Spring 2023 ORIE 5741 Project Proposal

1. **Project Topic:** Predicting Future Asset Performance with Historical Data
   Team Members: Zhiting Zheng (zz485), Jianing Guo (jg2353), Jingxuan Liu (jl3627)

2. Problem Statement

2.1 Question Identification

In this project, we want to use historical data of our interested asset (stocks or crypto currencies) to predict whether they will have positive excess returns in the future. We may apply different machine learning models, such as SVM, Random Forest, AdaBoost, etc., and compare the asset performance under different models.

2.2 Significance

Our project aims to shed light on the investment strategies for investors in the financial market. While investors need to make informed decisions based on accurate predictions of asset price moving directions to decide whether to long or short an asset, this is a complex and challenging task as the financial market is influenced by multiple factors and it is hard to explore this complex logic through a fundamental approach. Therefore, if we can develop a machine learning model to predict asset price moving direction with high accuracy and in real-time, we will help investors to process vast amounts of data in real-time, to make accurate and up-to-date predictions and to make wise investments.

3. Dataset

3.1 Data Source

We will download the data for several our interested stocks from Yahoo Finance or may download the data from the Exchange for crypto currencies.

3.2 Data Description

Our raw data will be historical price-volume data of our interested assets with frequency of a day (for stocks) or an hour (for crypto currency). We will choose several stocks or crypto currencies, each of which has several fields such as datetime, open price, high price, low price, close price, volume, and turnover, etc.

3.3 Data Rationality

- Accessibility: Our potential data can be obtained from the source described above, so investors can not only replicate our results but apply our model with data in the future.
- Relevance: As we want to use historical data to predict the future asset performance, the historical data that we described before is obviously the reasonable dataset. In addition to the raw historical data downloaded from the source, we may also compute several technical indicators as input of our machine learning model.
- Consistency and Accuracy: The sources we use are the most popular and well-known platforms that can provide accurate historical data for financial investors. Besides, the data we use are recorded based on the historical truth, so it is obviously consistent.