Overview

We have a set multi-armed bandits inspired by real data (obtained from our Dynamic Creative Optimization system).

Stochastic with reward either 0 or 1 (Bernoulli bandit)   +

Budget-limited (from 500 up to 30000 pulls)

10 stationary cases

10 non-stationary cases (multiple segments with constant probability)

Practical Work

1. **Download the code**, test it executes successfully
2. Implement and measure the performance of bandit policies:
   a. Epsilon-greedy
   b. UCB1
   c. Thompson sampling
   d. (advanced) UCB-Tuned
3. Test different algorithms' parameters
4. Own ideas for further improvement?

Use **Main.py** to run and configure your experiment

Add your bandit policies to **Agent.py**