Report 2: Coding cognitive bias in investing decisions.

Overall Strategy

We chose a quantitative research strategy for this demanding project. This decision is not arbitrary, but rather firmly embedded in the core of our project. Because our goal is to understand the complicated world of financial markets and investor behavior, a quantitative approach provides us with the precision and objectivity required to work in such a numerically rich environment.

Our journey begins with the gathering of facts. We collect financial market data and investor activity data as a typical step, but also as a critical basis. This data is the heart of our reinforcement learning model; it contains insights about market patterns, investment decisions, and, most crucially, indicators of cognitive biases that we hope to incorporate into our model.

The following step is to cleanse the data. Given the importance of data in our project, maintaining its cleanliness and correctness is a must. The quality of this data has a direct impact on the performance of our reinforcement learning model, influencing how well it learns and, ultimately, predicts investor behaviour.

We proceed to develop our preliminary reinforcement learning model now that we have clean data. This is where cognitive psychology and machine learning begin to intersect. We work hard to embed cognitive biases into the model's framework, ensuring that the model does more than just recognize patterns—it recognizes and incorporates the psychological factors that drive investor behavior.

Following the initial design, we do a thorough data analysis. This stage is about uncovering the cognitive biases lurking within the data, not just analyzing it. We seek numerical proof of these biases in investor behavior and market movements, which will help us enhance our model and make it a realistic mirror of real-world investor psychology.

Finally, we concentrate on improving the algorithm. The importance of our quantitative approach becomes obvious once more. We don't refine solely on gut instinct or anecdotal information. Instead, we employ cold, hard data—statistical feedback and rigorous testing—to improve our model.

Overall, our quantitative research strategy is inextricably linked to our project. It provides us with the numerical and statistical tools we need to address the particular problems of our project, bringing us closer to our objective of developing a robust, cognitive biases-informed reinforcement learning model for optimizing financial investment decisions.

Creating a reinforcement learning model with cognitive biases to examine the irrationality of trading decisions within a single trading firm.

1. Requirements for Data Identification: For our project, we identified and gathered two types of data: Financial Market Information: We've compiled a detailed dataset from the individual trading firm in question, which includes historical daily closing prices, trading volume data, and market indexes information. This information will form the foundation of our model, providing the market context in which the firm operates and makes trading decisions. We've compiled a thorough dataset on the oil market. This covers historical daily closing crude oil prices, trading volumes, and market indexes. These data points will form the foundation of our model, providing the market backdrop within which the firm functions and makes trading decisions.

Understanding how the corporation reacts to price swings in oil could provide useful insights into their decision-making processes and potential cognitive biases at work."
Concentrating on a specific market, such as oil, provides for a more precise investigation of key elements influencing that market. Geopolitical events, production levels, environmental rules, and so on are examples of such influences. A more focused strategy could lead to more robust and useful findings about the firm's trading conduct.

Data on Investor Behavior: In our suggested model, I will operate as the decision-making agent interacting with the environment, in this case, the financial market. I will make financial investing decisions based on the information supplied.
As an observer, the reinforcement learning system will examine my decision-making process. It will attempt to detect patterns, tendencies, or biases in my decisions in order to discover any cognitive biases that may influence my decisions.
For example, if my actions regularly indicate a specific cognitive bias, such as loss aversion (the strong preference to avoid losses over collecting equivalent gains), the algorithm will take note. The algorithm's purpose is to learn from my decisions and identify the cognitive biases that influence them.
A reward function will direct the algorithm's learning process. This function will assess the quality of my decisions, perhaps penalizing behaviors that correspond to known cognitive biases, so 'training' the algorithm to detect such biases.

The computer should get more skilled at spotting cognitive biases in my financial decision-making over time as a result of iterative decision-making and learning. The algorithm's ultimate purpose is to be a useful tool for self-reflection and progress, allowing me to acquire insights into my decision-making process and the cognitive biases that influence it.

This research combines behavioral finance and machine learning in an intriguing way. It has the potential to provide a deeper knowledge of how cognitive biases influence my personal financial decisions, as well as a path for minimizing these influences for more reasonable, objective investing selections.
However, it is necessary to recognize the difficulty of such a task. Designing a reinforcement learning model that recognizes cognitive biases effectively and reliably would necessitate substantial fine-tuning and modification. A continuous pattern of decision-making on my side will also be required to build a solid basis for the learning process.

2. Dataset Research: The financial market data was obtained from recognized financial databases such as Quandl and Yahoo Finance, confirming the accuracy and trustworthiness of the information. The trading data, which is an essential component of our research, was provided directly by the trading firm, providing us with significant insights into their decision-making process.

I plan to use data from the following sources for my research:
https://data.nasdaq.com/data/CHRIS-wiki-continuous-futures?keyword=oil%20
https://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm
https://finance.yahoo.com/quote/CL%3DF/history/

3. Dataset Quality Assessment:
The datasets collected have been properly checked for quality. The consistency of financial data across time has been validated, and trading data will be evaluated for completeness and trustworthiness. We made certain that the trading data accurately represented the firm's decisions and was not influenced by outside sources.

4. Data Gathering and Aggregation:
The datasets were collected and integrated into a central database once they had been vetted. We will ensure that they are arranged in a way that is beneficial to our project, allowing for easy access and analysis for later steps, particularly the training of the reinforcement learning model.

5. Revision and Iteration:
Following a first evaluation of the obtained data, we discovered that it is well aligned with the needs of our project. The financial data provides a solid market backdrop, while the trade data provides a detailed record of the firm's decisions, making it an ideal dataset for our reinforcement learning model.
Our next step will be to train the reinforcement learning model utilizing this data. The model will compare the firm's trading data to market data in order to learn from the firm's trading decisions, successes, and failures. The model should be able to recognize trends in the firm's behavior across time, and these patterns should be linked to known cognitive biases.
In short, the reinforcement learning model will "learn" the cognitive biases affecting the firm's trading decisions, providing us with a new perspective on the irrationalities that may be driving their investment behavior. As a result, the information we've gathered will be useful in attaining our project's core goal: developing a cognitive bias-informed reinforcement learning model to assess and optimize trading decisions.

6. Data Cleansing

Data cleansing is an important step in any data analysis effort, and ours was no different. Here's an outline of the data cleansing process we went through, as well as the insights we gained and the challenges we faced:
Duplicate Entries and Insignificant Data: Our initial step in data cleaning was to delete duplicate entries and data items that were irrelevant to our investigation. This required a careful evaluation of our dataset and the use of appropriate filters to exclude these entries.

Missing Values: We realized that certain crucial data points in our initial dataset were missing. For example, several trading days lacked missing volume or oil price data. We employed various techniques to handle issue, depending on the nature and quantity of missing data. To avoid drawing incorrect results, we eliminated entries with significant missing data from our dataset. We employed imputation approaches, such as substituting missing values with the mean or median of the appropriate data column, for minor occurrences.

Normalization of Data: Given the volatility of oil prices over the years, as well as the multiple external factors impacting these prices, it was critical to normalize our data. We were able to compare values on a consistent scale as a result of this. We used approaches like Min-Max normalization to ensure that outliers and big variances in raw data did not have an undue influence on our algorithm.

Data Consistency: We also discovered anomalies in data reporting, such as variances in date formats. This presented a problem because conflicting data can lead to inaccuracies during the analysis stage. To maintain data consistency, we standardized these entries.

Data Verification: Finally, we needed to confirm that the information gathered from multiple sources was correct and dependable. This entailed cross-referencing our data with other trustworthy sources to look for discrepancies.

In terms of what was uncovered, the data cleaning procedure showed the complexities of the trading firm's actions as well as the dynamics of the oil market. While certain trends and behaviors were visible in the raw data, the cleaned data provided a much clearer picture, allowing for more precise and trustworthy analysis. The difficulty was dealing with the massive volumes of data and ensuring that all discrepancies were found and corrected without introducing bias or inaccuracies into the dataset.
This difficult approach provided us with a clean, dependable dataset that is now ready for in-depth analysis and model training in the following stages of our project.