Stock Market Prediction and Analysis

Presented By: Shruti Suryawanshi, UMID- 41272646

Abstract-- **Stock Market Analysis and Prediction is the project on technical analysis, visualization, and prediction using data provided by Yahoo Finance. By looking at data from the stock market, particularly some giant technology stocks and others. Use pandas to get stock information, visualized different aspects of stocks, and finally looked at a few ways of analyzing the risk of a stock, based on its previous performance history. Will try to predict future stock prices through a Monte Carlo method.**

Index Terms—**Numpy, Pandas, Data Cleaning, Data pre-processing, Monte Carlo Method, Data Visualization, Data Prediction.**

I. INTRODUCTION

This project is to collect the stock information for some previous years and then accordingly predict the results for the predicting what would happen next. In this project, I am trying to review the possibility to apply two-known techniques which is Monte Carlo simulation and data-mining in stock market prediction. Extract useful information from a huge amount of data set and use data mining which can predict future trends and behaviors through Monte Carlo simulation. Therefore, combining both these methods could make the prediction much suitable and reliable. The most important for predicting stock market prices is Monte Carlo technique because it can learn nonlinear-mappings between inputs and outputs. It mainly focuses on Monte Carlo simulation technique to predict future prices.

Analysis of stocks using data mining will be useful for new investors to invest in stock market based on the various factors considered by the software. Stock market includes daily activities like Sensex calculation, exchange of shares. The exchange provides an efficient and transparent market for trading in equity, debt instruments and derivatives. The project encompasses the concept of Data Mining and Statistics. This project makes heavy use of NumPy, Pandas, Matplotlib and Data Visualization Libraries. The Data set I have got from Yahoo Finance of 6 months for 10 different companies.

II. METHODOLOGY

This section gives the detailed analysis of each step involved in the process of building this model. Each sub section is mapped to one of the stages in the project.

A. Data Gathering & Pre-Processing:

Yahoo Finance is a good source for extracting stock market data and is available for free. Using Pandas DataReader, I have extracted data of IT stock for 6 months and stored years for predicting future stock prices using time series analysis. The attributes that we are primarily concerned with are Open, High, Low, Close, Volume & Adjacent Close.

Next, I have gone over how to handle requesting stock information with pandas, and how to analyze basic attributes of a stock and play around with the FL (Foot Locker) Stock DataFrame to get a feel for the data. After seeing the DataFrame, went ahead and plot out the volume and closing price of the FL (Foot Locker) stocks.

B. Exploratory Analysis:

The moving average (MA) is a simple technical analysis tool that smooths out price data by creating a constantly updated average price. The average is taken over a specific period, like 10 days, 20 minutes, 30 weeks, or any time the trader chooses. There are advantages to using a moving average in your trading, as well options on what type of moving average to use.

Analyzing the IT Stocks together to determine any correlation relationship between them. The correlation coefficient of 0.6 indicates a strong correlation between the daily return values of all companies' stocks which I have used. This analysis can further be extended by first segmentation of stock based on industry and then further analysis can be done to determine the choice of stock within that industry.
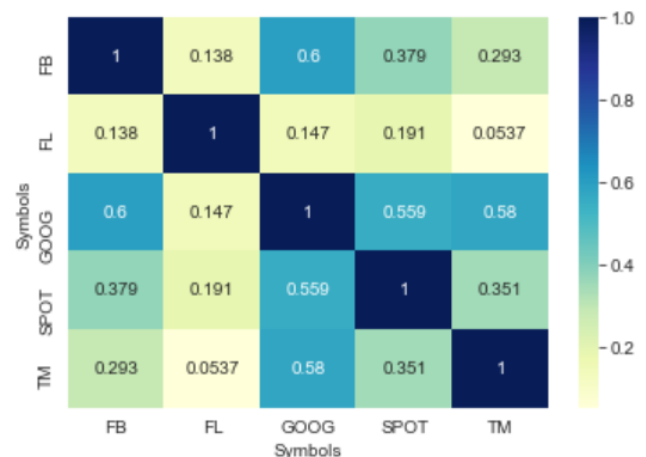

Fig. 1. Correlation Plot of Company Stocks

C. Risk Analysis Model:

I have determined the Value at Risk using the Monte Carlo method. Using the Monte Carlo for running multiple trials with random market conditions, we can evaluate portfolio losses for each random trial. Next, I aggregate all these simulations to determine the risky in the stock. Start with a brief explanation of what we're going to do. Geometric Brownian motion (GBM) method was used, which is technically known as Markov process. This implies that the stock price follows a random walk and is consistent with the weak form of the efficient market hypothesis (EMH): past price information is already incorporated, and the next price movement is "conditionally independent" of past price movements. This means that the past information on the price of a stock is independent of where the stock price will be in the future, basically meaning, you can't perfectly predict the future solely based on the previous price of a stock. The equation for geometric Brownian motion is given by the following equation:

$$\frac{\Delta S}{S} = \mu \Delta t + \sigma \epsilon \sqrt{\Delta t}$$

D. Analysis of various attributes:

1. Close:

The close is the end of a trading session in the financial markets when the markets close. It can also refer to the process of exiting a trade or the final procedure in a financial transaction in which contract documents are signed and recorded. The graph describes each day closing price of Nike stock.
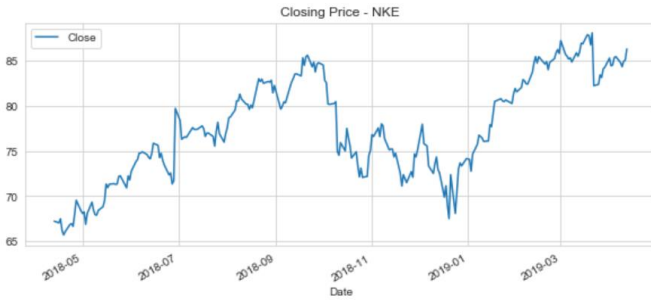

Fig. 2. Closing price of Nike Stock
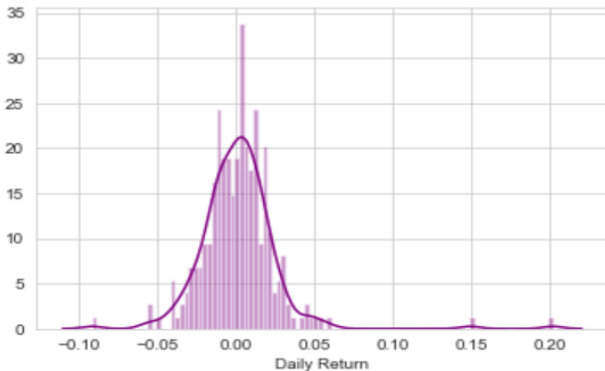
2. Daily Return Analysis:

Fig. 3. Histogram for Daily return analysis for Foot Locker Stock

The intraday return is one of the two components of the total daily return generated by a stock. Intraday return measures the return generated by a stock during regular trading hours, based on its price change from the opening of a trading day to its close. Intraday return and overnight return together constitute the total daily return from a stock, which is based on the price change of a stock from the close of one trading day to the close of the next trading day.

From the above histogram 0.05 empirical quantile of daily returns is at -0.0315. That means that with 97% confidence, our worst daily loss will not exceed 3.15%. If we have a 1-million-dollar investment, our one-day 5% VaR is 0.0315 * 1,000,000 = $31,500.

3. Volume:

It is simply the amount of shares that trade hands from sellers to buyers as a measure of activity. If a buyer of a stock purchases 100 shares from a seller then the volume for that period increases by 100 shares based on that transaction. Volume is an important indicator in technical analysis as it used to measure the worth of a market move. If the markets have made strong price move either up or down the perceived strength of that move depends on the volume for that period. The higher the volume during that price move the more significant the move. The Volume graph of Foot Locker stock has been shown below.
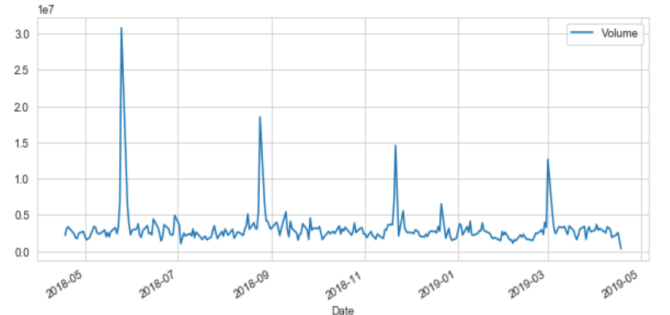

Fig. 4. Volume of Foot Locker Stock

4. Rolling Mean:

In stock applications a rolling mean is the unweighted mean of the previous n data. However, in science and engineering the mean is normally taken from an equal number of Data on either side of a central value. This ensures that variations in the mean are aligned with the variations in the data rather than being shifted in time. An example of a simple equally weighted running mean for a n-day sample of closing price is the mean of the previous n days' closing prices.

If those prices are Pm, Pm-1, Pm-2, …. Pm-(n-1), then the formula is

$$\bar{p}_{SM} = \frac{p_M + p_{M-1} + \cdots + p_{M-(n-1)}}{n}$$
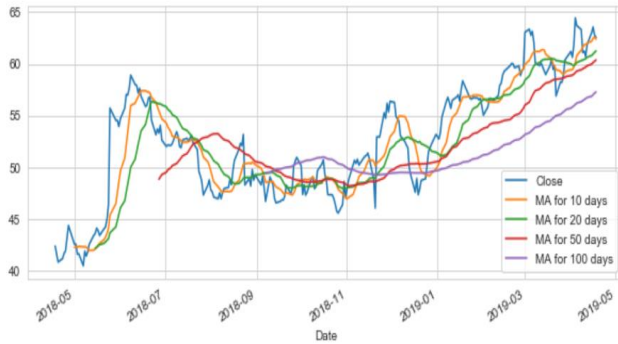$$= \frac{1}{n} \sum_{i=0}^{n-1} p_{M-i}$$

Fig. 5. Rolling Mean for Foot Locker Stock

And a simple plot for Foot Locker stock is as given above.

5. Tabular Form:

The tabular describes the various attributes discussed so far in terms of a table. It gives various discrete statistical parameter of each of these attributes. The parameters include count, mean, minimum, maximum, standard deviation, percentages etc.

	High	Low	Open	Close	Volume	Adj Close
count	252.000000	252.000000	252.000000	252.000000	2.520000e+02	252.000000
mean	93.743849	92.301468	92.982262	93.026111	8.278943e+06	92.053714
std	5.798188	5.708183	5.775474	5.735499	4.071143e+06	6.080820
min	82.519997	81.779999	81.959999	82.400002	4.026300e+06	81.043465
25%	88.164999	86.822502	87.462503	87.637499	5.913000e+06	86.128222
50%	95.545002	94.139999	94.764999	94.814999	7.478650e+06	94.026165
75%	98.295000	96.805000	97.589996	97.499998	9.363450e+06	96.931339
max	106.209999	104.809998	105.500000	105.559998	4.263130e+07	104.425194

Fig. 6. Tabular form of Walmart Stock

6. Risk Analysis:

Symbols	FB	FL	GOOG	SPOT	TM
Date					
2018-04-19	0.010459	-0.018388	0.014570	-0.009296	-0.004494
2018-04-20	-0.010827	-0.018252	-0.013552	0.032449	0.005759
2018-04-23	-0.002646	0.008072	-0.005135	-0.002588	-0.001238
2018-04-24	-0.037084	0.012133	-0.044470	-0.018793	0.003099
2018-04-25	0.000000	0.005514	0.001177	-0.011221	0.012590

Fig. 7. Percentage change in the stock values.

There are many ways we can quantify risk, one of the most basic ways using the information I have gathered on daily percentage returns are by comparing the expected return with the standard deviation of the daily returns (Risk).

7. Value at Risk:

In this I have defined a value at risk parameter for my stocks. I can treat value at risk as the amount of money we could expect to lose for a given confidence interval. There are several methods we can use for estimating a value at risk. Let's see some of them in action.

A. Value at risk using the "bootstrap" method

It is based on calculating the empirical quantile of a stock from a histogram. First, a histogram is constructed. From the histogram an empirical quantile (say 0.05) is calculated. This empirical quantile when multiplied with investment gives the expected loss or gain. Histogram for Foot Locker stock has been shown next.
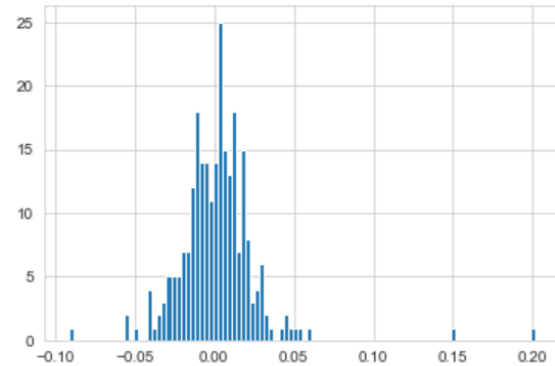

Fig. 8. Histogram of Foot Locker Stock

B. Value at Risk using the Monte Carlo method:

Using the Monte Carlo to run many trials with random market conditions, then I have calculated portfolio losses for each trial. After this, I have used the aggregation of all these simulations to establish how risky the stock is. I used the geometric Brownian motion (GBM), which is technically known as a Markov process.
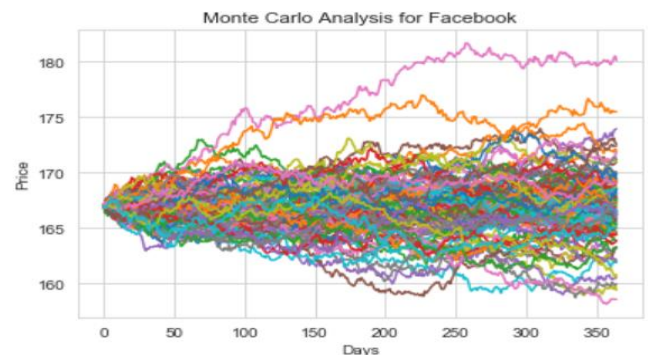

Fig. 8. Monte Carlo analysis for Facebook

This means that the stock price follows a random walk and is consistent with (at the very least) the weak form of the efficient market hypothesis (EMH): past price information is already incorporated, and the next price movement is "conditionally independent" of past price movements. This means that the past information on the price of a stock is independent of where the stock price will be in the future, basically meaning, you can't perfectly predict the future solely based on the previous price of a stock.

Now we see that the change in the stock price is the current stock price multiplied by two terms. The first term is known as "drift", which is the average daily return multiplied by the change of time. The second term is known as "shock", for each period the stock will "drift" and then experience a "shock" which will randomly push the stock price up or down. By simulating this series of steps of drift and shock thousands of times, we can begin to do a simulation of where we might expect the stock price to be.

C. LIMITATIONS

1. The analysis is not static.
2. The risk calculation and prediction changes with time.
3. Persons understanding stock market will only be able to analyze the graphs.
4. It involves statistics and probability which is not liked by many people, so it may turn out to be less interesting.

D. CONCLUSIONS

The paper presents extensive method of implementing Monte Carlo method for stock price risk analysis & prediction. The results found with Monte Carlo simulation helps to predict stock prices accurately on short term basis. This project implementation could guide investors to make profitable investment decisions.

E. REFERENCES

1. Y. Bengio. Practical recommendations for gradient-based training of deep architectures. In Neural Networks: Tricks of the Trade, pages 437478. Springer, 2012.
2. Y. Bengio and F. Gingras. Recurrent neural networks for missing or asynchronous data. Advances in neural information processing systems, pages 395401, 1996.
3. J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio. Theano: A cpu and gpu math compiler in python. In Proc. 9th Python in Science Conf, pages 17, 2010.
4. S. K. Mitra, Optimal Combination of Trading Rules Using Neural Networks, International Business Research, vol. 2, no. 1, pp. 86-99, 2009.
5. Daily Stock Market Forecast from Textual Web Data Wuthrich, B.; Cho, V.; Leung, S Permunetilleke, D.; Sankaran, K.; Zhang, J.; Lam, W. IEEE International Conference on Systems, Man, and Cybernetics, vol.3, p.2720-2725 vol.3, 11-14 Oct 1998
6. The Apache http client library is an open source Java library for working with HTTP. *http://hc.apache.org/httpcomponents-client/*
7. Neko HTML is an open source Java library for fixing HTML. *http://nekohtml.sourceforge.net/*
8. The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data Ronan Feldman and James Sanger Cambridge University Press, 11 Dec 2006
9. An Intelligent Forecasting System of Stock Price Using Neural Networks.
10. Baba, N.; Kozaki, M. International Joint Conference on Neural Networks, vol.1, pp.371-377, 1992.
11. Yahoo Finance https://finance.yahoo.com/