Proposal of Smart Money Strategy

Part 1: Objective and Research Questions

This project aims to determine whether we can construct a smart money strategy using 13F data and price data from 2000 to 2020. To achieve this, we aim to answer one or more of the following research questions:

- What can we learn from the holdings of top institutional investors using 13F data and track their trading patterns over time?
- Can we use machine learning techniques to predict the future performance of these holdings and develop a trading strategy based on these predictions?
- Can this trading strategy generate higher returns than the market over the long term?

Part 2: Introduction to 13F Filing and Price Data

The Securities and Exchange Commission's (SEC) Form 13F is a quarterly report that all institutional investment managers must file with at least $100 million in assets under management. It discloses their equity holdings and can provide insights into what the smart money is doing in the market. Our datasets consist of 13F filings and price data from 2000 to 2020, which contains over 100 million records. The 13F filing data contains a list of equity securities, the number of shares held, and the total value of those securities by institutional investment managers. Besides 13F, we also found the full dataset of prices of all the listed stocks in the United States. The price data includes the ticker and these securities' daily opening and closing prices, which also have over 50 million records.

Part 3: Challenges of Handling Large and Messy Datasets

Handling large and messy datasets presents several challenges, including data cleaning and preprocessing to remove missing or erroneous data, dealing with the large volume of data, and reducing the computational time required for analysis. For example, evening reading the CSV file of the two datasets costs more than 2 minutes. We must employ efficient algorithms and techniques such as parallel processing, multi-threading, and cloud computing to overcome these challenges.

Part 4: Importance of the Project

Constructing a smart money strategy using 13F and price data is an important project because it can identify investment opportunities, provide insights into the investment patterns of institutional investors, and contribute to developing more sophisticated investment strategies that can outperform the market over the long term.

The investment patterns of institutional investors can reveal valuable information about market trends and potential opportunities for investors. By analyzing the 13F data, we can identify which stocks and sectors are being bought or sold by institutional investors and adjust our investment strategy accordingly. The use of machine learning techniques to predict the future performance of these holdings can provide us with a competitive advantage in making investment decisions.

Moreover, as a class project, this is a great opportunity for us majoring in financial engineering to be exposed to the real-world scenario since the real-world data can be as massive as in this case. We cannot just grab an Algorithm taught in the textbook and run it. This time, we must look cleverly, extract necessary information from the data, and employ efficient algorithms.

Part 5: Methodology and Approach

Our approach involves cleaning and preprocessing the data, analyzing the 13F data to identify institutional investors' holdings, using price data to track their performance, and developing machine learning models to predict future performance. We will use these predictions to create a trading strategy that we'll backtest and optimize before deploying it in a live trading environment.

We will employ parallel processing, multi-threading, and cloud computing techniques to speed up data processing and reduce computational time. We will evaluate the performance of our strategy using various metrics such as the Sharpe ratio, Sortino ratio, and maximum drawdown.

In conclusion, this project requires technical expertise, efficient algorithms, and advanced machine learning techniques to analyze and interpret the large volume of data. The outcome of this project will be a smart money strategy that can generate higher returns than the market by leveraging the investment patterns of institutional investors. This project is important because it can identify investment opportunities, provide insights into the investment patterns of institutional investors, and contribute to developing more sophisticated investment strategies that can outperform the market over the long term.