Predicting the directionality of S&P500

Capstone Proposal

Yemitan Isaiah Olurotimi

October 13th, 2019

Proposal

Domain Background

S&P 500 is a stock market index that measures the stock performance of 500 large companies listed on stock exchanges in the united states. The index is widely considered as the best indicator of the day to day performance of the U.S stocks.

Over 50 years for data was analyzed, It was deduced that there is approximately 52% chance that the stock price will increase and 48% that it will decrease for any given day.

This project aims to use some machine learning techniques to predict the directionality of stock prices.

Problem Statement

Stock prices are usually described as a statistical process called “random walk” which means each days closing value is unpredictable and random.
The problem to be solved with this project is to apply machine learning to predict the directionality of a stock price for any given day using over 50 years of historical data.

Datasets and Inputs

The dataset for S&P 500 is publicly available and will be obtained from yahoo finance.
The time period for the data set is between October 1969 to October 2019 which is 50 years.
This is data contains 12,614 rows and 7 columns which are;

- Date: Date for the given data.
- Open: The stock opening price for the given date.
- High: Highest stock price for the given date.
- Low: Highest stock price for the given date.
- Close: The stock closing price for the given date.
- Adj Close: Adjusted closing price for the given date.
- Volume: Volume of stock traded for the given date.

Link: https://finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC

Solution Statement

The task for this project is to predict the directionality of stock prices of the next N day(s) in the future using historical data as its input.
Multiple ML algorithms will be used to predict the closing stock price. The algorithms that will be used are Long short term memory (LSTM), Autoregressive integrated moving averages (Arima) and Prophet. These models will be used individually and compared to see which performs best.

Benchmark Model

The benchmark model will include multiple sources such as trading economics, forecast chart, and the use of random forest since stock prices are perceived as a random walk.

- Trading Economics https://tradingeconomics.com/spx:ind/forecast
- Forecast Chart https://www.forecast-chart.com/index-sp-500.html
- Random forest
- Financial forecast centre https://www.forecasts.org/data/data/SP500.htm

Evaluation Metrics

In this project, the model will be evaluated using Mean squared error and mean absolute percentage error.

Mean squared error measured the average of the squared of the errors (Average of the difference between the estimated values and actual value)

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

Mean absolute percentage error is a measure of prediction accuracy of a forecasting method. It is used for regression problems.

$$\mathrm{M} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|,$$

Project Design

The project is intended to be built with a frontend (react) and Backend (Flask). The backend will be responsible for the prediction model and API.
For the time being, the project will be implemented with Jupyter notebook.
Outline of the workflow;

- Collecting Data
- Cleaning and pre-processing data
- Splitting data to Training, validation and testing set.
- Use the various models, architecture and tune the parameters
- Train model
- Test model
- Compare model results with the benchmark.

References

- https://en.wikipedia.org/wiki/S%26P_500_Index
- https://en.wikipedia.org/wiki/Stock_market_prediction
- https://www.fool.com/knowledge-center/what-is-the-sp-500.aspx
- https://finance.yahoo.com/quote/%5EGSPC/
- https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd-deep-learning-techniques-python/

- https://towardsdatascience.com/machine-learning-techniques-applied-to-stock-price-prediction-6c1994da8001
- https://machinelearningmastery.com/time-series-forecasting-performance-measures-with-python/
- https://otexts.com/fpp2/accuracy.html
- https://en.wikipedia.org/wiki/Mean_squared_error