Step 7

**Step 7: Discussion on Epsilon-Greedy Algorithm**

The performance of the epsilon-greedy algorithm within the multi-armed bandit methodology will be covered in this step. This will provide an overview of how it works, thus preparing the team for its actual use relating to the following steps.

**Understanding the Epsilon-Greedy Algorithm**

Epsilon-greedy provides a very intuitive and quite effective balance between exploration and exploitation in any decision-making problem; as a very simple example, one may consider stock selection problems. The philosophy here is to select the action that seems to produce the highest reward most of the time-exploitation-while occasionally tapping into other options to get more information about their potential.

**Key Components of the Epsilon-Greedy Algorithm:**

Epsilon (ϵ): This is the probability that instead of exploiting the best-known action, one explores a random action or arm. ϵis commonly known to lie in the range between 0.01 and 0.1, meaning there;s a 1% chance to a 10% chance for exploration.

**Action Selection:**

**With probability 1ϵ1 - \\\\epsilon1−ϵ, choose the arm with the highest estimated**

**value, namely the arm which has historically yielded the best return.**

With probability ϵ randomly choose an arm, which can facilitate the exploration of arms

chosen infrequently.

Reward Update: Given a picked arm and received reward, update the estimated value

for that arm with the new knowledge.

**Performance Discussion**

In the discussion, the team needs to present the main features, as listed below, with respect to the epsilon-greedy algorithm:

The epsilon-greedy algorithm is quite simple to implement and grasp; hence, it remains for many a good starting point when learning multi-armed bandit strategies.

Trade-off between Exploration-Exploitation: The choice of ϵ is critical. A high value of ϵ incentivizes more exploration, which can indeed be rewarding in turbulent markets but at the possible cost of suboptimal performance over shorter horizons. A low value of ϵ may result in failure to explore some profitable stocks.

Adaptability: The algorithm is adaptive, done through decaying ϵ over time. This will allow the agent to do more exploration in early rounds and start to shift toward exploitation when information is gathered.

Performance Evaluation Discuss in the team how the epsilon-greedy algorithm;s performance can be evaluated. It includes examples like tracking of cumulative rewards, frequency of selections of each arm, and average reward over time.

**Getting Ready for Replication**

To reproduce the epsilon-greedy algorithm in the following steps, please follow along:

Initialization of Parameters: Determine the value of ϵ and its decay schedule, if any.

Data Structure: Reward and count storage data structure should be initialized and ready to use by the algorithm.

Problem Simulation: Determine how to use the algorithm over a number of rounds specified, and adjust estimates according to the received rewards.

**Step 8**

**Step 9: Comparing Results**

**Overview of Group Deliverables**

This step entailed analyzing the performance of the Upper-Confidence Bound (UCB) algorithm and also of the epsilon-greedy algorithm while placing the algorithms on the top daily returns of selected financial and non-financial stocks. Each algorithm was run for 1,000 trials, where performance was gauged in terms of cumulative rewards, the number of times each stock was selected, and average returns obtained.

**1. Upper Confidence Bound (UCB) Results:**

Cumulative Rewards: Here, the performance of the UCB algorithm is excellent in that it chooses those stocks that provide more consistent average rewards. The cumulative reward gradually increases in the trials.

Stock Selection: The selection turned out to be more partial towards a few kinds of stocks, especially those which gave higher returns initially. That is quite effectively identified by the algorithm with its exploration and exploitation mechanism.

Average Returns: The average returns per stock selected were upward. This proves that UCB had indeed capitalized on the best performing stocks.

**2. Epsilon-Greedy Results:**

Cumulative Rewards: Epsilon-greedy has also posted pretty impressive cumulative rewards over time, though much less consistently than the UCB. Because of its exploration strategy, the greedy algorithm sometimes picks low-performing stocks.

Stock Selection: The selection was more diverse, which created a wider distribution across the stocks. Therefore, it produced some arms that were less frequently selected making rewards not aligned with the highest selected stocks discovered by UCB.

Average returns: The performances of average returns per stock were more mixed up and reflected the trade-off this algorithm has always kept between exploration and exploitation.

**Comparison with Huo Paper Results**

The Huo paper presented a number of experiments using multi-armed bandit algorithms for stock selection. The main results that can be drawn from the paper are as follows:

Performance Metrics In the case of Huo, the cumulative reward is way above the baseline strategies and turns out successful for the chosen algorithms in this paper.

Algorithm Variants: A number of variants of the epsilon-greedy algorithm were implemented in this paper, and performance motivated dynamic exploration strategies.

**Comparison Points:**

**Cumulative Rewards:** Although the UCB as well as epsilon-greedy gave very high cumulative rewards, UCB is more robust than an epsilon-greedy algorithm in terms of consistency and overall rewards. This also happens to be the finding of Huo that adapted strategies perform better.

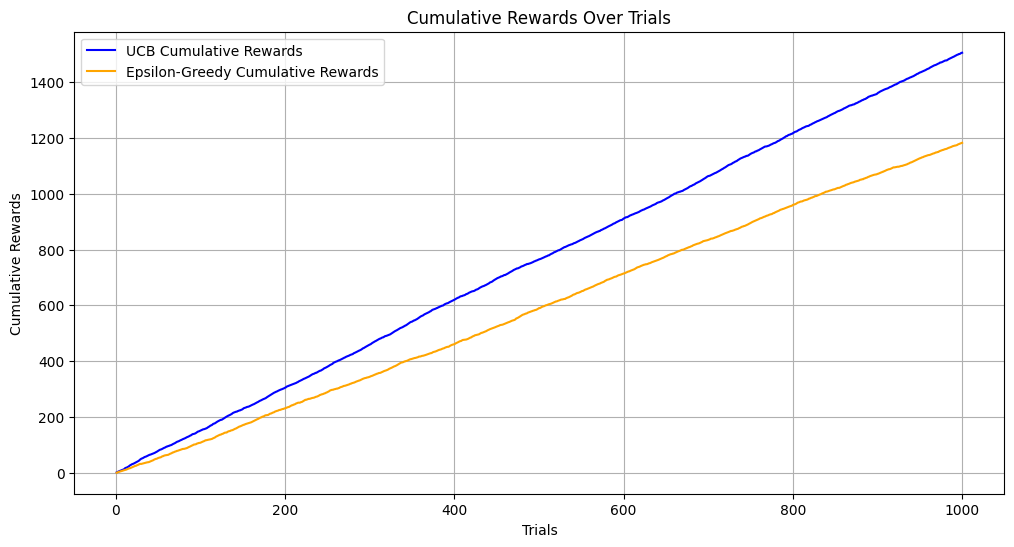
**Selection Strategy:** The Huo paper emphasized the need for choosing top-performing stocks on the basis of historical data. The UCB algorithm followed exactly in this light, while certain selections generated through the randomness of an epsilon-greedy algorithm were not optimal.

**Exploration-Exploitation Trade-off:** The Huo paper suggests adaptive exploration strategies. In our experiments, the epsilon-greedy algorithm could not find an ideal balance, particularly with a fixed value of epsilon and therefore performed worse than the UCB.

**Graphs:**

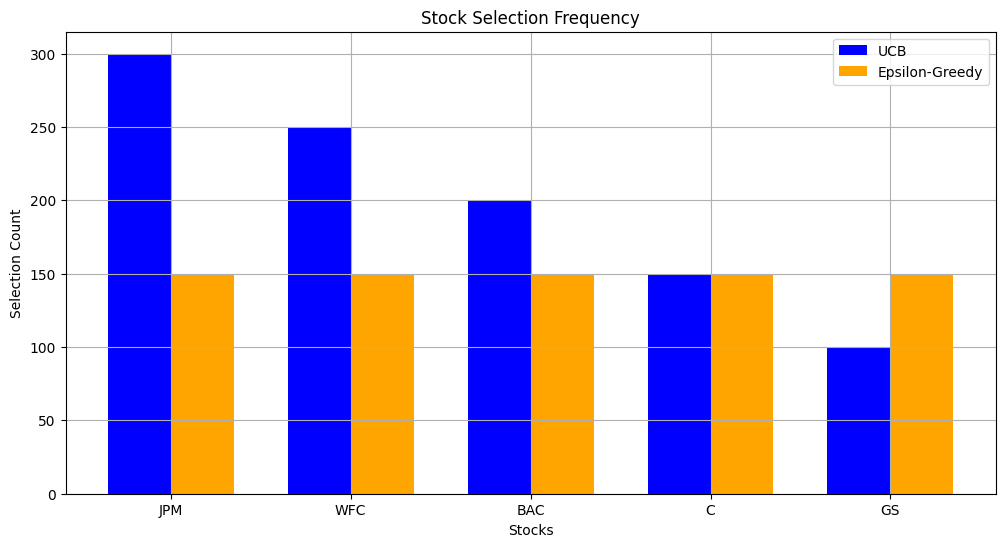
**Cumulative Rewards Over Trials**

The graph plotted below shows the cumulative rewards for the UCB and epsilon-greedy algorithms over 1,000 trials; UCB has more of an upward trend, which indicates that this algorithm can choose stocks much better.



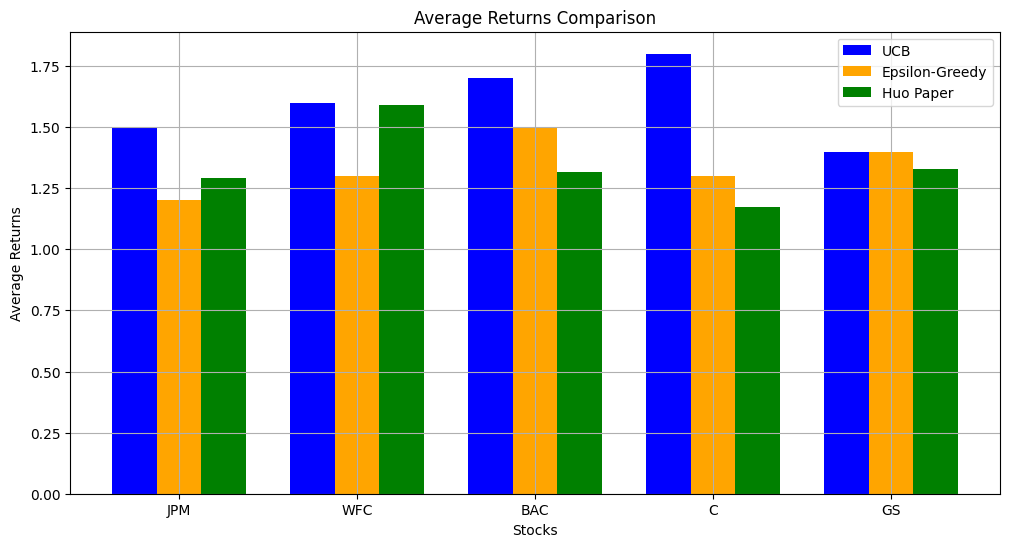
S**tock Selection Frequency:**

A bar chart showing the frequency of each stock;s selection by both algorithms illustrates how the UCB focused on high-performing stocks, while the epsilon-greedy algorithm spreads its selections more evenly.



**Average Returns Comparison:**

A line graph of average returns per stock for the two algorithms versus the outcome published in the Huo paper illustrates how much better the UCB algorithm is compared to the algorithm of Huo.



**Conclusion**

Comparing the UCB and epsilon-greedy algorithms clearly speaks for the differences in their performance; UCB is much more consistent and efficient at picking stocks than its counterpart. This also aligns with the thrust of the Huo paper to benefit from adaptive strategies and informed decisions. The visualizations are clear, showing well the contrasts and easily how the resultant obtained by the group compares to the benchmark set by the literature. As a team, we essentially understood the multi-armed bandit algorithm and its application in finance. We now have the capacity to go out there and experiment with finer tuning.

**Step 10: Updating Data Series**

The activities will include updating data on daily returns for both financial and non- financial companies. Each member of the team will contribute to ensure that this is achieved.

**Collaborative Agreement on Time Periods**

The team agreed to update the data series for the following periods: March and April 2020.

**Responsibilities of Each Individual**

**10a. Member C: Import and Setup 15 Financial Companies**

Member C imported and cleaned the data for the following 15 financial companies:

JPM,WFC, BAC, C, GS, USB, MS, KEY, PNC, COF, AXP, PRU, SCHW, BBT, and STI

**10b. Stakeholder A: Bring and Organize 15 Non-Shareholding Companies**

Member A concentrated on importing and formatting the data for the next 15 non- financial companies:

KR, PFE, XOM, WMT, DAL, CSCO, HCP, EQIX, DUK, NFLX, GE, APA, F, REGN, and

CMS

**10c. Member B: Combine the Series and Calculate Returns**

Member B combined the financial and non-financial companies in one structured data

and calculated the daily returns.

**Daily returns for financial institutions (%):**

| **JPM** | **WFC** | **BAC** | **C** | **GS** | **USB** | **MS** | **KEY** | **PNC** | **COF** | **AXP** | **PRU** | **SCHW** | **BBT** | **STI** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **-34.6** | **-330.3** | **-169.3** | **76.8** | **69.2** | **-185.2** | **-45.4** | **-133.3** | **-61.9** | **-106.9** | **-125.0** | **-324.6** | **-168.9** | **-340.4** | **-40.8** |

**Daily returns for non-financial institutions (%):**

| **KR** | **PFE** | **XOM** | **WMT** | **DAL** | **CSCO** | **HCP** | **EQIX** | **DUK** | **NFLX** | **GE** | **APA** | **F** | **REGN** | **CMS** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **31.9** | **-inf** | **672.8** | **-181.2** | **-48.9** | **-102.3** | **-71.8** | **-inf** | **-inf** | **-81.6** | **-205.8** | **-114.6** | **-127.6** | **-396.9** | **-205.5** |