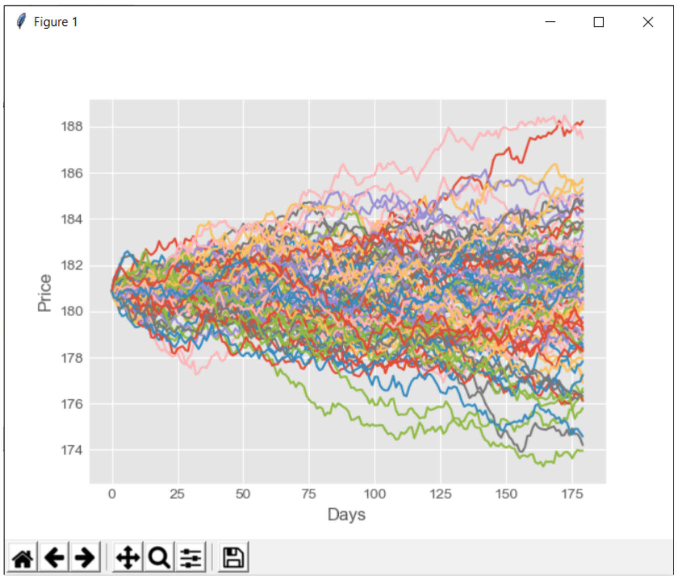


**Figure 11: Facebook Risk Prediction**

stock.

## 4. ETHICAL ISSUES WITH DATA ANALYTICS

The stock market data is publicly available and easily accessible over the internet (Yahoo Finance provides us around 60 years of stock market prices). The collection of this finance data is unrestricted and does not evade any individual's or company's privacy, so we can say that it is ethical to store or access this data.

Almost all finance firms keep building models which can make accurate predictions about the stock market. To build these models, they are always on the look-out for the best features to feed their models. Along with historical stock prices, these features may include data from social media (Twitter, Facebook, LinkedIn), product purchase data, customer expenditure trends, employee satisfaction data and more. When this kind of information is used for predicting the fluctuations in the stock market, it is mostly done without the information provider's consent which is ethically incorrect.

## 5. FUTURE SCOPE

Currently, we are making use of multi-layer perceptron neural network. In future, we can use more refined configuration or hyperparameter to predict the stock price. These might include the number of hidden layers, number of neurons in the hidden layer. We can combine various other algorithms along with neural network for better results. As it is a time series problem, we can make use of the Recurrent Neural network, Long Short-Term Memory RRN in future, as RNN and LSTM-RNN are most suitable for time series problem. In future, we can also take various other factors into consideration, i.e., sentiment analysis, festive season such as Christmas might affect the prices of the stock due to increase in sales, globalization also plays a significant role. Presently, no country's stock market index is independent of other country's stock index. For instance, if the stock market index of the United States is affected severely, it will affect other country's indexes as well.

## 6. REFERENCES

[1] National Stock Exchange of India Dataset, URL: www.quandl.com/data/NSE-National-Stock-Exchange-of-India

[2] https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877

[3] Stock Market Dateset, URL: www.finance.yahoo.com/quote

[4] Nasdaq Dataset, URL: www.quandl.com/data/NASDAQOMX-NASDAQ-OMX-Global-Index-Data

# APPENDIX

## A. INDIVIDUAL CONTRIBUTIONS

Following is the individual contribution of team members contributing to the project team Big D:

1. **Abhishek Shakwala**

   Phase 1:

   In this phase, I came up with an idea of Stock Market Prediction and discussed my idea with Pratik and Viral in detail. I also contributed to the topic Employee Attrition Prediction suggested by Pratik and Viral.

   Phase 2:

   In this phase, I researched about gathering the data for stock market prediction and came up with Quandl.com resource. I also contributed towards cleaning and pre-processing of data such as removing NaN, reducing the attributes which were not useful for our prediction, transforming data by adding %Gain/ Loss and 90Days Moving Average.

   Phase 3:

   In this phase, I performed the investment risk analysis using bootstrapping and monte carlo method for risk analysis.

2. **Pratik Bongale**

   Phase 1:

   In this phase, I came up with an idea of Employee Attrition Prediction and discussed the idea with Abhishek and Viral. I also contributed to the topic Stock Market Prediction suggested by Abhishek.

   Phase 2:

   In this phase, I researched about gathering the data for stock market prediction and came up with Yahoo Finance resource. I also contributed towards cleaning and pre-processing of data such as removing redundant data, removing null records, outliers and invalid characters.

   Phase 3:

   In this phase, I performed the S&P 500 returns prediction using Multi-Layer Perceptron Neural Network.

3. **Viral Parmar**

   Phase 1:

   In this phase, I came up with an idea of Employee Attrition Prediction and discussed the idea with Abhishek and Pratik. I also contributed to the topic Stock Market Prediction suggested by Abhishek.

   Phase 2:

   In this phase, I researched about gathering the data for stock market prediction and came up with Yahoo finance and Google finance resource. I also contributed towards cleaning and pre-processing of data such as identifying missing values for the attributes, removing redundant data, removing null records and outliers.

   Phase 3:

   In this phase, I performed the correlation analysis between various technology companies. I also worked closely with Pratik to get the S&P 500 returns prediction.

## B. USER INSTRUCTIONS

1. **Analyzing Correlation:**

   Filename: correlation.py

   Dependencies: Pandas, Matplotlib, Numpy, Seaborn, DateTime

   Run the python file correlation.py

2. **Running ANN:**

   Files provided:

   stock_market_prediction_ANN: Build and run Artificial Neural Network to predict returns on S&P 500 index.

   clean_stocks_dataset: Clean the original dataset(Drop irrelevant attributes)

   prepare_stocks_dataset: Prepare cleaned dataset for analysis(Derive necessary attributes)

   Dependencies:

   Python 3.6, Tensorflow, Pandas, NumPy, Matplotlib

   The original dataset file is too large( 50MB) and can be found here:
   http://files.statworx.com/sp500.zip

   We did not submit the original dataset, however, we have provided the cleaned dataset. To observe the progress of the neural network trying to fit the data, run the python file: stock_market_prediction_ANN

   The output will produce a plot showing the progress or training and once the network is trained, the Sum of Squared errors will be printed.

3. **Investment Risk Analysis:**

   File Description:

   MSFT.xlsx : The original excel file after fetching the data from quandl.com

   Transform_Clean_MSFT.xlsx: File returned by the program after cleaning "NaN" values and redundant data.

   Transform_Unclean_MSFT.xlsx: File returned by the program after generating two new attributes: % Gain/ Loss and 90DayMA.

   StockReading.py: Program that reads the excelsheet, generates two new attribute, removes redundant data and removes "NaN" values.

   riskInvestment.py: Program that reads the data from Yahoo finance and performs bootstrapping and monte carlo method for analyzing the risk of investment into the technology stocks.