The Hong Kong University of Science and Technology
MATH 4995 Capstone Project for Data Science
Fall 2021 Project Proposal
Project Topic: G-Research Crypto Forecasting (Kaggle)

Student:
Leung, Ko Tsun (SID: 20516287)
Cheng, Tsz Yui (SID: 20595441)
Yang, Po-Yen (SID: 20561878)

Project Overview:

- https://www.kaggle.com/c/g-research-crypto-forecasting/overview

Problem Identification:

The cryptocurrency market has been skyrocketing recently, and it is estimated that over $40 billion worth of cryptocurrencies are traded every single day. Cryptocurrencies have become one of the most popular and trending assets for speculation and investment, however, it has been proven to be wildly volatile, where a person can make a fortune and become a millionaire in one day, and lose all his assets the day after. While a few people have made a great fortune through the fast-fluctuating prices, others have been experiencing losses. Hence, we would like to try whether we can predict some of these price movements in advance and forecast short term returns in 14 popular cryptocurrencies through machine learning techniques.

Background Knowledge:

The simultaneous activity of thousands of traders ensures that most signals will be transitory, persistent alpha will be exceptionally difficult to find, and the danger of overfitting will be considerable. In addition, since 2018, interest in the cryptomarket has exploded, so the volatility and correlation structure in our data are likely to be highly non-stationary. The successful contestant will pay careful attention to these considerations, and in the process gain valuable insight into the art and science of financial forecasting.

Model and Methods:

Given a time-series based dataset of cryptocurrency prices, we can use the hierarchical time series model, which is to train a model for all time, models per weekday, model per day, etc. and ensemble them together. Multiple popular models will also be applied and we are going to conduct research about the effectiveness of these models. For example, LSTM is popular among quantitative finance fields, as it can capture the relationship between previous cryptocurrency prices. Since cryptocurrency price is a time-dependent variable, multiple time-series prediction models may also be adopted, like ARIMA, GARCH, and Dynamic linear model. We aim to create innovative methods, bringing more insights to the quantitative finance field.

Data Source:

- Kaggle : https://www.kaggle.com/c/g-research-crypto-forecasting/data

They have amassed a dataset of millions of rows of high-frequency market data dating back to 2018 which we can use to build our model. Once the submission deadline has passed, our final score will be calculated over the following 3 months using live crypto data as it is collected.