Prediction Stock Price Movement with Machine Learning

WQD7005 Data Mining Semester 2 2018/2019

Name: Lim Kaomin, Leslie
Matric: WQD180076

Problem Statement

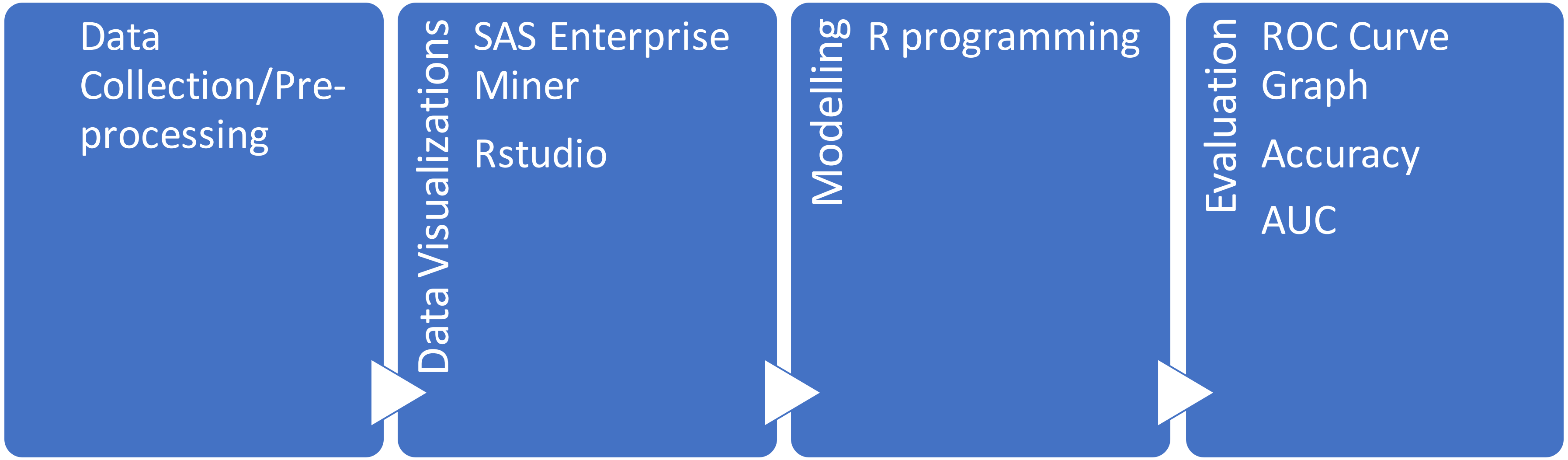
Stock prices are extremely volatile due to internal and external factors such as insider trading, internal development of companies for internal factors and political climate, interest rates for external factors.

Many enterprises, investors want to earn high returns from their investments and need to determine if a stock would be worth investing.

Objectives:

1) Gain Insights into qualitative and quantitative attributes
2)Use Machine learning methods to predict stock price movements
3) Compare the machine learning models

Workflow



Data Collection/Pre-processing

Data Visualizations

- SAS Enterprise Miner
- Rstudio

Modelling

- R programming

Evaluation

- ROC Curve Graph
- Accuracy
- AUC

Data Collection/ Pre-processing

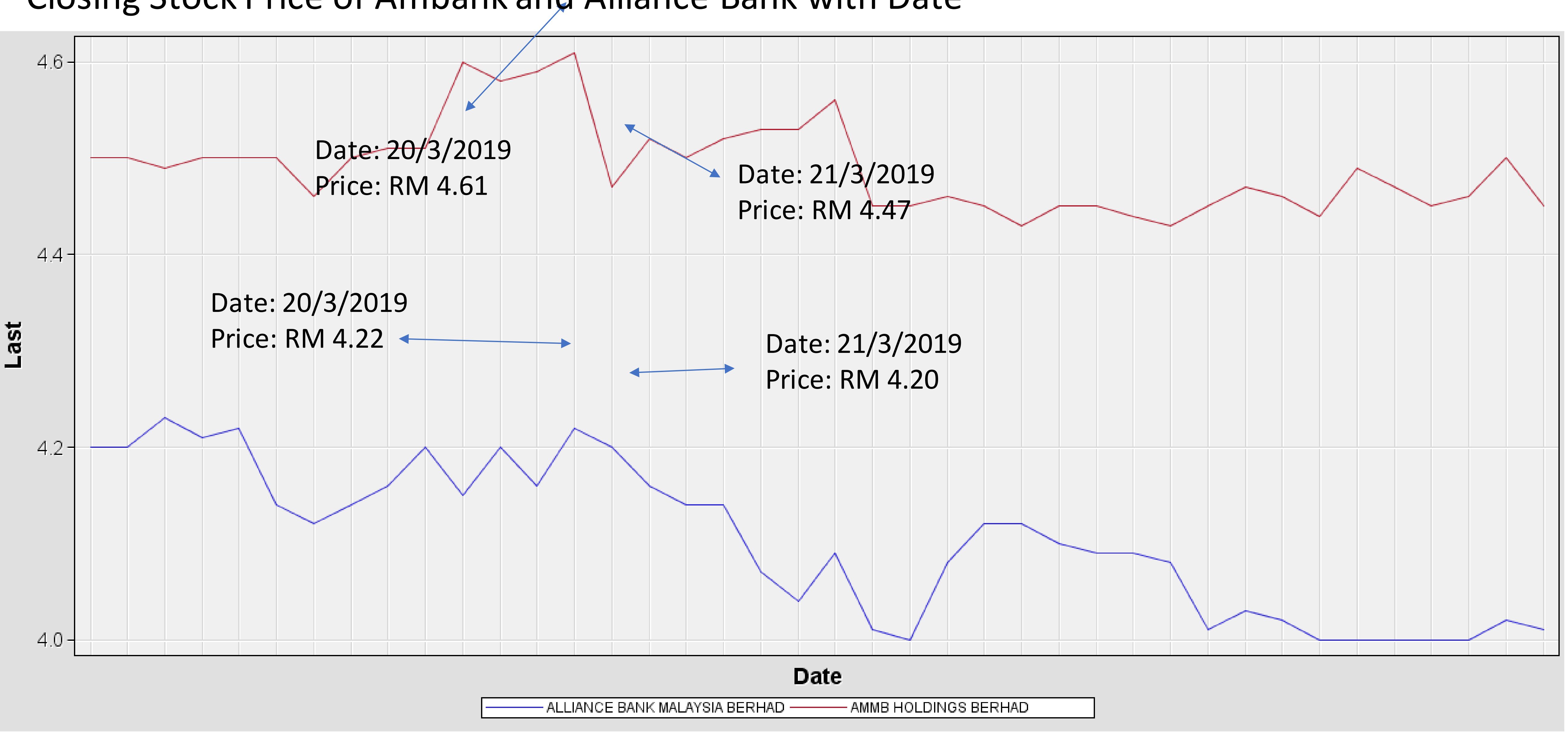
Stocks info are extracted from
https://www.thestar.com.my/business/marketwatch/stocks/?qcounter=
using python and stored in Xampp.

News headlines are extracted from
https://www.klsescreener.com/v2/news
using python and stored in Xampp.

Both data sources will be then combined in Excel file as CSV format

Data Visualizations

Closing Stock Price of Ambank and Alliance Bank with Date



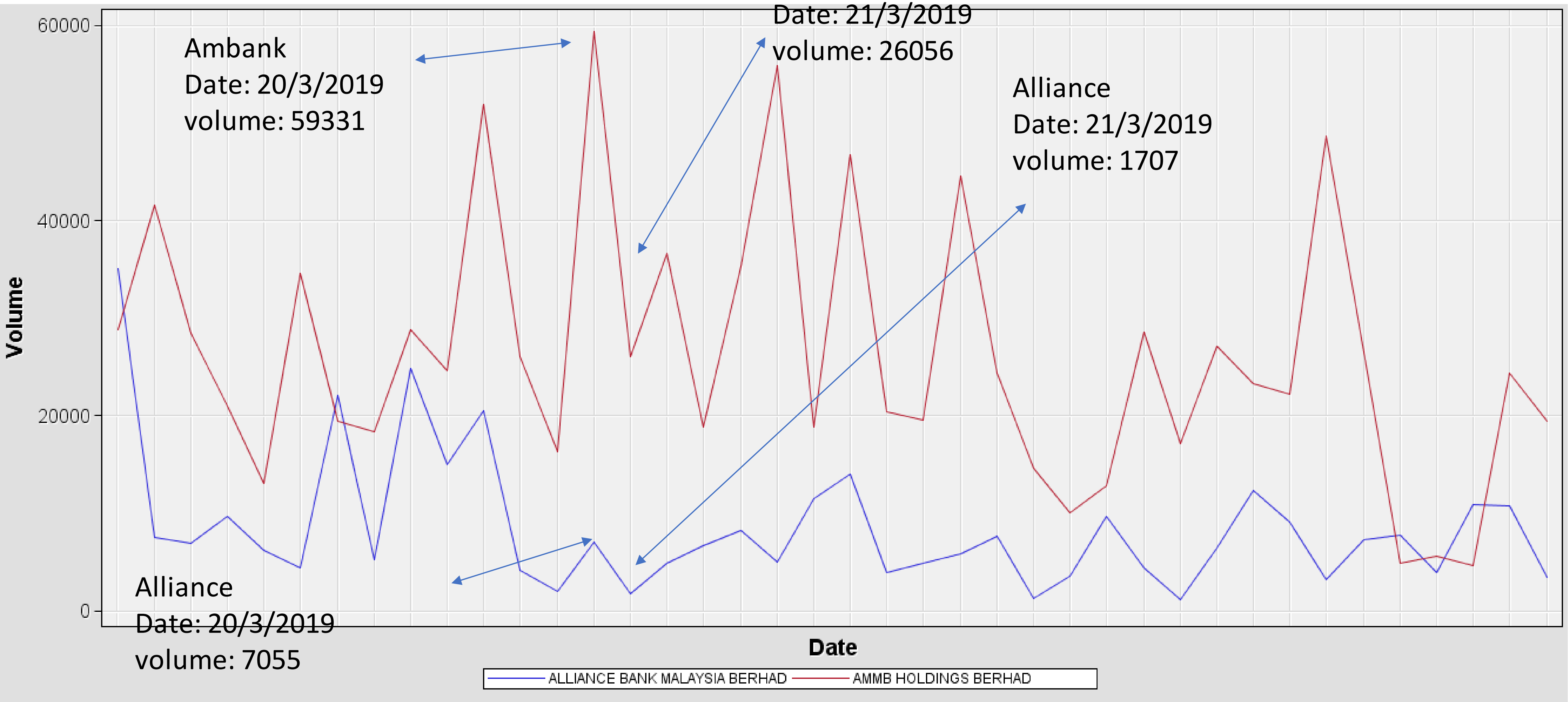
Date: 20/3/2019
Price: RM 4.61

Date: 21/3/2019
Price: RM 4.47

Date: 20/3/2019
Price: RM 4.22

Date: 21/3/2019
Price: RM 4.20

Volume Traded of Ambank and Alliance Bank with Date



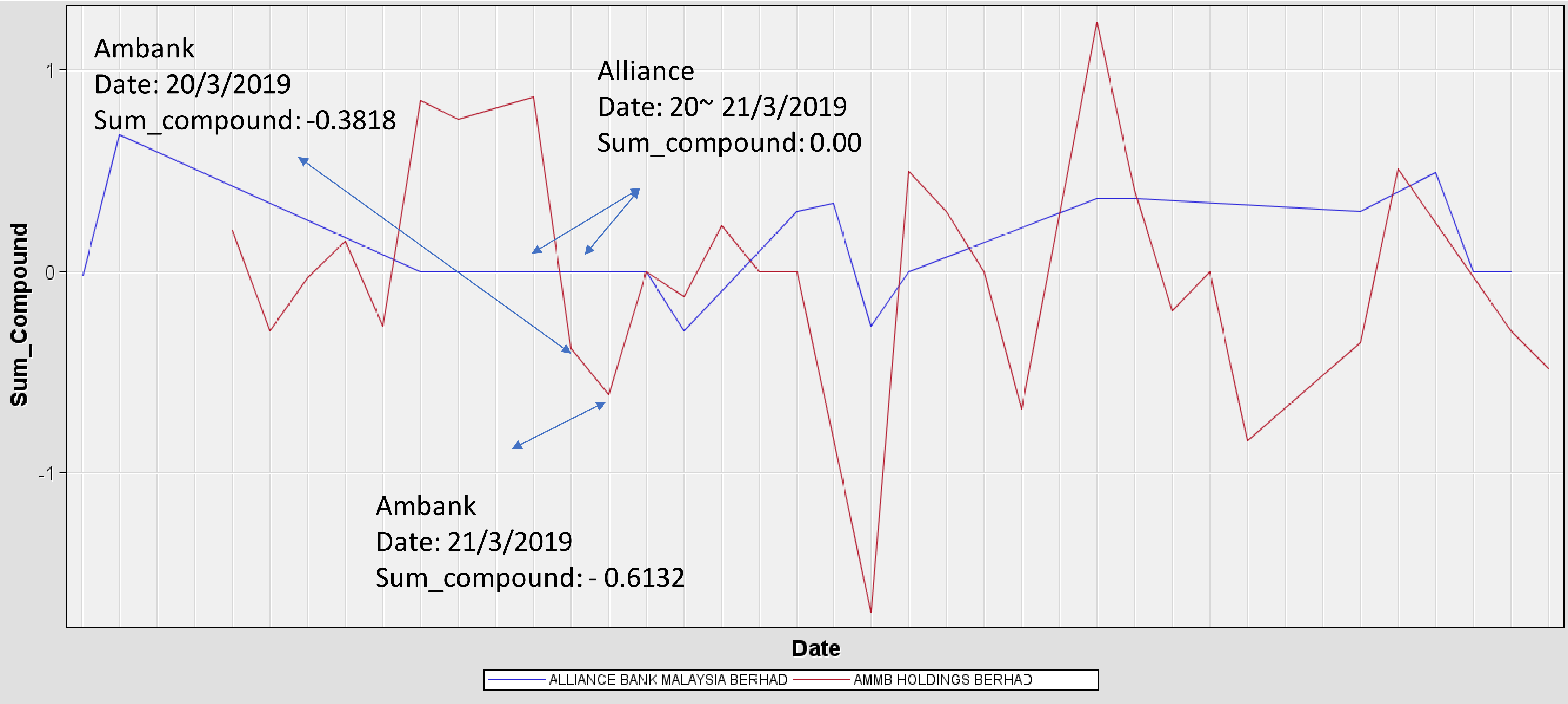
Ambank
Date: 21/3/2019
volume: 26056

Ambank
Date: 20/3/2019
volume: 59331

Alliance
Date: 21/3/2019
volume: 1707

Alliance
Date: 20/3/2019
volume: 7055

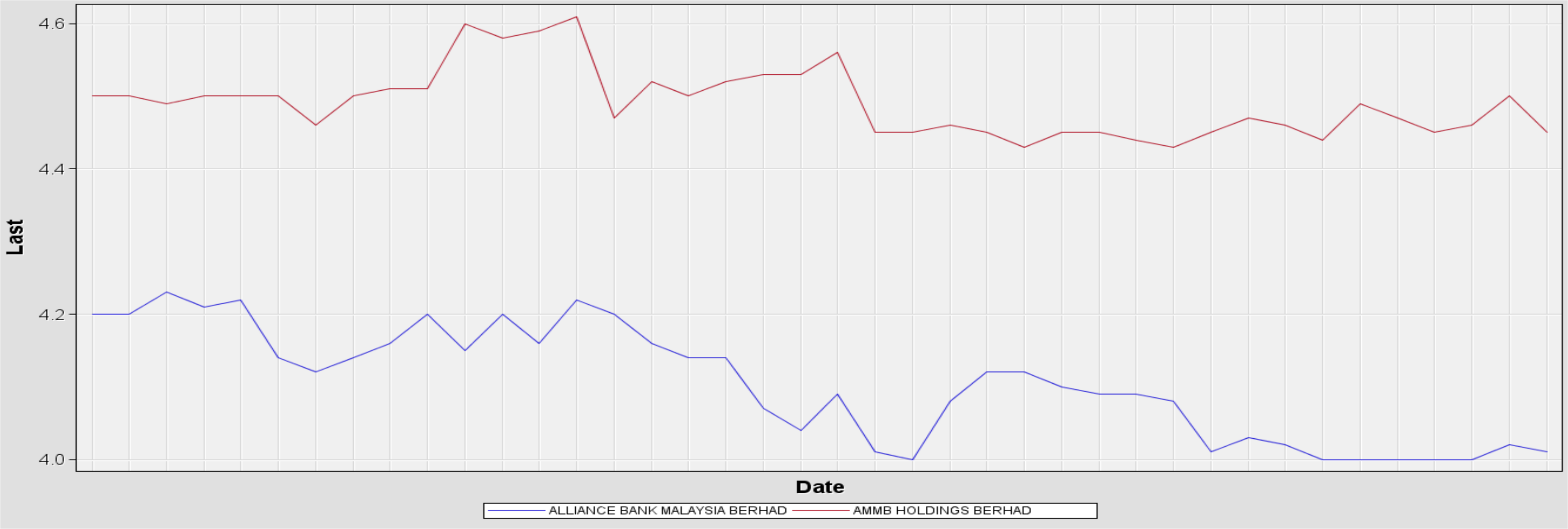
News Sentiment Score of Ambank and Alliance Bank with Date

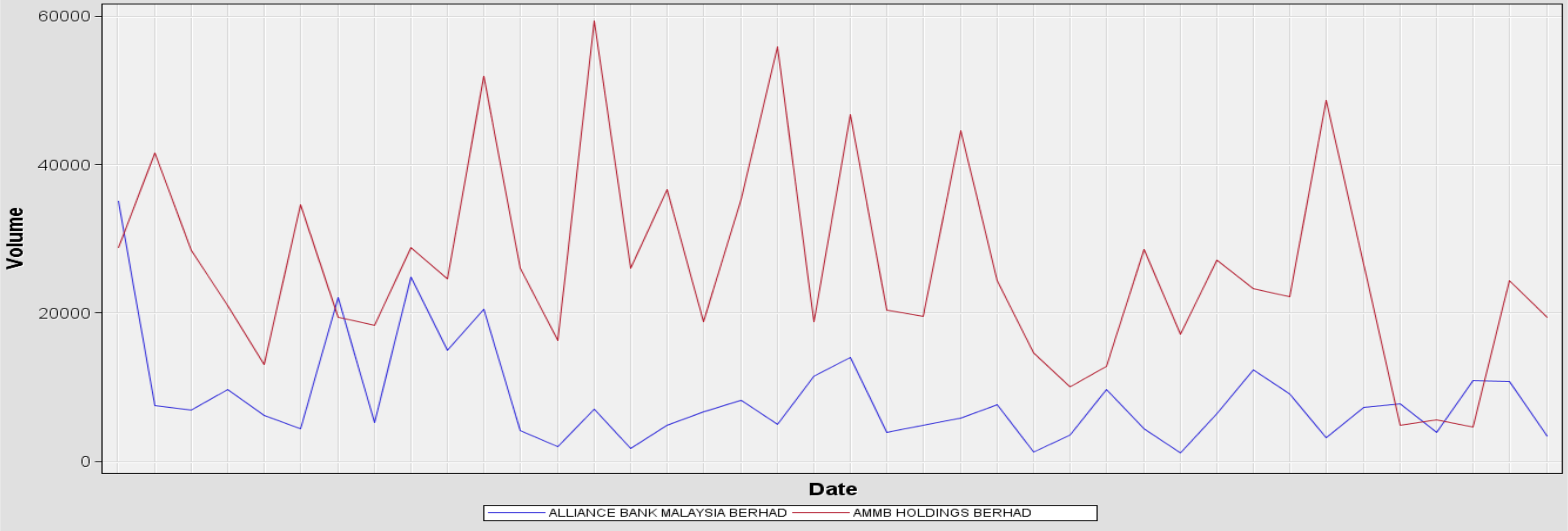


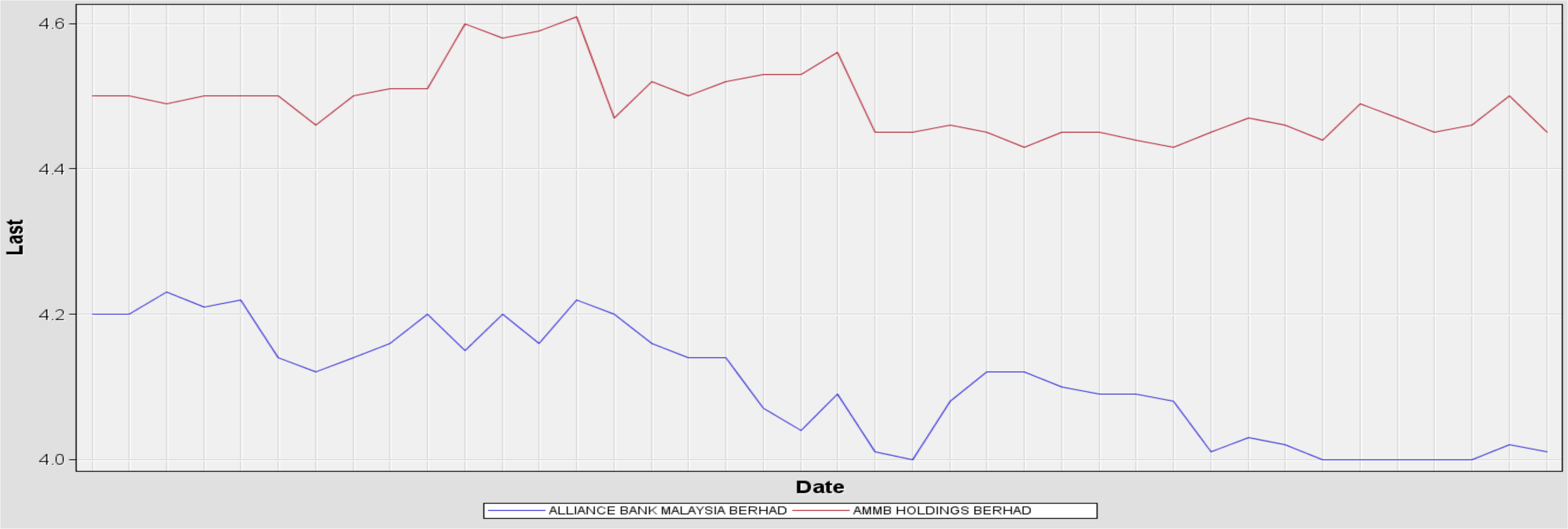
Ambank
Date: 20/3/2019
Sum_compound: -0.3818

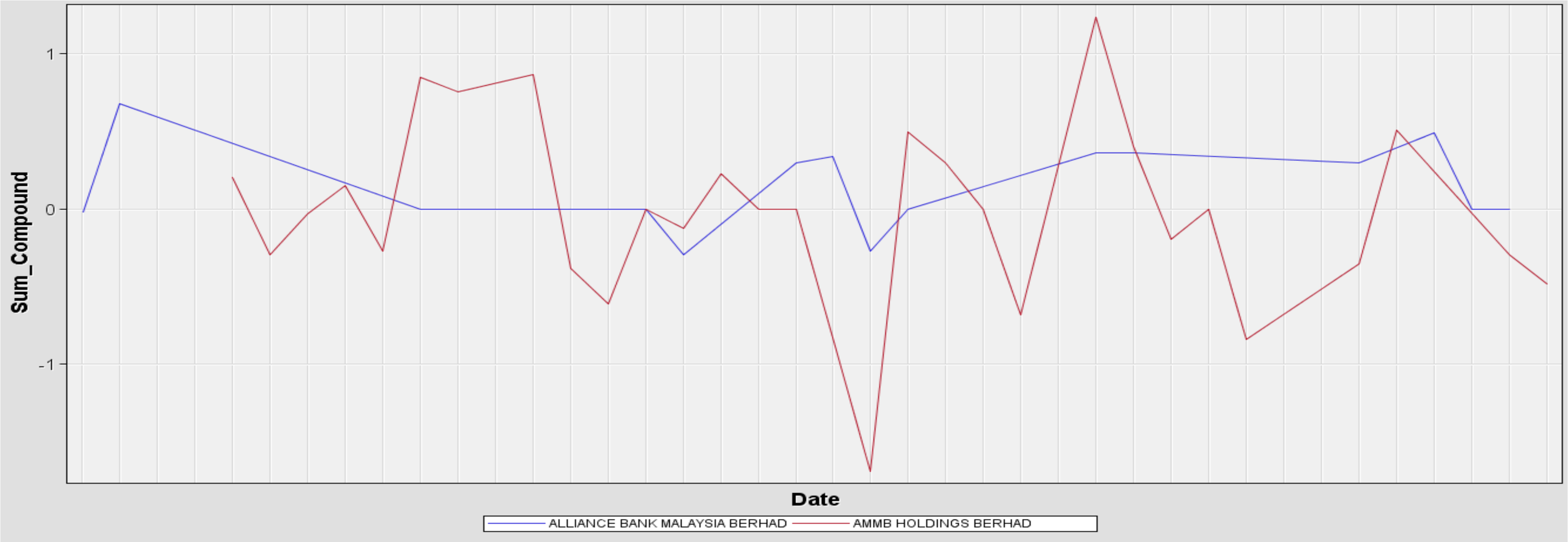
Alliance
Date: 20~ 21/3/2019
Sum_compound: 0.00

Ambank
Date: 21/3/2019
Sum_compound: - 0.6132

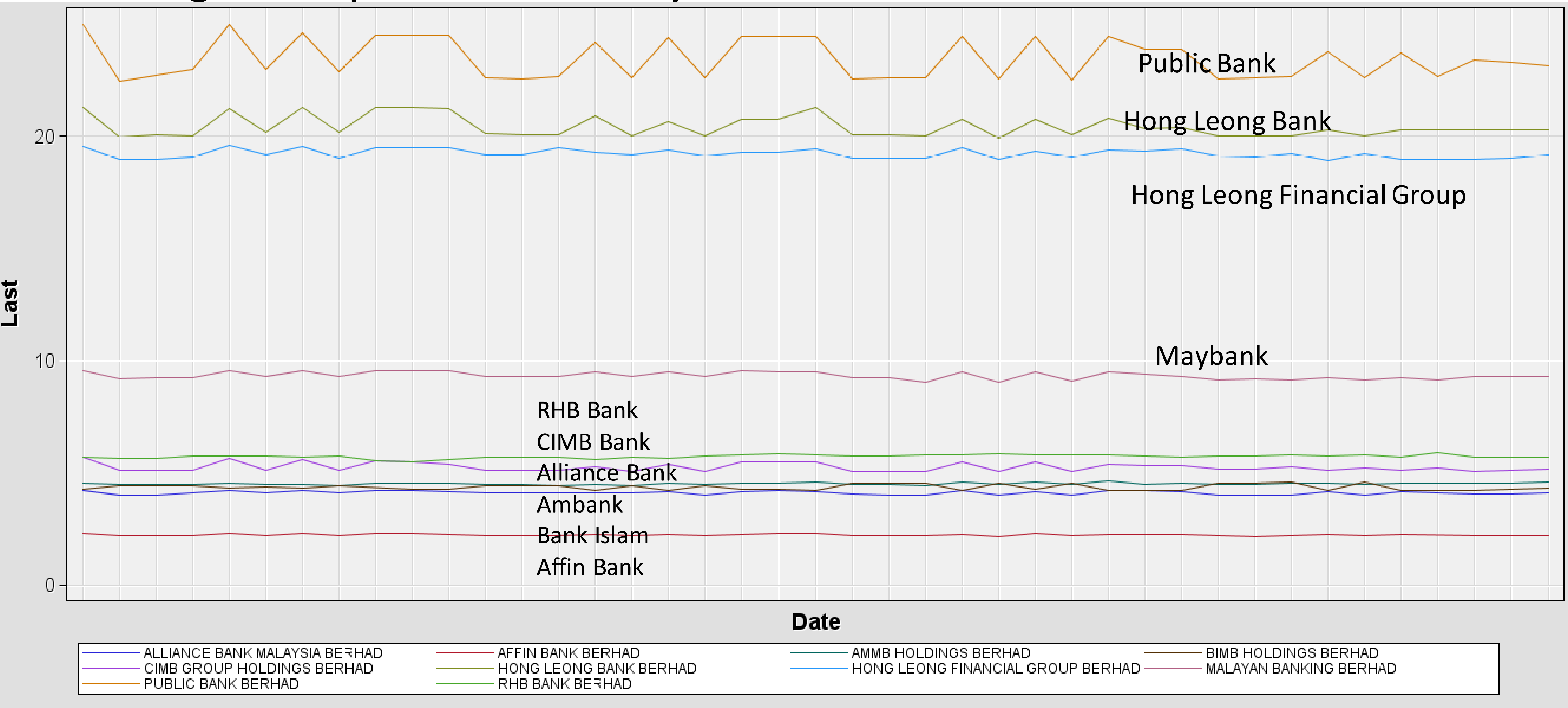

4.6
4.4
4.2
4.0
Last
Date
ALLIANCE BANK MALAYSIA BERHAD
AMMB HOLDINGS BERHAD


60000
40000
20000
0
Volume
Date
ALLIANCE BANK MALAYSIA BERHAD
AMMB HOLDINGS BERHAD


4.6
4.4
4.2
4.0
Last
Date
ALLIANCE BANK MALAYSIA BERHAD
AMMB HOLDINGS BERHAD


1
0
-1
Sum_Compound
Date
ALLIANCE BANK MALAYSIA BERHAD
AMMB HOLDINGS BERHAD

Closing Stock prices for Malaysian Banks



Public Bank

Hong Leong Bank

Hong Leong Financial Group

Maybank

RHB Bank

CIMB Bank

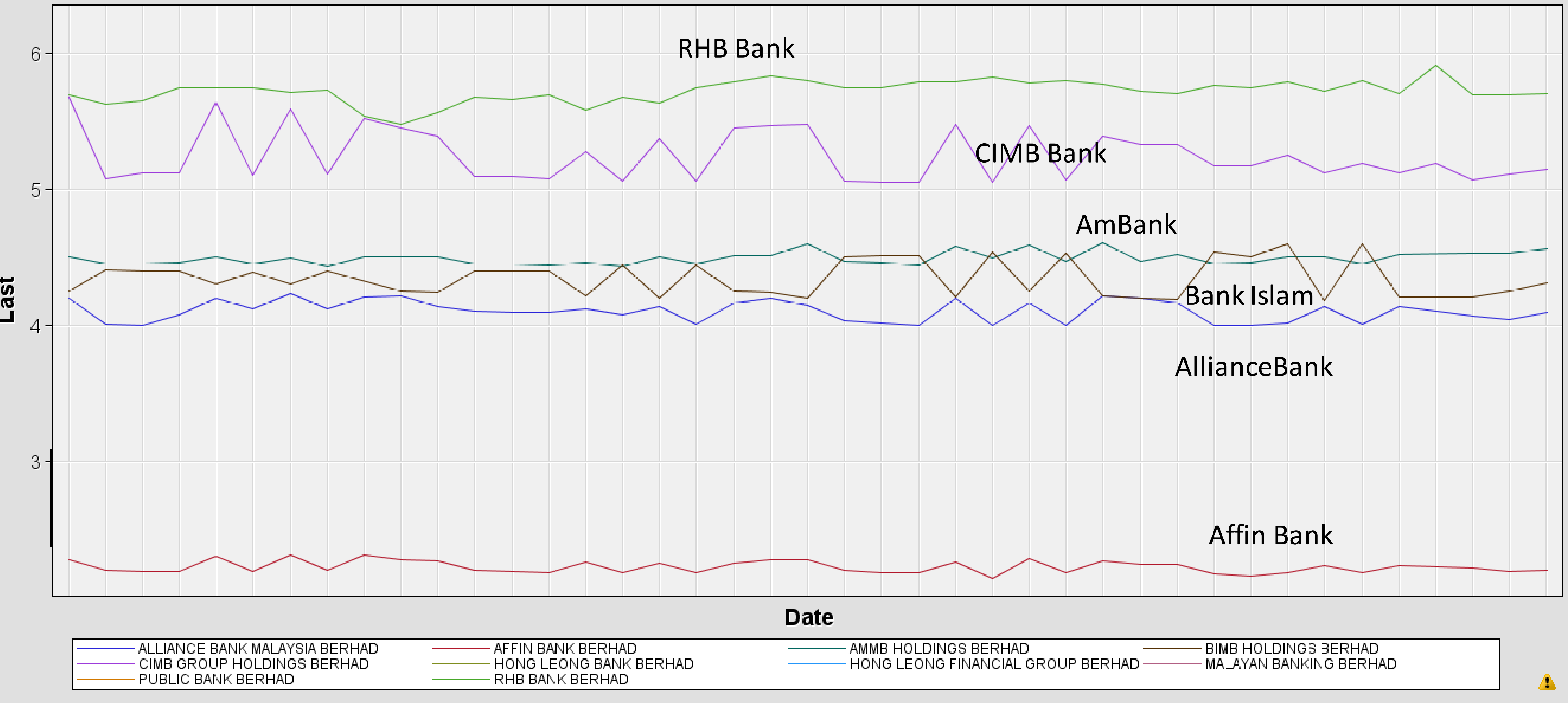
Alliance Bank

Ambank

Bank Islam

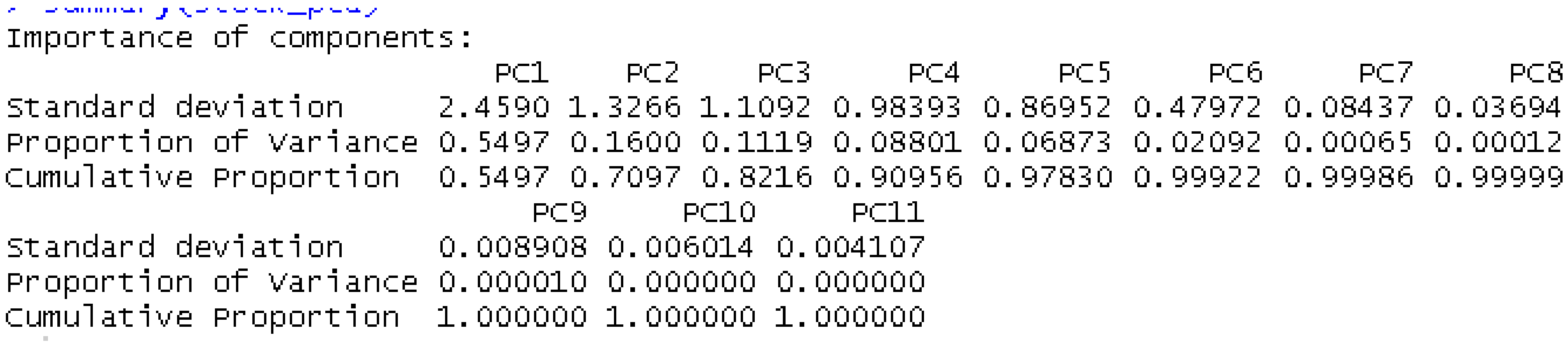
Affin Bank

Closing Stock prices for Malaysian Banks

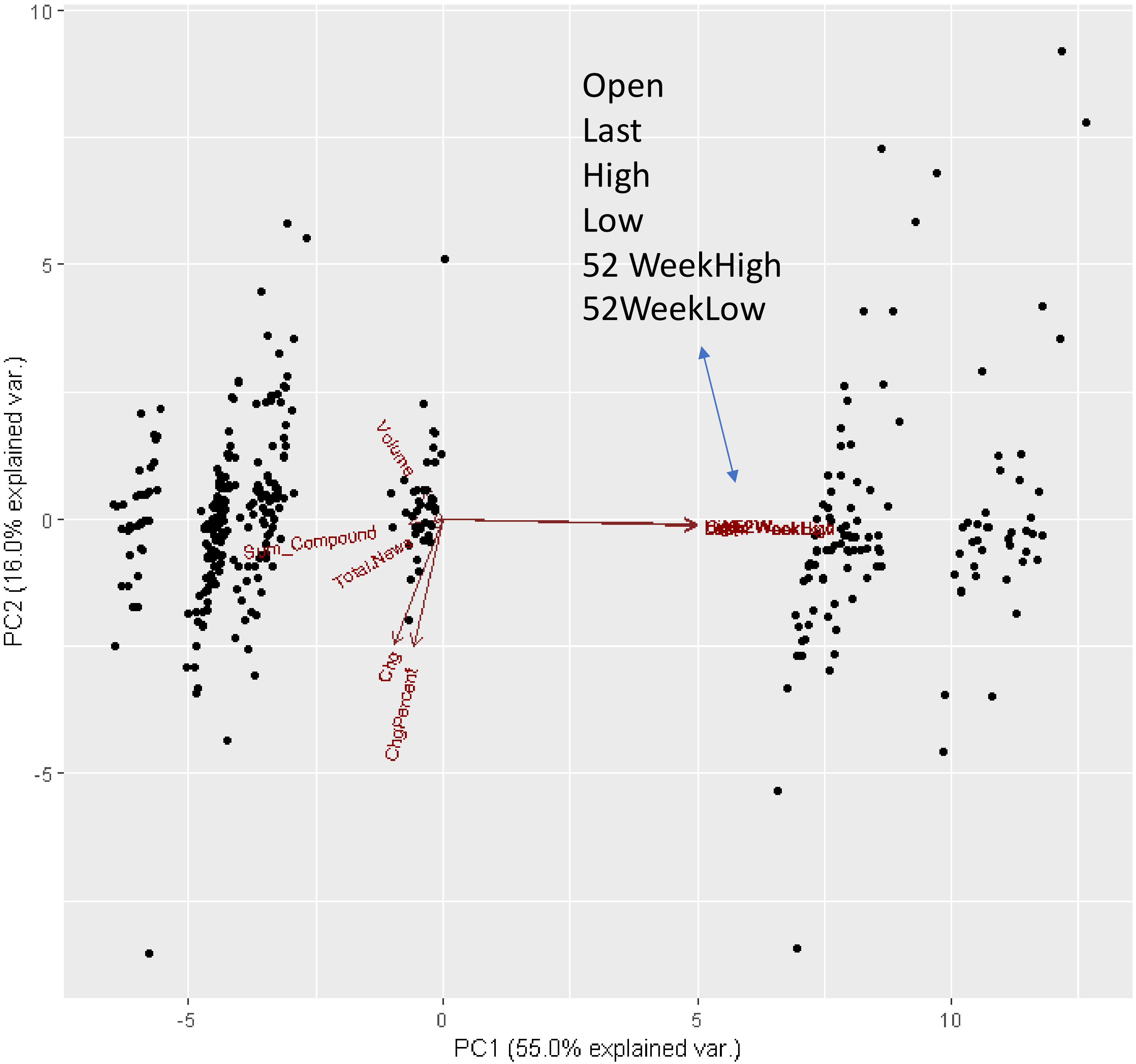


PCA Analysis

Study the correlation between the attributes

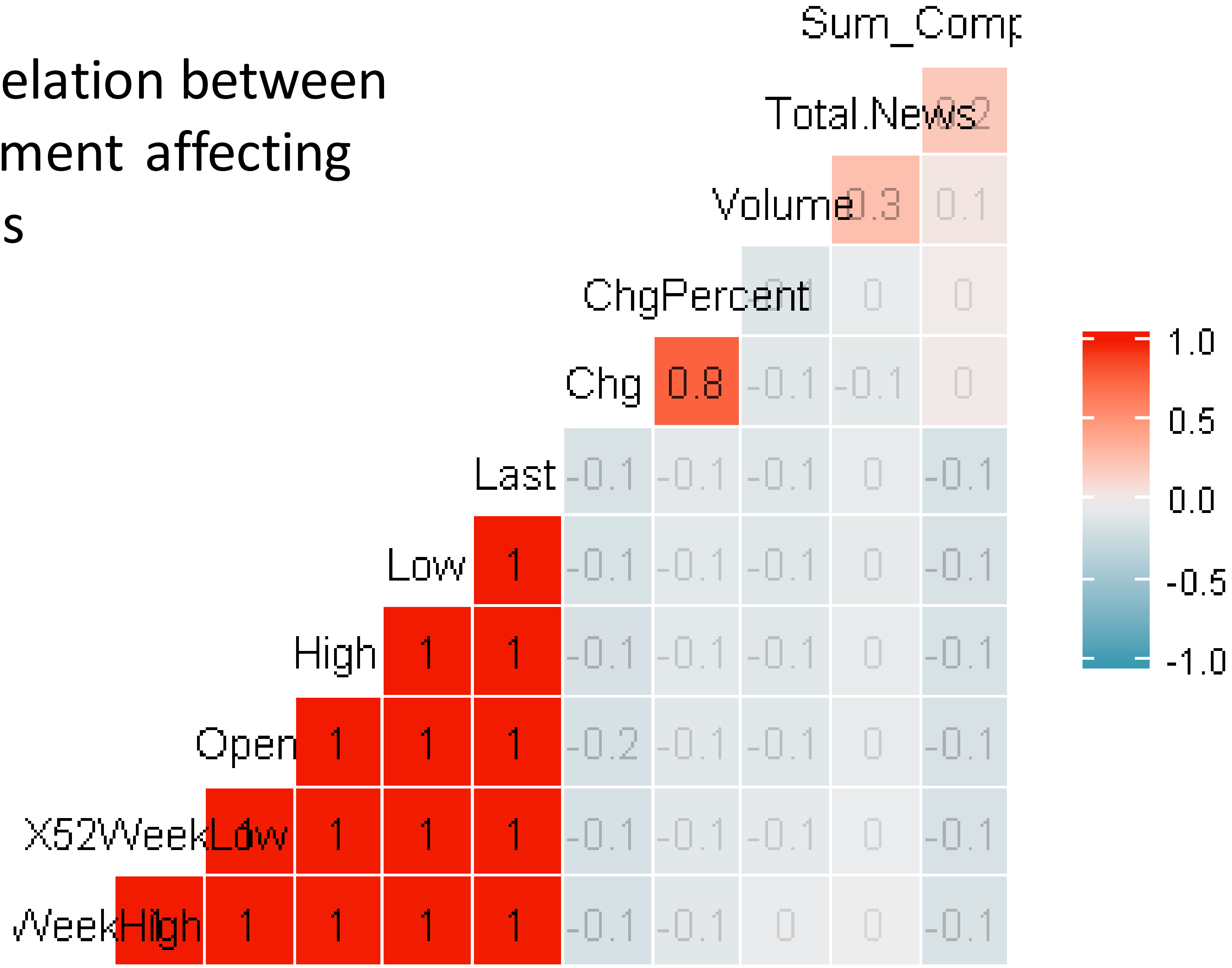


```
> summary(stock_pca)
Importance of components:
                          PC1    PC2    PC3     PC4     PC5     PC6     PC7     PC8
Standard deviation     2.4590 1.3266 1.1092 0.98393 0.86952 0.47972 0.08437 0.03694
Proportion of Variance 0.5497 0.1600 0.1119 0.08801 0.06873 0.02092 0.00065 0.00012
Cumulative Proportion  0.5497 0.7097 0.8216 0.90956 0.97830 0.99922 0.99986 0.99999
                            PC9     PC10     PC11
Standard deviation     0.008908 0.006014 0.004107
Proportion of Variance 0.000010 0.000000 0.000000
Cumulative Proportion  1.000000 1.000000 1.000000
```



Open
Last
High
Low
52 WeekHigh
52WeekLow

Weak Correlation between news sentiment affecting stock prices



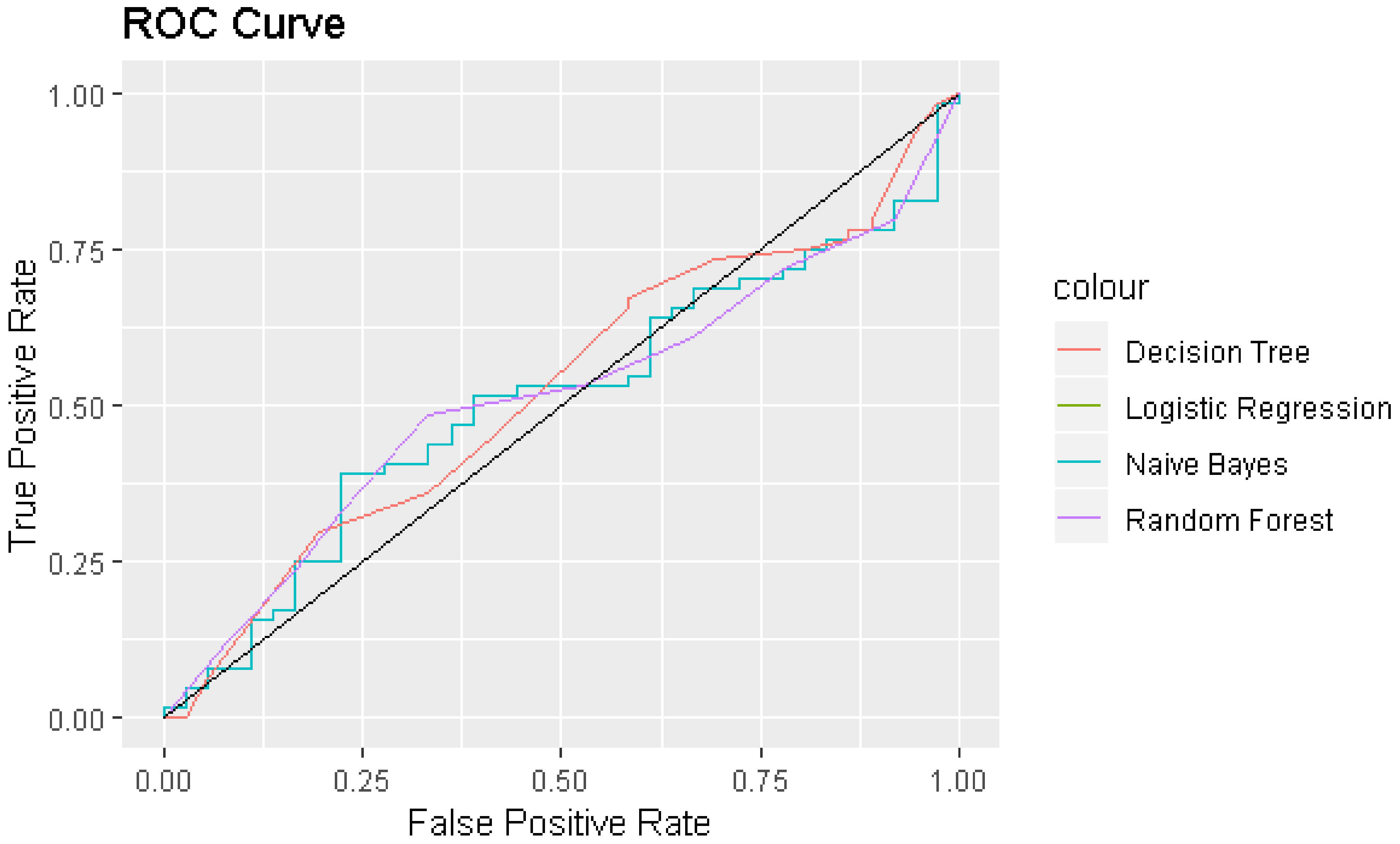
Modelling

R programming Languange

Logistic Regression
Naïve Bayes
Decision Tree
Random Forest

The 'last price', 'price change' and 'percentage of price change' are removed from the dataset, the target variable is the 'PriceLabel'.

Evaluation


ROC Curve
True Positive Rate
0.00
0.25
0.50
0.75
1.00
False Positive Rate
0.00
0.25
0.50
0.75
1.00
colour
Decision Tree
Logistic Regression
Naive Bayes
Random Forest

Results

Model	Accuracy	Area Under Curve
Logistic Regression	46%	0.5128
Naïve Bayes	54%	0.4918
Decision Tree	58%	0.4755
Random Forest	57%	0.5130