Junhe Zhang, DATS 6401, Summer 2020

Final Technical Report

## Reveal and visualize the important Factors that affect Stock Price Trend

**Abstract**

By the end of 2019, there were more than 625,000 users who have been registered a stock account in the U.S stock market and began to trade in stock. The stock market is a place where full of opportunities and risks. Some people are rapidly growing their wealth by trading stocks, others have lost a lot of money in the stock market. The success of Warren Buffett proved that the stock market gives an intuitive feedback on one's vision and knowledge of stocks. In other word, how to gain profits from stock trading can be studied and analyzed. The purpose of this project is to use visualize and machine learning tools to study company’s complex financial data and provide end-user a better perspective of those financial data in the future stock investment.

**Objectives**

The primary objective of this project to visualize the financial data that extracted from big tech-companies’ finance reports which are listed on the U.S. stock exchange from 2010 to 2019. Because of a company's financial information can most directly reflect the operation of a company, these data are also tightly bound with their stock price. It would be very useful to build a machine learning data model that reveal the effects to their stock price of each financial factor and the hidden coefficient. Finally, visualize the top 10 important factors’ coefficients of the data model and give end-user a good inspiration in their future stock investment. The financial data from this project will come from Yahoo finance website.

**Functional Requirements**

This project must be able to construct a reliable data model, which has its accuracy to be at least 60 percent, from all the financial data of Apple Inc from 1989 to 2020. The financial data and stock price data need to be organized in a way that visualization can be efficiently create and end-user will be able to use the visualized data as assistant to draw their conclusions. The end-user will be able to interface with a web page where they could do exploratory data analysis through the interaction functionality that provided with each graph on both the company’s financial data and stock price data and study their relationship. End-user will be able to navigate through each webpage section and go to the raw data using various links. The interactions include the ability to selection different financial terms and display detailed data on mouse hover over the graphs.

**System Architecture and Description**

The raw data was downloaded in the form of CSV files from YAHOO finance website for both Apple Inc’s financial data and its stock price data. Since there are 5 percent missing value in datetime column of stock price data, to fill the missing datetime data in stock price dataset they were generated using the datetime library in python and then use the forward fill function in pandas to complete the missing data point with the previous date’s stock price. Both data sets were then combined using the datetime as the primary key into a consolidated raw data set that could be preprocessed. The entire data preprocessing process were done using Pandas library, which including table join, handle missing values, data scaling and so on. The cleaned data were then been used to build the data model.

**Development Platforms**

The back end of this project was build using python Flask library, front end was a webpage that build with HTML, CSS, and Java Script. The HTML was used a webpage structure which organize the contents of the webpage. The webpage was then beautified with Cascading Style Sheets (CSS) and Bootstrap version 4.0. Visualizations and graph interactions will be created using D3.js and Google visualization API, which will be discussed more specific in the following section.

**Proposed Visualizations**

The D3 Java Script library and Google visualization API will be use in this project. The webpage starts with a scatter plot which will be used to show the correlation between Apple Inc’s financial data that including the top 10 important factors among over 50 financial data columns that were used to build our data model and Apple Inc’s stock price data. So that end-user could explore those data as much as they want to assist their study on stock market. The second section will be a line chart which will be used to show the trend of stock price and the top 10 financial factors over 30 years from 1990 to 2020. This will help end-user to discover both trends and magnitude of all eleven data columns, the top 10 important financial data and stock price. The third section will be another scatter plot that display the actual stock price compared with the predicted stock price that using the trained data model. Since the data model was trained using the first 20 years data points, the prediction will be done on the latest 10 years data point which is from 2010 to 2020, then plot both actual and predicted stock price on the same coordinate system with quarters as x-axis and stock price as y-axis. This will show end-user how accurate of our data model is, so that they could trust our coefficient analysis in the next section. In the coefficient analysis section, there will be a bar chart to show the top 10 factors’ coefficients in our data model. These coefficient values could be acquired using the expand Scikit-learn library in python.

**Experimental Analyses and Conclusions**

**A screenshot of a cell phone

Description automatically generated**The data first revealed that both Apple Inc’s stock price and its financial data were very low in magnitude. The growing rate of each data columns was nearly nothing before 2005, soon after 2005 Apple’s stock price were soaring as well as its financial data. This is all because of the great profits that made by iPhone, which is a typical case of a rising stock in a profitable company. The trained data model has 90 percent accuracy which is very reliable. From its coefficient values some financial terms like Gross Profit has great positive effects on stock price, other terms like Operating Expense has negative effects. When the positive terms have greater magnitude than negative terms, the stock price would increase in that year and vice versa. This indicates that in long term stock investment, investors should pay more attention to a company’s financial report statues that have high influence (high coefficient in our study) on stock price and use them as important factors on their investment.