

Summary of Assignment (from Milestone 1 to 5)

Title: Application of Data Mining in Price Movement of Cryptocurrency

Team Member:

1. Tan Chang Jung
2. Tan Sia Hong

Github link: https://github.com/changjung1995/WQD7005_Data_Mining

In Milestone 1, I was crawled the price of cryptocurrencies from <https://coinmarketcap.com/>. The first 20 top cryptocurrencies was crawled by using Python BeautifulSoup. There are 6 columns exist in the dataset, which are close, high, low, open, volume and market capacity. The dataset was cleaned, and the data types were changed accordingly. Lastly, the cleaned dataset was save as csv files.

In Milestone 2, the crawled data are required to store in the data warehouse. We need data warehouse instead of RDBMS database as data warehouse is designed to separate the big data analysis and query process. This is designed to facilitate decision making by consolidate and analyze information at different levels. I was installing Hortonworks Sandbox, as Hortonworks comes with Apache Hadoop and Apache Hive on top of the Hadoop. Hortonworks is an open source, as it faster the deployment process. At the end of installation, I was successfully load the data into Hive database with mysql as database schema.

In Milestone 3, I was connected to Hive database via Python by using *pyodbc* library. First, it is required to create a data server name to using Microsoft ODBC data source, as the connection to Hive was done through ODBC connection. After connected with correct host and port, as well as its authorization, I was able to access Hive database by using Python.

In Milestone 4, I was doing descriptive analytics and predictive analytics on the cryptocurrencies dataset. From the descriptive analytics, it is obvious to see that, when there is one cryptocurrency increased in price, so do the other cryptocurrencies, or vice versa. Furthermore, the analytics told that the price of cryptocurrencies are fluctuated from time to time. In the predictive analytics, a Long-short Term Model from recurrent neural network, was trained to predict the future price one-with one day ahead.

In Milestone 5, we were using python Dash to create the dashboard for web application. Dash is an open source framework on top of Flask. The web application was built to visualize the price of cryptocurrencies, as well as predicting the closing price for cryptocurrencies of next day. For the Android mobile application, we have converted our web application into mobile application by using Android Studio.