Project Summary

Data preprocessing

The data set used in this project was collected from Yahoo Finance and CoinMarketCap. Totally 7 independent variables were collected and used in this project, they are: AR, BR, Volume, VIX, Market Capacity, Vol and MC ratio. Computing the market reflection index AR BR to scale the marketing reflection by its formula and insert them into the BTC data frame. Then collected market afraid of investing index VIX from Yahoo Finance. Because the market was not opened in some certain dates, so used the past day VIX as the VIX in the closed dates. Inserted into the data frame as well. Finally, collected market capacity number from CoinMarketCap and inserted into the BTC data frame. In the end, the BTC data frame contained 1937 rows and 8 columns (Date, Price, AR, BR, Vol, VIX, MC, VMR).

Overview of BTC historical data

Used plotly to visualize the historical data of Bitcoin in the past 5 years to provide a clear idea about the fluctuation and the price changes. In addition, separated all independent variables to two parts: Natural number variables and Market reflection index. Visualized all variables and run into the next section.

Label Assignment

Assigned label based on the market reflection index. Then, came up with the investment suggestion labels from Market Status labels. Inserted all of the labels into the data frame.

Model application

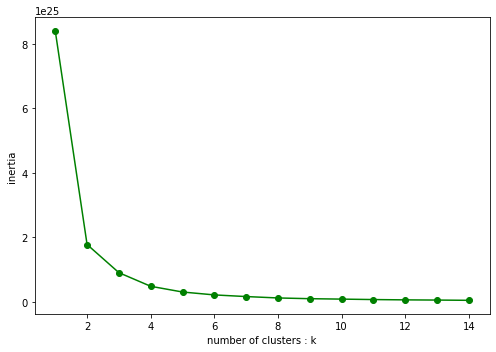
A total of 4 classifers (kNN, Logistic regression, Naïve Bayes, Gaussian SVM) were used to predict the market status label in 2021. One cluster (K-means) to separate the historical BTC data in to K stages.

By computed error rate of different number of k in kNN classifier, all k values rolled out the same error rate, so k = 10 was used to predict the labels. The summary of the result for each classifier is shown in the following table,

|  |  |
| --- | --- |
| Classifiers | Accuracy |
| kNN | 50.877% |
| Logistic Regression | 34.857% |
| Naïve Bayes | 50.446% |
| Gaussian SVM | 8.772% |

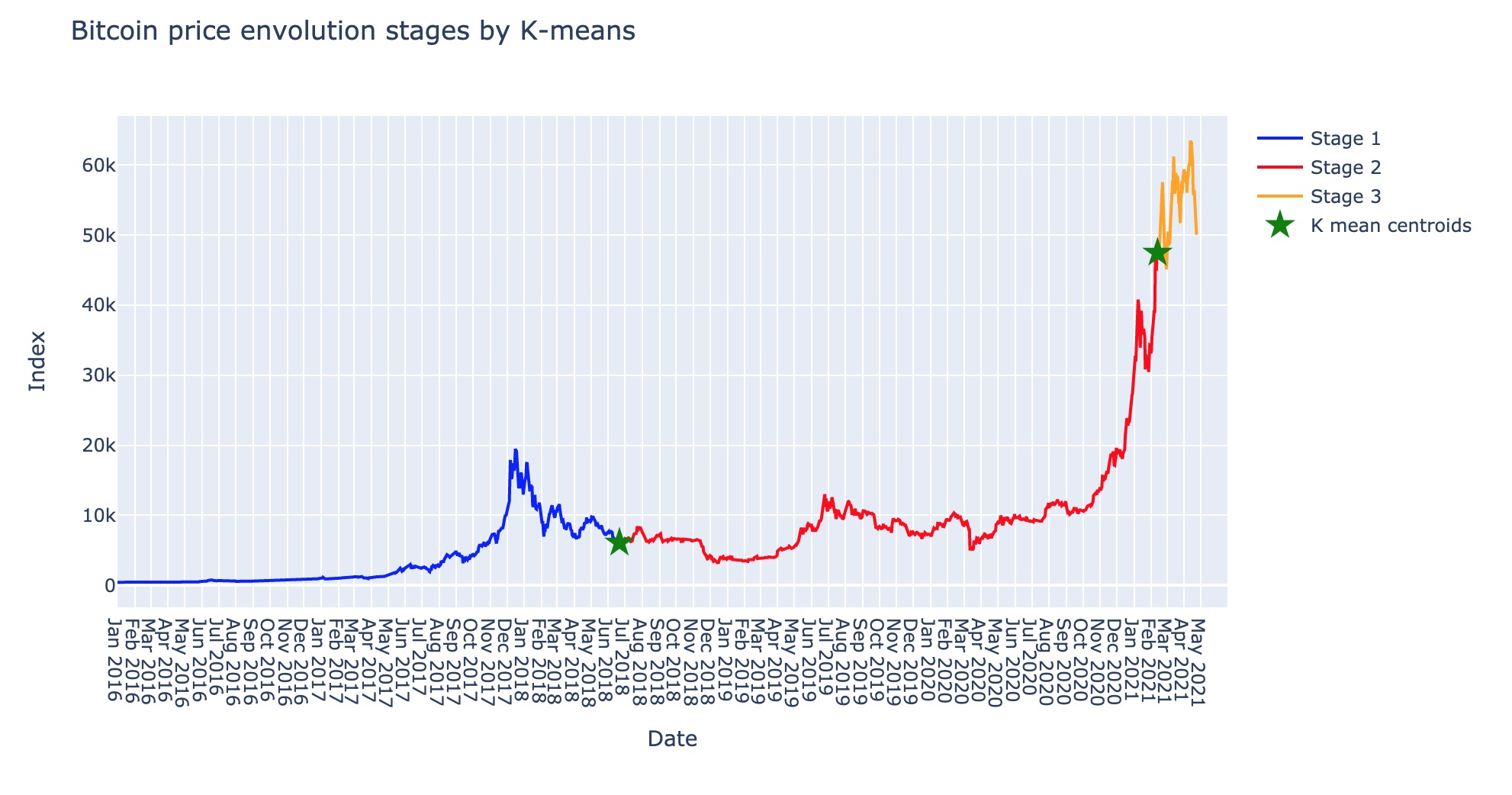
As a result, kNN classifier predicted the best accuracy among four classifiers.

Try k = 1 to k = 15 for K-means to find the inflection point.



From the diagram, it is clear to see that the optimal number of clusters is k=2, since at k=2, there is no significant decrease in loss function. Two centroids were given by K-means cluster.

The Bitcoin price evolution would be separated to three stages by these two centroids. Finding the closest two time slot for two centroids as the bound of each stage to plot.



Trading Demo

Five trading strategies were used. Real label, By k-NN, by Logistic Regression, by Naive Bayes, By Gaussian SVM. Sat train set as BTC from 2016-01-01 to 2020-12-31 and test set as BTC from 2021-01-01 to 2021-04-23. Applied four classifiers to Investment suggestion labels and predict the labels for the test sets. Then, applied the related trading strategy by using predicted label. The accuracy for the prediction of investment suggestion labels and final return was shown in the following table.

Classifier accuracy:

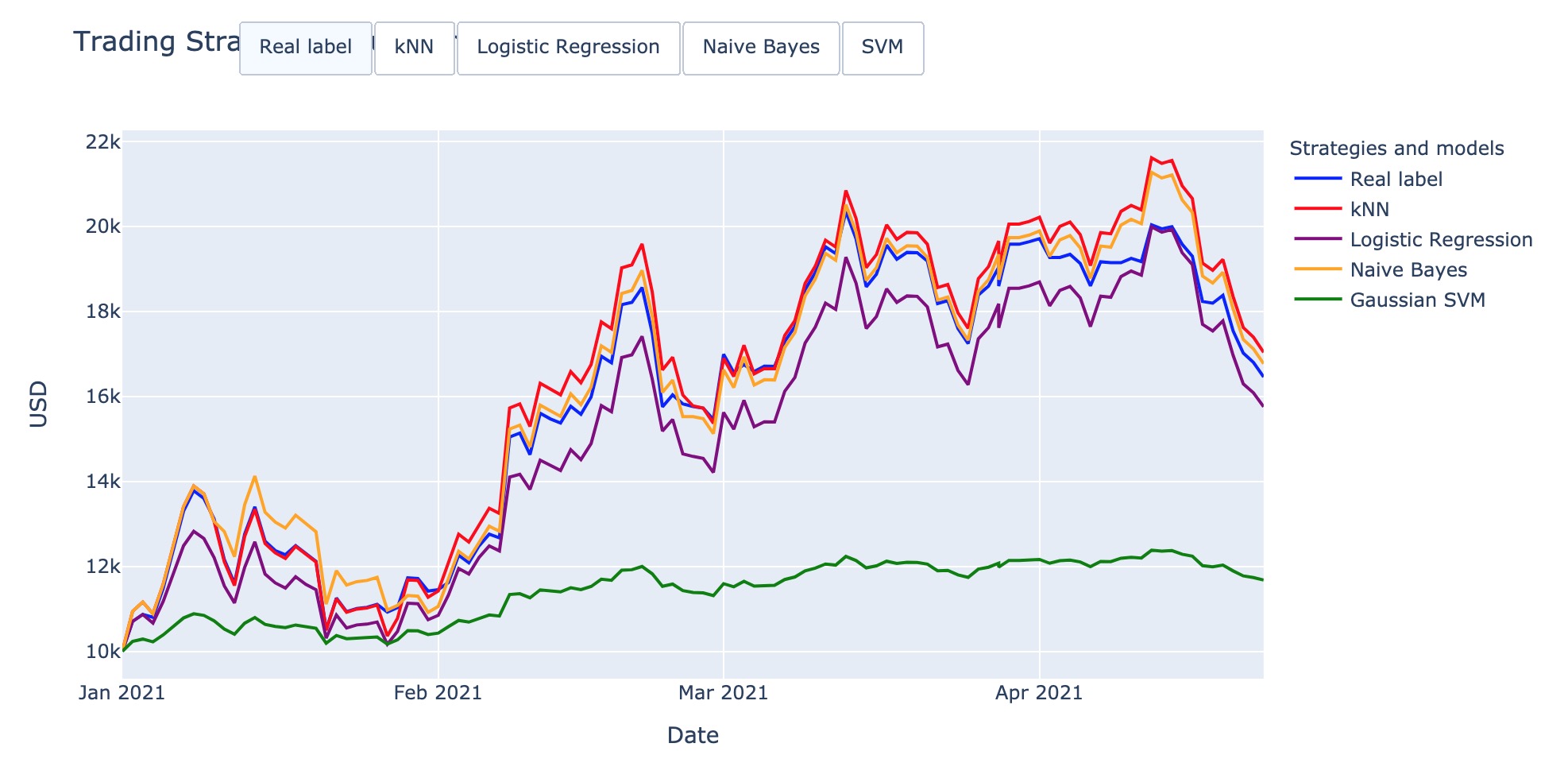
|  |  |
| --- | --- |
| Classifiers | Accuracy |
| kNN | 54.386% |
| Logistic Regression | 49.077% |
| Naïve Bayes | 52.723% |
| Gaussian SVM | 8.772% |

It is clearly to see that, kNN still hold the highest accuracy for investment suggestion label.

Final return by initial invests $10,000:

|  |  |
| --- | --- |
| Trading Strategy | Final return (USD) |
| Real label | 16456.99 |
| kNN | 17039.09 |
| Logistic Regression | 15755.46 |
| Naïve Bayes | 16768.45 |
| Gaussian SVM | 11676.02 |

kNN finally gained the highest money in the first 4 months in 2021 in trading demo.



To sum up, investing is a high-risk behavior, it needs to be considered carefully and need to be modeled and computed a lot before investing. This project visualized the huge fluctuation of Bitcoin price, and find the best model kNN to predict the market status, investment suggestion and the final return in trading demo. More independent variables need to be considered for more complicated prediction such as BPI, ROC, and etc. This project, successful analyzed the historical data and find some rules for the bitcoin price. It clearly provides the relationships between price and other variables. And was able to apply the models to find the tendency.

Reference

<https://support.futunn.com/en-us/topic166#:~:text=(AR)%20and%20willingness%20indicator%20(,the%20market%20The%20degree%20of>