**S&P 500 Stocks analysis**

Nada Aljumah

**Abstract**

The goal of this project was to use time series models on a group of stocks historical data to predict the future price and therefore its investment worthiness.

**Design**

The data is provided by S&P 500 and contains several features to be examined but since only timeseries models are used, everything except closing prices has been ruled out.

Analyzing stocks via machine learning models would improve individuals’ financial decisions and ensure educated actions has been taken. This is by getting the best parameter for each model before taking outputs into consideration.

**Data**

Data has been extracted from S&P 500 index: it’s a daily record of 500 companies for 5 years.

Since the data is time series, closing prices is the only feature to deal with. Yahoo finance package used to extract tickers for the S&P stocks.

**Algorithms**

*Feature Engineering*

1. dropna()**.**values to rule out any NA value from closing prices data.
2. Prepare data for RRN by converting test and train data into the matrix with step value that contains n number of elements.

*Models*

Recurrent Neural Network to map historical prices sequence to a predicted sequence.

An autoregressive integrated moving average to predict future trends based on past values.

Stepwise to get better estimation when selecting ARIMA model order.

*Model Evaluation and Selection*

For each stock, an ARIMA and a RRN is applied.

The metric used for evaluation is Mean Squared Error (MSE); the average of the squared forecast error values.

**Tools**

* Numpy and Pandas for data manipulation
* Scikit-learn for modeling
* matplotlib.pyplot for plotting
* yfinance for stocks tickers

**Communication**

In addition to the slides, my project has been also uploaded on my repo.