Predict Stock Market (PSM)

Introduction

A Stock market, equity market or share market is the aggregation of buyers and sellers of stocks, which represents ownership claims of business, Stock market and share price changes based on economy, international reputation, war, and so on. Therefore, Investors are willing to know the future and predict stock market

In this project, I'm going to analyze data related to 30 famous stock market and predict their price and up and down based on price history using analytical model called time series. In addition I will visualize my data in different ways to help portfolio managers convince investor in the better way.

Literature Review

Stock Market prediction has always had a certain appeal for researchers. While numerous scientific attempts have been made, no method has been discovered to accurately predict stock price movement. The difficulty of prediction lies in the complexities of modeling market dynamics. Even with a lack of consistent prediction methods, there have been some mild successes. Stock Market research encapsulates two elemental trading philosophies; Fundamental and Technical approaches. In Fundamental analysis, Stock Market price movements are believed to derive from a security's relative data. Fundamentalists use numeric information such as earnings, ratios, and management effectiveness to determine future forecasts. In Technical analysis, it is believed that market timing is key. Technicians utilize charts and modeling techniques to identify trends in price and volume. These later individuals rely on historical data in order to predict future outcomes. Most existing literature on financial text mining relies on identifying a predefined set of keywords and machine learning techniques. These methods typically assign weights to keywords in proportion to the movement of a share price. These types of analyses have shown a definite, but weak ability to forecast the direction of share prices.

Dataset

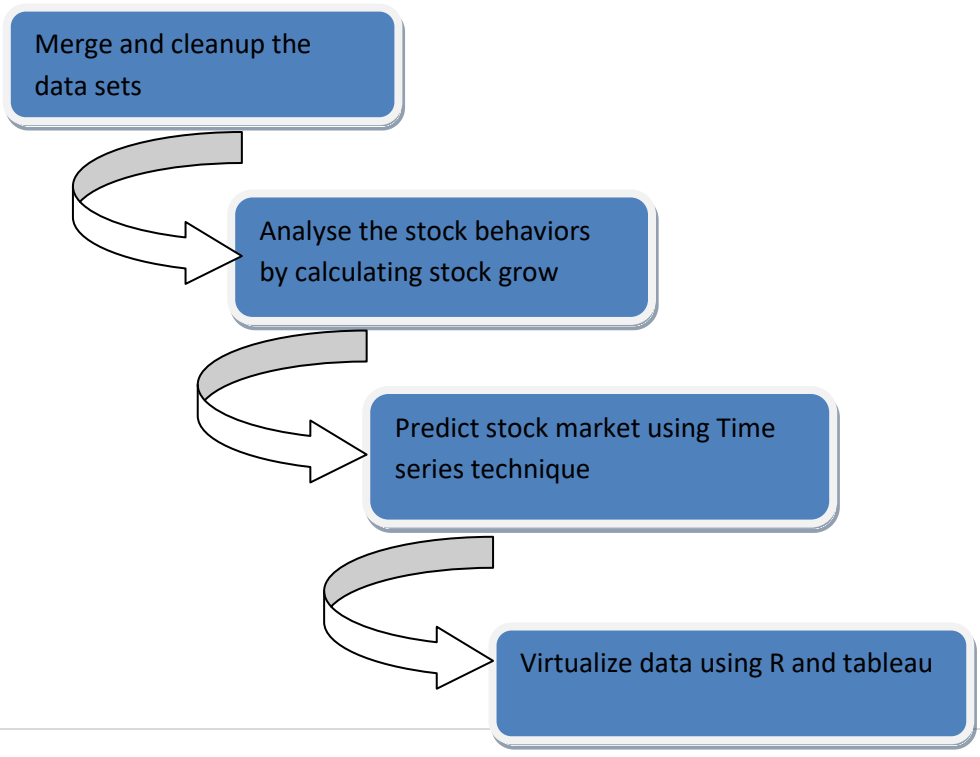
In this project I'm going to use stock price history from Yahoo finance and S&P 500. I would use around 500 stocks which have at least 30 stock traded for at least 10 years, so I'm sure that 10 years of price history is available. My data set contains Date, symbol, open price, low price, high price, close price, and volume and contains around 1,200,000 rows in CSV format

https://quantquote.com/historical-stock-data

http://www.nasdaq.com/symbol

- Date: Represents the date of the market price, format is DATE, YYYYMMDD.
- Symbol: each company or stock in the market is known with symbol. This is a unique value that identifies equity in the market, its CHAR variable that could be a combination of alphabet and number.
- Open price: this is a numeric entity that shows price of each unit of stock when market opens, this price is usually is closed to previous day closing price.
- Low price: this item shows lowest price of each unit in a day, this items changes frequency while market is open, however my data set record the lowest one. This is a numeric attribute with 3 decimal.
- High price: this item is opposite of low price, hence shows highest unit price in one day. This is also numeric attribute with 3 decimal.
- Close price: this item records the latest share price during a day and do not use in financial calculation such as rate of returns (ROR). Same as above, this is a numeric with 3 decimal.
- Volume: This attribute shows the number of shares in total that has been traded within a day; Volume is an important element for investor to analyze the stock market.

Approach



Step 1: <Merge and cleanup data set>

I've downloaded around 500 individual data file (in CSV format) for different stock as of 10 years ago, I wrote a script in Perl to read each file and write into one consolidated file. In addition, I used the previous price for missing price value, if price is missed on 1st day I evaluated as zero.

Step 2: <Analyze the stock behavior by calculating stock growth >

In order to analyze stock reaction to market, I had to calculate several stock market finance components such as Stock growth, Rate or Return, and graph them. I used R for this task. I'll share code in Github as well.

Step 3: <Predict stock market using time series>

I used time series method to analyze the stock market changes and predict the future. In fact, I used previous day price to predict stock market price for following day. I have used R to train my sample data

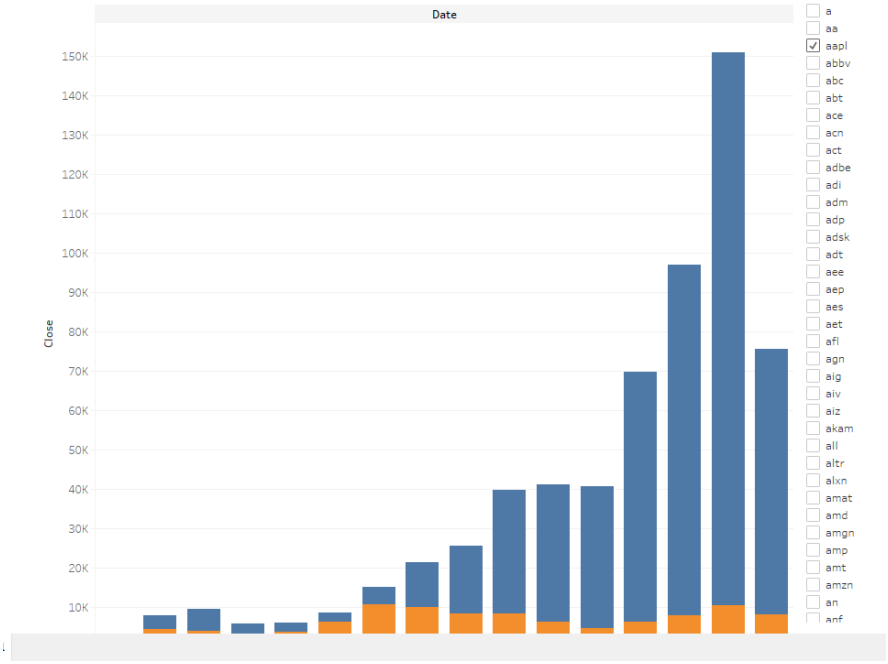
Step 4: <Virtualize data using R and Tableau >

Beside R, for graphs and visualization of data, I'm going to export data into Tableau and create dashboard for my advisor and let me to review the market price and compare several stock market with each other.

Step 5: <Conclusion and result >

Result would be clean and processed data in target files, several models and charts in R that shows stock market evaluation and prediction. Visual chart in Tableau that shows :

Sample:



Reference:

1. https://en.wikipedia.org/wiki/Stock_market

2. R. Bisoi, P.K. Dash, “Prediction of financial time series and its volatility using a hybrid dynamic neural network trained by sliding mode algorithm and differential evolution”, Int. J. on Information and Decision Science, Inderscience Publications, U.K., Vol. 7, no. 2, pp.166-191, 2015.

3. Ortega, Edgar; Yalman, Onaran (December 4, 2006). "UBS, Goldman Threaten NYSE, Nasdaq With Rival Stock Markets". Bloomberg.com. Retrieved 2011-05-31.

4. Hong Keel Sul,Alan R. Dennis,Lingyao (Ivy) “Trading on Twitter: The Financial Information Content of Emotion in Social Media”, 2014 47th Hawaii International Conference on System Science

Saeid