Data Mining Assignment WQD7005

Student ID: 17202174

Student Name: Ng Kah Yong

# Introduction

This assignment covers the data mining process when solving real-world problems, which is to predict share price and compare the relationship between the share price and KLSE index – the main trading board in Malaysia. The share chosen is Maybank (MBB). We will go through the process from data acquisition, data storage and retrieval, data processing, data interpretation and insights communication to deployment.

# Data Acquisition

This is the first stage of the data mining process where we crawl real-time data from the web using Python. We crawl share information and KLSE index information from Yahoo Finance in order to get the price and index value by date.

Source for MBB share: <https://finance.yahoo.com/quote/1155.KL/history?period1=946857600&period2=1583539200&interval=1d&filter=history&frequency=1d>

Source for KLSE index:

<https://finance.yahoo.com/quote/%5EKLSE%3FP%3D%5EKLSE/history?period1=946857600&period2=1583539200&interval=1d&filter=history&frequency=1d>

The data retrieved is stored in separate data frame with columns ‘date’ and ‘close’.

# Data Storage and Retrieval

The data retrieved is stored into Hive database which acts as the data warehouse for permanent storage. To initialize Hive database, enter this command in terminal:

./Downloads/hive/apache-hive-3.1.2-bin/bin/schematool -initSchema -dbType derby

To enter into Hive, enter this command in terminal

Hive

To create Hive table with the necessary columns and import data into this table directly from Hive, enter this command in terminal:

CREATE EXTERNAL TABLE IF NOT EXISTS mbb(

stockdate STRING,stockprice STRING)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ';'

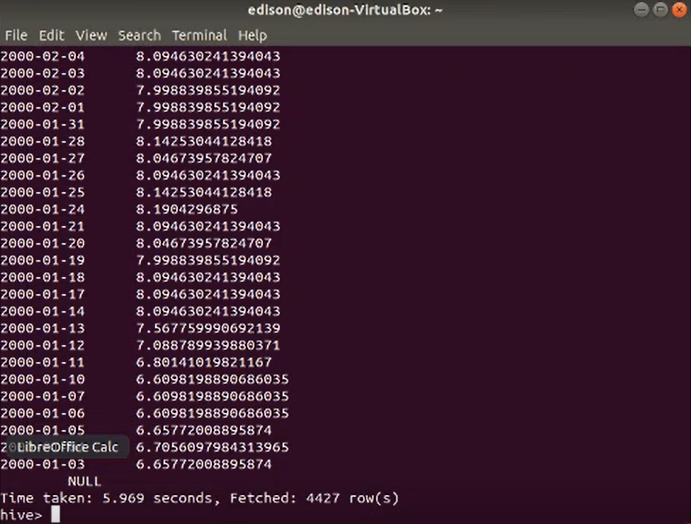
STORED AS TEXTFILE

LOCATION './input/datamining';

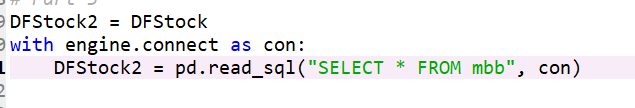
# Data Processing

Data stored in Hive can be retrieved for further processing. To retrieve the data inserted into Hive table, enter this command in terminal:

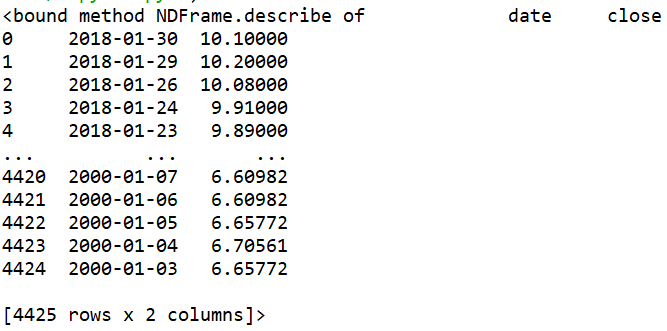
SELECT \* FROM mbb;



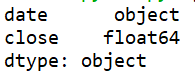
Or in Python, use following command to read data from Hive table:



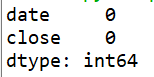
To view the structure of the data frame, use ‘describe’ function. There are 2 columns (date and close) with 4425 observations:



To check the data type of columns in data frame, use ‘dtypes’ function:



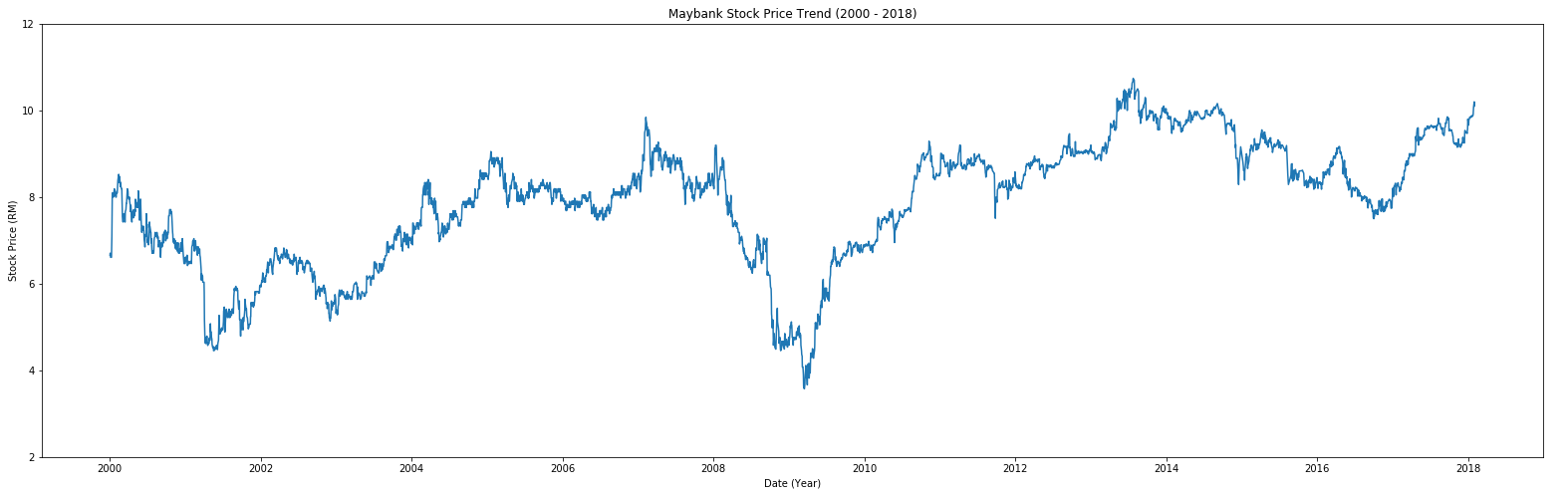
To check for null values, use ‘isnull’ function and use ’sum’ function to calculate total count of null value. There is no null value in the data:



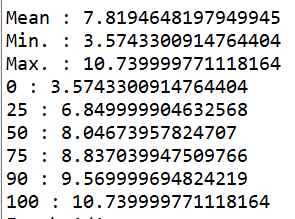
As the data is clean, for example no null value or invalid value, data cleaning step can be skipped.

# Data Interpretation and Insights Communication

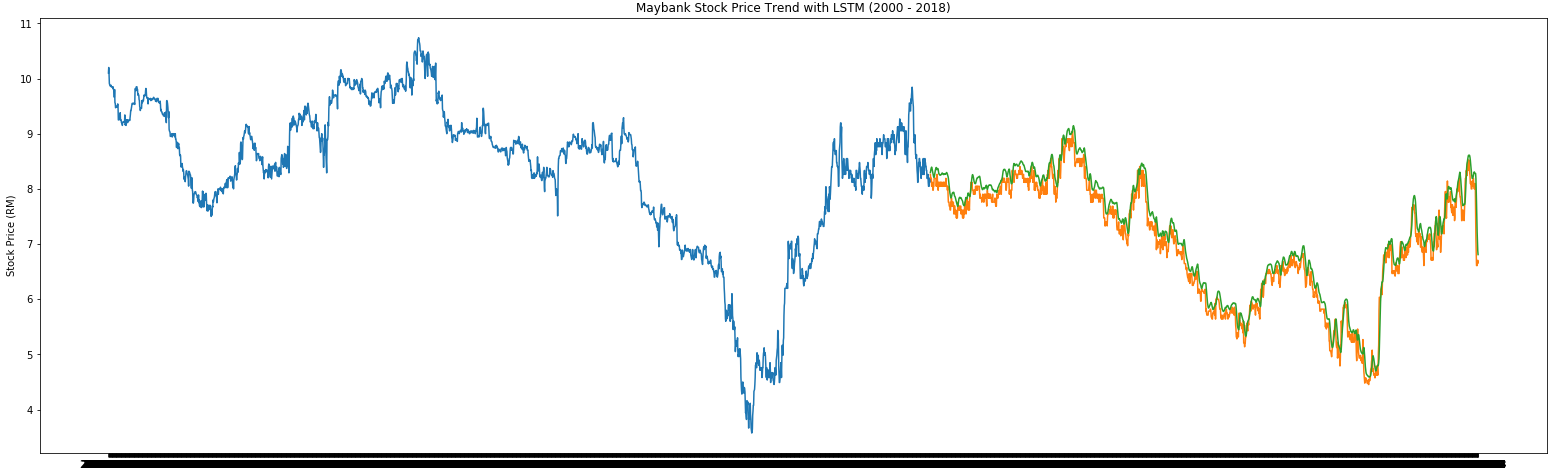
Perform trend analysis by plotting line graph:



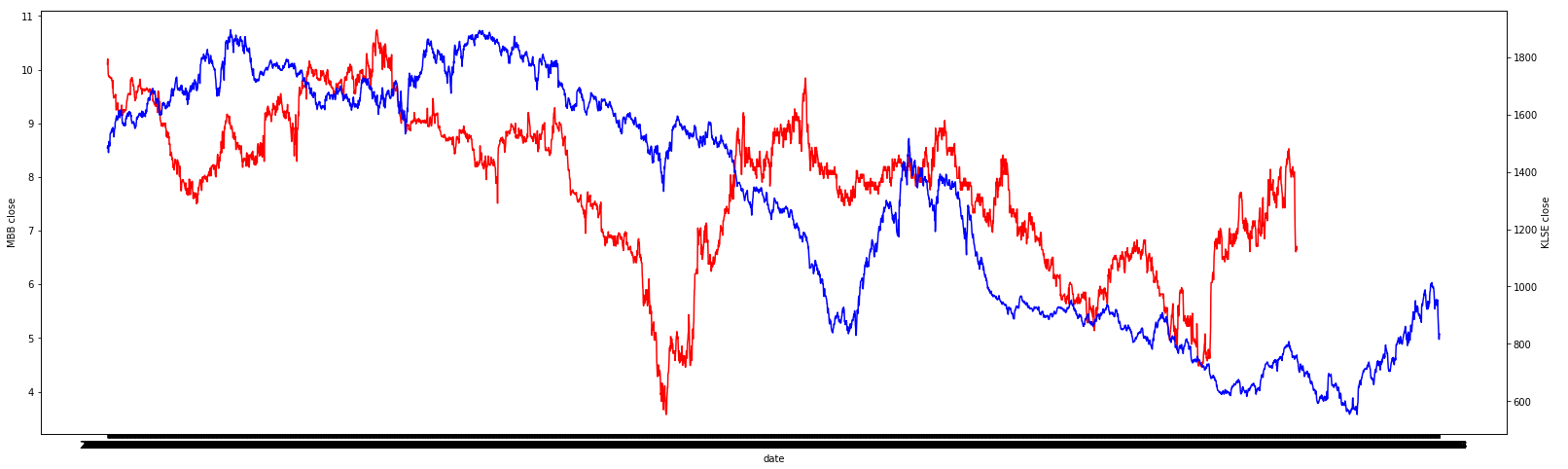
Perform statistical calculation on the data, which include mean, minimum, maximum and percentile at 0, 25, 50, 75, 90 and 100. There is no extreme value / outlier and there is no zero in close price:



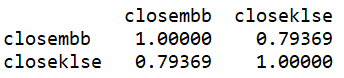
Use Long Short Term Memory algorithm to predict stock price based on dataset provided. The dataset is split into train and validation dataset at the ratio of 60:40. MinMaxScaler is applied to close price to transform the data onto the scale of 0 to 1 before performing prediction using LSTM machine learning model. Once prediction is executed, close price is transformed back to its original form using inverse\_transform function. The predicted price is compared to the real data from validation dataset and the predicted price is close to the actual close price:



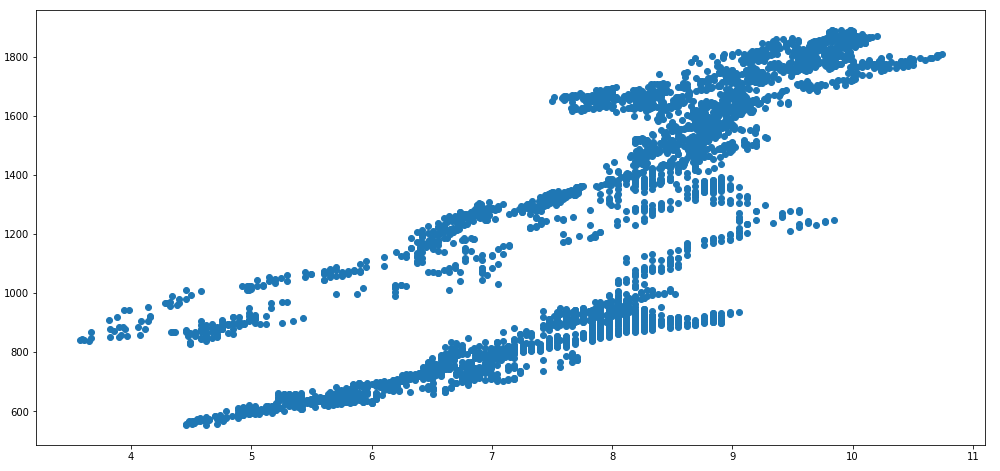
To compare MBB share price against KLSE index, first use line chart to compare their trend. Red line represents MBB’s share price while blue line represents KLSE’s index. The trend of both MBB share price and KLSE index have similar pattern whereby both increase and drop in tandem, although MBB share price happens at a slightly later timeframe:



Next, to find out correlation between MBB share price and KLSE index after merging them into single data frame. There is relatively strong positive correlation between MBB share price and KLSE index at 0.79369:



Use scatter plot to visualize the correlation between MBB share price and KLSE index. The plotted scatter plot supports the relatively strong positive correlation between MBB share price and KLSE index:

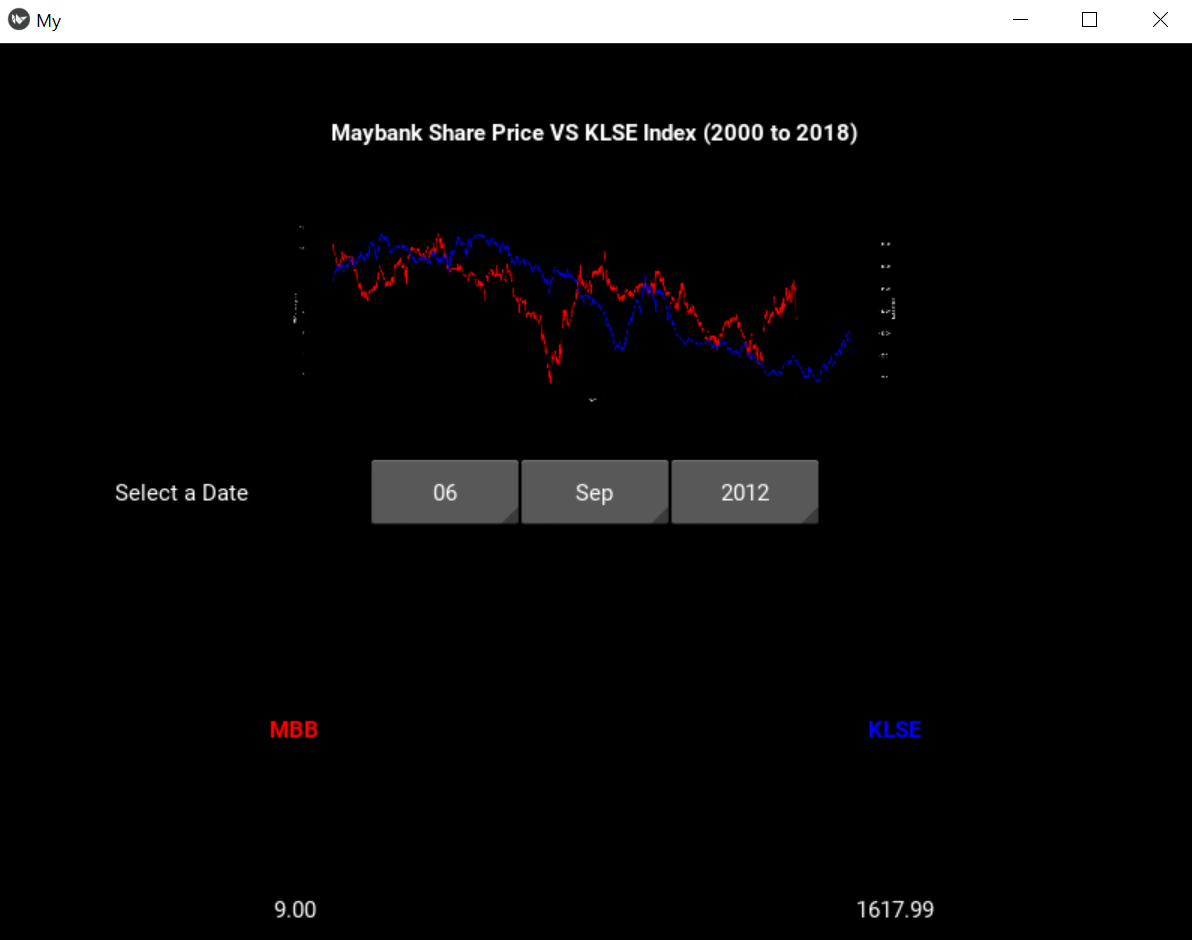


# Deployment

In the final step, the analysed data is deployed on mobile app using Kivy. The mobile app screen will shows the name of the project, the comparison chart between MBB share price and KLSE index, a date input where user is able to select the date to view the actual share price and KLSE index on a particular date.



When a valid date is selected:



# Conclusion

In conclusion, the assignment followed through the data mining process from data acquisition up until deployment for the user. There is a positive correlation between MBB share price and KLSE index and LSTM was used in the share price prediction.