Project Group 21 03/16/2018

Solution Design and Data Collection:

The goal of the project as stated in the project statement is to come up with a model to predict the price of a certain stock based on various indexes and ETFs in which this stock is a component. We have selected the US Banking Services Industry as our target industry and we will be analyzing JP Morgan (NYSE: **JPM**) to perform our analysis. The goal is to predict based on the performance of various indices whether the stock price will go up or down during the current day.

We will be having 4 different datasets.

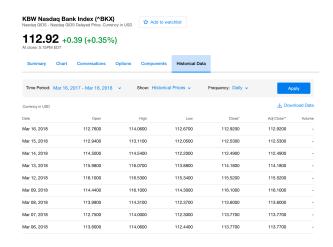
- 1- Target Stock: JP Morgan Stock Dataset, a daily stock dataset with variables such as High, Low, Close for a period of 1 year. Available on Yahoo Finance API for R in the R Package Quantmod.
- 2- Index 1: Dow Jones Bank Index (DJUSBK) is a popular benchmark index for top banking companies in the USA. JPM is a component of this index and has a weightage of 4.57%. The dataset for this index will also be based on 1-year time series data with High, Low and Close for each day. Data for this is available from various sources such as Quantmod or Investing.com.
- 3- Index 2: Financial Select Sector SPDR ETF (**XLF**) is a prominent exchange traded fund comprising of various companies in the financial services sector in the USA. JPM is a component of this ETF and carries a weightage of 11.49%. The dataset for this index will also be based on 1-year time series data with **High, Low and Close** for each day. Data for this is available from various sources such as Quantmod or Investing.com.
- 4- Index 3: KBW Bank Index (BKX) is a banking industry benchmark index comprising of 22 banking firms. JPM is a component of this index and has a volume of 19.88 MM. The dataset for this index will also be based on 1-year time series data with High, Low and Close for each day. Data for this is available from various sources such as Quantmod or Investing.com.

After obtaining the data from each individual dataset, we need to perform a few adjustment operations to get price-based weighting as each component stock of the index has a different price. Hence, increase coherence, we multiply volume with price of the stock on that day to generate actual percentage weight value of the stock for that day.

We then plan to use ARIMA (Auto-Regressive Integrated Moving Average) to forecast our stock price by training the ARIMA models on our benchmark indices. We will be optimizing our algorithm along the way and will be experimenting with various ML models like neural networks as well. For neural networks we will be using TensorFlow for R.

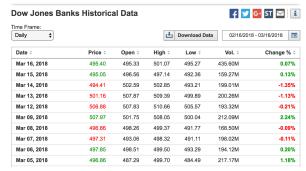
Project Group 21 03/16/2018

Snapshots of Datasets.



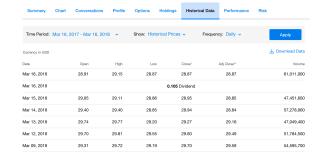
KBW Bank Index (BKX)





DJ Banks Index (DJUSBK)





Financial Select Sector SPDR ETF (XLF)