Module Code: CS3AM

Assignment report Title: Stock price prediction

Date (when the work completed): fill this out

Actual hrs spent for the assignment: 30

Name: Lyndon Fernandes

Student number: 30000205

What have I done so far:

* Built a data collection pipeline in python using the yfinance module which is a module that gets data using yahoo API.
  + Gathered a list of 20 stocks in a weighted split of different sectors reflected in the SP500
  + Concatenated stocks from different sectors and created datasets based on sectors.
* Checked for Null values and duplicates as well as dtypes
  + For all stocks present checked for null values and duplicates.
    - Dataset was fairly clean and had no null values or duplicates
  + Checked Distribution of each stock for any outliers that stood out to me.
    - Need to finish but the dataset looks fairy clean again
  + Formatted dtypes properly so save storage and improve efficiency in data handling.
    - Change datetime UTC-> ‘America/ New York’ time zone
    - Make sure all datetime records are in ISO8601 format
    - Change all floats from float64 -> float32
    - Ints from int64 -> int32
    - And any name objects to string

What I need to do next:

* Visualisations
  + Look at the data closely generate time series plots
  + Look at correlation between different stocks see if any stocks behaved the same or similar
    - Check between different industries
* Descriptive statistics
  + See which stock changed it’s mind the most check for volatility, variance, std
* Feature generation:
  + Look at percentage change for adjusted close, will be a good feature for non-deep learning model
  + Rolling or moving averages? Maybe daily or weekly see what happens

Which models will I use?

Deep learning model: LSTM

Non-deep learning model: (Random Forest Regressor [might not be ideal]), ARIMA