Predicting Apple Stock Market Buy Action

# **Gloria Moore Summer 2021** [github.com/gloriaypradal](https://github.com/gloriaypradal)

Name of the repository for this project is: buyapple

# Which Domain?

* [Yahoo Finance - Stock Market Live, Quotes, Business & Finance News](https://finance.yahoo.com/).

The web scraping process of the yahoo finance has not been so simple, there are tables that have different codes, and the time factor (for future predictions) is something that I have not figure out yet, however, there is a good resource on kaggle about dealing with data with time series and API’s

* <https://www.kaggle.com/> using this data set and read it on my notebook was complicated at the first attempt because my computer takes too much time to read all the rows on the file, there are 138 columns on the train file (this one is going to be used to extract the coded features) and It has 5.77GB, so it has been a struggle as well to read it. This file will be used just with a restricted number of observations because of its lenght
* [API Marketplace - Free Public & Open Rest APIs | RapidAPI](https://rapidapi.com/marketplace) I have not started to work with this API yet
* [Inflation Calculator | Find US Dollar's Value from 1913-2021 (usinflationcalculator.com)](https://www.usinflationcalculator.com/): This website has been the easiest to scrape so far, the table is short and simple.
* The data set about oil prices was bought on Macro trends and it’s a .csv file, in the moment, I am cleaning the data set and designing what is going to be the time frame to use. It is a simple file just 2 columns with date and prices

# Which Data?

The data available on yahoo finance it is being collected using BeautifulSoup, but because of all the information on this website, this task is taking long time, the complication on this particular is the time series involved in all this process, however, for this project I will used historical data, but the complication will be for the prediction

[Home | MacroTrends (dpdcart.com)](https://macrotrends.dpdcart.com/) : Price Oil daily from 1986, will be used as a feature, this data set was bought by me on the website. .csv file with only 2 column to analyize, not missing values, the cleaning and processing of the data hasn’t been started.

Cleaning and processing the “dates”columns on each file is the task that its taking too long

For the Consumer Price Index data, the website was scrapped and data is on a data frame, there are non- missing values

A picture containing text, screenshot, indoor, computer

Description automatically generated

For daily temperature data, I haven’t started to work with this APi yet

# Research Questions? Benefits? Why analyzes these data?

Question: May I know when to buy or not in a future an Apple Stock, based prices of closest competitors, oil prices, gold process, temperature of the day and Disney stock price?

In this section the complication will be the prediction to the future, I am not very knowlagable on time series management, but there are a lot of helpful resources on Kaggle and other websites:

<https://www.kaggle.com/italosimoes/jane-street-market-predictions>

<https://www.kdnuggets.com/2020/01/stock-market-forecasting-time-series-analysis.html>

This section hasn’t been touched yet, I am still working on collecting and processing the data, this tis the most detailed task so far

# What Method?

Most of the data will be collected using web scrapping, with BeautifulSoup. This library is very easy to understand, and I think that there is a lot of resources on the internet that has been helping me on collecting the data specially on yahoo finance.

My approach with time series is going to be the hardest part, as I mentioned, I am not very knowledgeable on this matter, but I am already reading and reading some examples about this.

My approach in the biggening is to use dates as observation, and predict as a target to buy or not but depending on the maximum price that would be previously set to buy, in this case, if I am predicting a value (US$) I might have to test 2 types of models, predicting with a linear regression the value of the stock and then setting a rule to decide to buy or not, or transform the prices over the maximum and under or equals the maximum as a 0 or 1 variable and use a classification model to predict to buy or not, but this last approach won’t let me know the price that I actually buying at. So, this is in analysis.

# Potential Issues?

* Time series knowledge
* Size of the files (Janet Street Market Data set)
* Time, time is always a risk factor because of all the data processing time. Level of details necessary to make this project applicable.

# Concluding Remarks

I am not an expert on stock market, my husband is giving me the theory of how it works, and I am dealing with the data, I think that predicting the value of a stock is something that has been done a lot, but having or setting the best factors and model for an specific stock trading will create the strategy for those who are interested on this type of transactions. With this project I intent to create the best model possible to predict an action to buy or not a stock, however the stock market is affected for so many factors that are out of my scope and can result on my model not being the most realistic one. An interesting extension of this project can be a sentiment analysis on twitter or other social media about recent political news and relate this with the value of the stocks

References

[1] <https://www.investopedia.com/articles/stocks/12/history-apple-stock-increases.asp>

<https://www.usatoday.com/story/money/markets/2012/10/09/apple-stock-must-own/1609157/>