**Stock Price Trend Forecasting**

**Background:**

Stock price prediction is a challenging problem that widely studied by researchers from various fields. The biggest challenge of studying the stock market is the volatility in prices caused by their sensitivity to financial and economical noises. This volatility makes it difficult to apply the traditional statistical techniques including time series and regression models. Recently, machine learning algorithms like support vector machine, random forest, reinforcement learning, deep learning or decision tree learning have become popular among researchers and financial institutions to explain the movement of stock market. This problem is challenging though appealing to many researchers and traders because even a slight improvement could make a significant increase in profit for the stock holder.

**Goals:**

In this project, I want to study and apply machine learning techniques to predict the stock price pattern of a single stock. My attempt is to predict whether the stock price will increase or decrease sometime in the future compared to its value on a given day. In particular, I plan to discuss tree based models like random forest, boosting along with logistic regression in detail to predict the movement of a single stock. While predicting the movement, I also explore various variables that might have an effect on stock market or that might be used as an indicator of stock price movement.

**Data Source and Data Wrangling:**

To apply different models and to compare their efficiencies, only one stock, AT&T (T) is chosen from 1/1/2010 to 4/25/2017. The first data set contains daily stock information (Open, Low, High, Close, Volume, Adjusted Close, Dividend Date). It is extracted from Yahoo!Finance using pandas\_datareader library. The dividends information is also extracted from Yahoo!Finance and the earning data is extracted from busystock.com.

The following features are some candidates to explore the movement of stock prices and predict the future prices:

* n-day price change eg. n in {1, 2, 5, 10, 20, 270}[[1]](#footnote-0). It can be net price change or percent change (variable abbreviation: PCh)
* Kaufman’s efficiency ratio [link](http://etfhq.com/blog/2011/02/07/kaufmans-efficiency-ratio/), [link2](https://www.marketvolume.com/technicalanalysis/efficiencyratio.asp), [link3](http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:kaufman_s_adaptive_moving_average) (variable abbreviation: Effr)
* n-day moving average (variable abbreviation: MA)
* n-day momentum (rate-of-change) (variable abbreviation: ROC)
* Stock moving direction i.e. 1 if closing price that day is higher than the day before, and −1 if the price is lower than the day before (variable abbreviation: PrDir)
* Calendar date eg. 2 for earning record date, 1 for ex-dividend date, 0 for regular date (variable abbreviation: Action)
* Day of the week (variable abbreviation: Day)
* n-day volume difference (variable abbreviation: Volumelagged).

The initial features selection for modeling part is based on the literature and expert recommendations.

The main data is loaded into pandas dataframes via pandas\_readerlibrary. Thus, it is ready to use. There are no missing values in the data set except the holidays on which the stock market was closed. Thus I start immediately preparing the data set for modeling part. The first thing before modeling is to do an exploratory data analysis. The next section uses visualisation and transformation to explore the data in a systematic way.

**Exploratory Data Analysis:**

I start exploring data with a time series visualization of closing price. It is clearly evident from below graph that there is an overall increasing trend in data along with some variations. The AT&T stock price shows overall increases about 30% from 2010 through 2017. The steepest increase in prices happens the first half of the 2012 and 2016. On the other hand, from the second half through 2016, the prices shows steady trend despite some ups and downs.

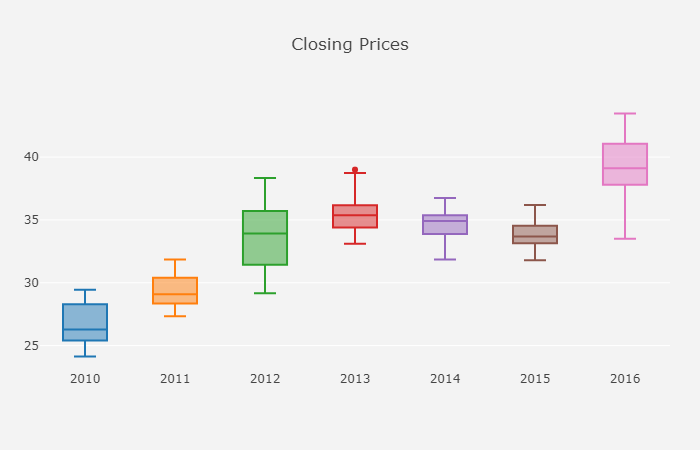


The following two plot shows the monthly and yearly rolling average of prices. Both plot support our previous observations.

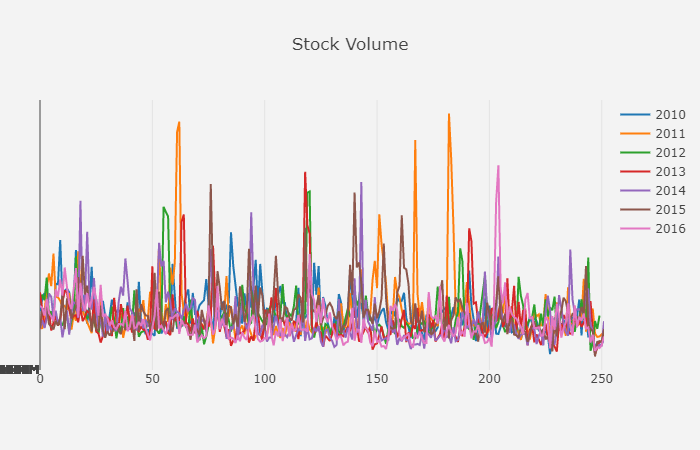




Next, box plots of closing prices for each year is shown. This plot provides a bigger picture how closing prices behave with respect to each year. Except 2013 through 2015, A&T stock prices show an increasing trend with a highest performance in 2016 and lowest performance in 2015. Another observation seen in the plot is the dispersion of stock prices. While the stock prices show small variability and close to each other in most of the time, the dispersion is big in 2012 followed by 2016.



The last plot shows the volume of AT&T stocks for each year. The volume of stock looks stationary overall although it has some strikes in each year.



For more descriptive statistics, please refer ATT-ExploratoryDataAnalysis.ipynb jupyter notebook.

1. n in {5, 10, 20, 270} represent one week, two weeks, one month and one year. [↑](#footnote-ref-0)