Week 5

**WEEK 5 – MULTI-ARMED BANDITS – AD OPTIMIZATION**

Consider the dataset **"Ads\_Clicks,"** which contains information about user interactions with advertisements over time. An advertising company is running **18 different ads** on a webpage, all targeted toward a similar audience. The dataset records whether a user clicked at a given time step. Each column corresponds to a specific ad, where YES(**1) indicates that the ad was clicked**, and NO(**0) indicates that it was not**. Consider the attached csv fi each adle.

1. Define the multi-armed bandit (MAB) problem in the context of ad optimization, considering how an agent selects among multiple ads to maximize clicks.
2. How does the exploration-exploitation trade-off influence decision-making in this scenario?
3. Implement the ε-greedy algorithm to optimize ad selection and compute the total rewards after **2000-time steps** for:**ε = 0.05** and **ε = 0.2**
4. Compare the effect of different ε values on total rewards and action selection.
5. Implement the UCB method with an exploration factor **c = 2.0** and compute total rewards after **2000-time steps**.
6. How does increasing or decreasing the exploration factor **c** affect the performance?
7. Analyze how the estimated action values (Q-values) compare to the actual optimal action in both **ε-greedy** and **UCB** methods.
8. Which approach leads to a better approximation of the optimal action?
9. Evaluate how the performance of **ε-greedy** and **UCB** changes when the time horizon is extended to **5000-time steps** instead of **2000-time steps**.
10. Does a longer time horizon reduce the impact of exploration parameters (**ε or c**) on total rewards?