Stock Prices Indicator Project Proposal

Overview:

Planning is a main point in any investment, so in order to put good plan that can success the current business .Investors want to know the expected future profit and loss to make decision depend on it, so data and machine learning play a big role in making this information available in a high accuracy level.

Any Decision Maker will take his decision depend on market forecasting for his product or service or any new item he wants to direct his investment to it, offer good stock prices indicator is a very critical issue since decision maker will give instruction to all department (production, development, marketing) depend on it.

Problem Statement:

The goal is to achieving an accurate stock price indicator using yahoo finance data for training and testing in order to determine the stability of financial market and the efficiency portfolio.

The potential algorithms that will be used is DeepAR since ARIMA model perform good only for linear data.

Solution Statement:

1.Data Loading Steps:

-Data loading from yahoo finance data using ranaroussi loader which download the data in csv format.

-Save Data in csv file

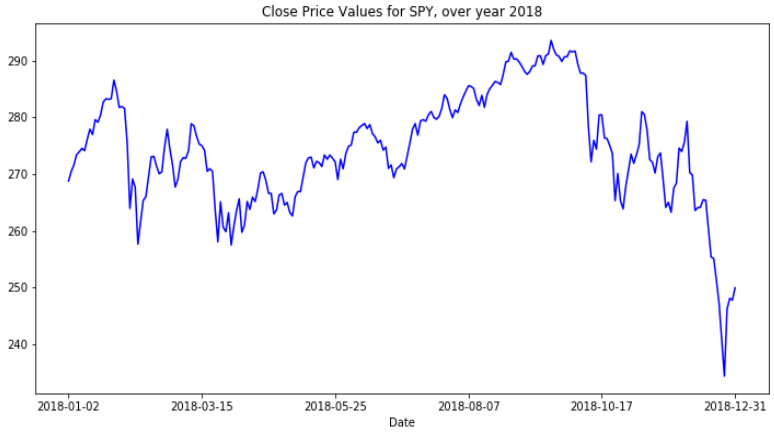
Sample of data saved

Date	Open	High	Low	Close	Adj Close	Volume
1/2/2018	267.84	268.81	267.4	268.77	254.1127	86655700
1/3/2018	268.96	270.64	268.96	270.47	255.72	90070400

1/4/2018	271.2	272.16	270.54	271.61	256.7978	80636400
1/5/2018	272.51	273.56	271.95	273.42	258.5092	83524000
1/8/2018	273.31	274.1	272.98	273.92	258.9819	57319200
1/9/2018	274.4	275.25	274.08	274.54	259.5681	57254000
1/10/2018	273.68	274.42	272.92	274.12	259.171	69574300
1/11/2018	274.75	276.12	274.56	276.12	261.0619	62361500

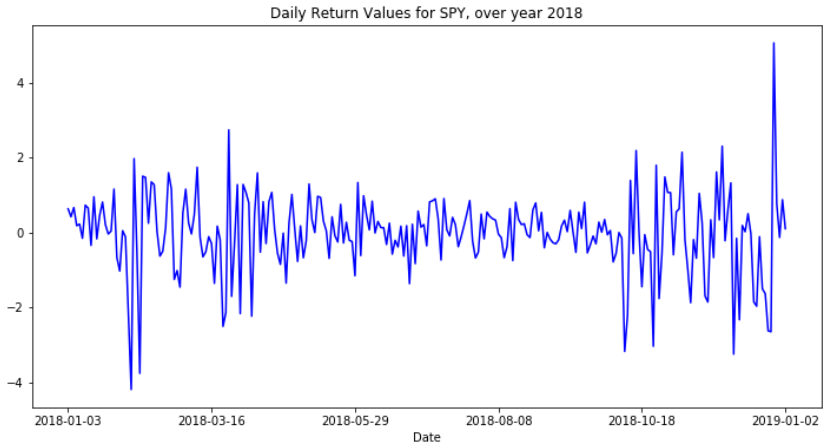
2.Data Exploration Analysis:

-convert data to series by making datetime is index.



- check for missing period and data then decide best way for fill it like by feature mean value.

-calculate daily return and save it to dataset since it gives better indication of market changes than actual price value in addition to its stationary nature.



-using correlation plots to can determine best value of lag.

-using plots to can know type of time series data according to distribution shape.

3.Alogorithm:

Implement sagemaker built in algorithm DeepAR since it built on RNN which gives very good result in case of stock indicator, so we transfer the data to json format since it is the acceptable format for DEEPAR.

-Define the model by using base sagemaker estimator and DeepAR image container.

-Tuning the hyperparameter in order to active best parameter for my data pattern.

-Train the model after determine the test data.

-Deploy the model to create endpoint

-Format the input data for test the send to predictor

-Decode output from indicator

-Plotting the data with quantile space of 80% to can explore result.

-Calculate RMSE for model evaluation.

-Repeat the steps after deploy on future data that not used in training to be sure the result is in the same accuracy level.

-After that retrain model on whole data and deploy it to be ready for future task.

Evaluation Metrics

The main evaluation matrix is RMSE and graphs that demonstrated the deviation between prediction and actual value since RMSE will not provide information about what prediction result for every point in data distribution.

Benchmark:

Main point in Benchmark is to achieve result best than the result can be obtained by SARIMAX model which not perform well with spiks in data and result is go to sharp increase and decrease far from actual data.

From my research for a good value for RMSE in case of stock price indicator I found it in the range between 1.0 and 1.2, so my goal is to achieve this range or less.