Stock Price Forecasting Analysis and Visualization

John Q. Li
York University
Toronto, Canada
johnwill4g@gmail.com

Rakesh Chauhan
York University
Toronto, Canada
rakesh1.chauhan@gmail.com

ABSTRACT

One of the most interesting (or perhaps most profitable) time series to predict are, arguably, stock prices. Accurately predicting how the stock market will perform is one of the most difficult things to do. There are so many factors involved in the prediction – physical factors vs. physiological, rational and irrational behavior, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy.

This work is focused on visualization side using Machine learning techniques to analyses data. Because data need to be cleaned as well as need to make in appropriate quality to visualize effectively. For a layman to make decision on complex stock data based on market sentiment is much easier with visualization. The stock price data at any moment are historical data of a publicly listed company. We will implement a mix of machine learning algorithms to predict the future stock price of a choosing 3 stocks as google, bitcoin and gas price future. The core idea behind this article is to showcase how the data value been clearly systematic mapped by using data D3JS visualization and Python graphics to ensure data quality. The Tools: d3.js + Python libraries and Statistical methods.

AUTHOR KEYWORDS

Data Analysis, Predictive Modeling comparison, D3JS, Interactive Visualization, Data Scientist, , Statistical Modeling, and Fundamental Analysis, Correlation, Standard Error, Risk comparison, Machine Learning.

INTRODUCTION

Visualization of the stock price data, which continually changes every moment and predicting price, is a challenging and exciting endeavour. Providing visualization can aid in its effective utilization in terms of interpretation and knowledge discovery for decision making. In this article, we attempt to apply Visualisation to qualitative as well as quantitative stock price forecasting data. Our motivation is to provide an easy view to understand and give the tool to make an informed decision. We are studying *Information Science and Technology* with an interest in data science. We attempt to combine machine learning and statistical modelling techniques to produce stock price forecast visualization. This paper uncovered lots of unknown insight, including the stock risk/return position and the potential value at risk and some other useful stock references, which may be key indicators.

All those founded detailed insights based on data analysis with visualization will vitally guide and assist investors with their easy decisions to minimize their decision risks and cost.

EXISTING WORK

The data visualization concepts are designed to understand a large amount of hierarchical data, including financial.
The visualization ability to overview of large
amount of data provides the ability to learn new dynamics and easy decision-making on data set, for example, by Applying Data Visualization Techniques for Stock Relationship Analysis 2019 Jie Huaa, Mao Lin Huangb, Guohua Wangc, Mouataz Zreikad [1].
There has been a lot of interests and article published in the field of Visualization and Data Science. There is enormous interest in the field of Data Science including visualization community
http://ieeevis.org/year/2019/info/papers-sessions

In this study, two aspects emphasized on:
• Discovering potential relationships among Apple, Microsoft, Google and Amazon stocks.
• exploring the correctness as well as the forecast of price and visualized relationship representation.

PROPOSED APPROACH

The approach we used in this study:
1) Stock raw data collecting from the Yahoo Finance
2) Applied Filter to reduce dataset and Data cleansing collected data
2) Data Analysis by applying statistical and machine learning techniques using Python library to make quality data ready for Visualization
3) Stocks price forecast visualization using D3JS

DATA COLLECTION

We claw the stock datasets from Yahoo finance web service. For our analysis purpose, we define the dataset time range from Jan 1st, 2018, to the current date. It would have around 500 records Which are good enough for a brief machine learning training and testing purposes. We can see all the attributes as the following table. The fact is that the current information within the dataset is far beyond our expectations; the daily changes, the averages etc. are not given to us.

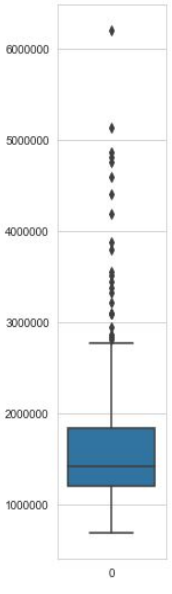
In the Stock market, following the essential intraday tips is a common practice before starting the trading day. However, the trading strategy changes with time and the concurrent events play a huge role in its working. The trading indicators are used to understand the market. Trading indicators are used to maximize returns.

- Uncertain risks and relation from trading:
- Unknown future direction and movement
- Significant potential risks due to volatility
- Lack of reliable reference or indicators
- The dependency between Volume and daily price changes?
- The most correlated stocks choice
- The most risk-sensitive stock choice for comparison.

These are the vital pointers shared by trading indicators. These essential but beneficial tips help in assessing the market conditions and allow traders to make better decisions concerning trade positions.

We can see from the original dataset tables.

DATA ANALYSIS



volume Daily Change

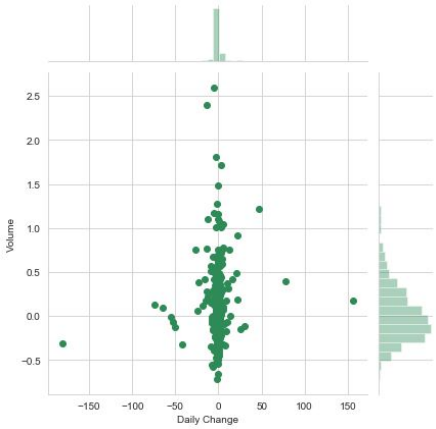
We know the fact that large volume stock trading in a short period would cause a rapid impact on its stock price. Would there any dependencies among the volume and the daily price change?

The first step in our analysis was cleansing the dataset, volume identification. The Daily Changes value calculated

from the difference between Open and Close value. The rest of the characteristics have a very similar distribution and ranges. Would there be any dependencies among the volume and the daily price change? To test the dependence, the data need to be Standardized between 0 to 1, as shown above. We have the following hypotheses:

H0: the two samples are independent.

H1: there is a dependency between the samples.



By using the Chi-Squared Test with a 95% confidence interval, we concluded that there is no sufficient evidence to prove there is a dependency between samples. And the paired chart of volume and daily change doesn't show any correlation between them.

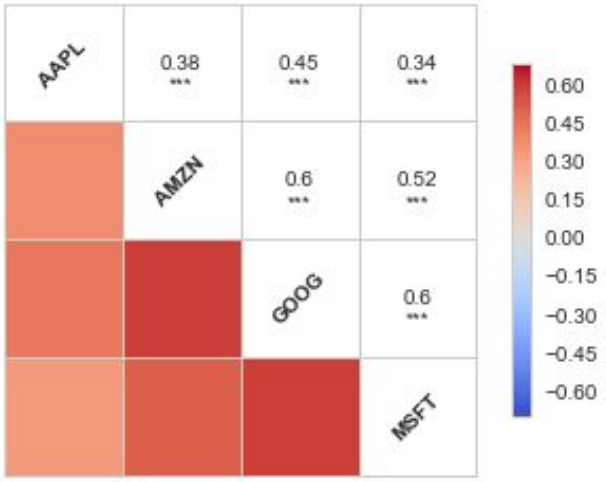
FUNDAMENTAL ANALYSIS

The stocks in the same domain class would possibly have any correlation between them. If so Which stock among them would be more sensitive to the risk, which stock is less painful? If we saw the stock price changes, It would offer some reference about what could happen next for the rest of the stocks belong to the same domain that the user was interested in. Finding the stock change indicators is important. The Most sensitive and correlated stock data could be a risk indicator. This paper selects Apple, Microsoft, Google and Amazon as an illustration to finding the indicator stock in the leading high tech domain. The most correlated stocks choice. The most risk sensitive stock choice for comparison. Our next step was to identify the correlations among stocks.

This paper selects Apple, Microsoft, Google and Amazon as an illustration to finding the indicator stock in the leading high tech domain.

The data need percentage change standardized. From the pair grid charts below, we can visualize a specific correlation that existed among all those stocks.

The generated heat map chart with coefficient r valued, we quickly suggest that the top 2 correlated pairs of stocks are Google and Amazon, Google and Microsoft. And between Microsoft and Amazon, a strong bond also exists.



	Coefficient, r	
Strength of Association	Positive	Negative
Small	.1 to .3	-0.1 to -0.3
Medium	.3 to .5	-0.3 to -0.5
Large	.5 to 1.0	-0.5 to -1.0

It is essential to note the risk involved with the expected return for each stock visually before starting the analysis. The standard deviation and 95% confidence level have calculated the returns and risk scarlet chart here. We can find that Amazon stock price is risky, and Microsoft has the best return potential so far.

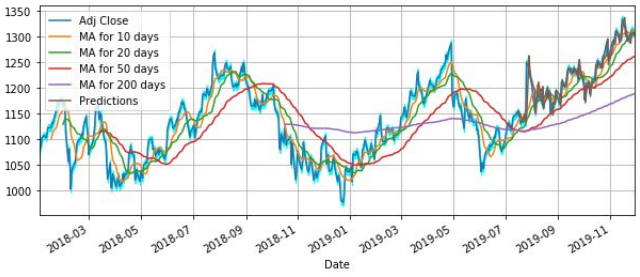
Summarize the heatmap and above risk/return chart. The Amazon stock can be used as an indicator to buy Google or Microsoft stocks. Since it's prices are more sensitive to the risks.

Technique forecasting analysis:

- Linear Regression
- k-Nearest Neighbors
- Auto ARIMA
- Prophet
- Long Short Term Memory (LSTM)

Identify the best prediction model from the list for our stock. We are using the visualization to justify the best solution.

Meanwhile, the 10 20 50 200 day Moving Averages values calculated. After applying the chosen LSTM model, our stock chart looks like the following.



It was interesting, to pointed out the peaks and throughs within out charts as following.

FORECAST INVESTMENT RISKS

We can predict future potential prices now. People need to know the chance of the value at risk for a particular range of future.

We used Monte Carlo to run many trials with random market conditions. We calculated portfolio losses for each experiment. We used the aggregation of all these simulations to establish stock risk.

We used the histogram to demo all the predicted values from the Monte Carlo analysis result at 365 days from now. And using Z test to define the value at risk with a 99% confidence interval.

For every initial stock purchased, there is about $21.16 at risk 1% of the time, as predicted by Monte Carlo Simulation.

D3JS SOLUTION

The stock price is time-series data. If a reader going to gamble, whether Google will gain or lose tomorrow, what are the chances? The user may make decision-based on this analysis, for example. What is prediction accuracy for this stock at this moment? The difference among the Moving averages, stock price and predicted price? and so on.

DATA VISUALIZATION

To provide visualized results for analysis purposes, we adopted two types of a graph in this experiment based on data analysis fundamental analysis, technical analysis and risk analysis using Python. We visualized data on D3JS platform, and we offered the future forecast stock price within the time series with the history and various moving averages lines. We also provided the trends prediction accuracy level to help the user to make their selections. Through our D3 platform, we can zoom in the specific period to figure out the value differences and observe the price volatility within this short period.

Meanwhile, we can also compare the short time price change with long-time stock price lines. The user may learn a lot of new angles by applying their perspective with provided stock price forecast Visualization. We believe that with all this information and references our interactive data solution has offered, the user will be able to make a better investment decision judgement.

CONCLUSION

Historical data is a reliable resource. Data may become more potent if you analyze them thoroughly. This paper demonstrates how robust the data analysis could be — especially combined with machine learning. Through our project, starting from the dataset, we manipulate the data according to our expectations. We generate new attributes; using to prove the correlation and dependence with statistics hypothesis; identify the risk/profit extreme values with charts; forecast the future status with machine learning. Step by step, we are discovering more useful information, minimize the risks, informing the conclusion and using data visualization to make easy understanding, learn new and in-depth about data and smooth decision-making.

ACKNOWLEDGMENTS

We thank York University professor Enamul Hoque Prince, as well as online community specifically Visualization, and data science publications for excellent knowledge, who wrote and provided helpful learning information. Particularly D3JS.

REFERENCES

1. Applying Data Visualization Techniques for Stock Relationship Analysis Jie Huaa, Mao Lin Huangb, Guohua Wangc, Mouataz Zreikad [online] Available at: https://www.pmf.ni.ac.rs/filomat-content/2018/32-5/32-5-45-6898.pdf [Accessed 7 Dec. 2019]

2. Bostock, M. (2019). *D3.js - Data-Driven Documents*. [online] D3js.org. Available at: https://d3js.org/ [Accessed 6 Dec. 2019].

3. Learning, D. and codes), S. (2019). *Predicting the Stock Market Using Machine Learning and Deep Learning*. [online] Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd-deep-learning-techniques-python/ [Accessed 6 Dec. 2019].

4. Medium. (2019). *Using the latest advancements in deep learning to predict stock price movements*. [online] Available at: https://towardsdatascience.com/aifortrading-2edd6fac689d [Accessed 6 Dec. 2019].

5. Brownlee, J. (2019). *Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras*. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/ [Accessed 6 Dec. 2019].

6. Medium. (2019). *Stock Analysis in Python*. [online] Available at: https://towardsdatascience.com/stock-analysis-in-python-a0054e2c1a4c [Accessed 6 Dec. 2019].

7. Medium. (2019). *Stock Prediction in Python*. [online] Available at: https://towardsdatascience.com/stock-prediction-in-python-b66555171a2 [Accessed 6 Dec. 2019].