Aleksei Evgrafov

F29AI

H00427637

Question 6 (1 point)

Q-Learning Agent against each of the provided agents 50 times

Against Defensive Agent:

Wins: 43 Losses: 0 Draws: 7

Against Aggressive Agent:

Wins: 50 Losses: 0 Draws: 0

Against Random Agent:

Wins: 50 Losses: 0 Draws: 0

In this part the train() method implements the Q-learning algorithm to iteratively improve the agent's decision-making in this environment. The agent explores actions using an epsilon-greedy strategy updates Q-values based on rewards and discounted future values using the Q-learning update rule, and gradually shifts from exploration to exploitation by decaying the epsilon parameter. After training, the extractPolicy() method generates an optimal policy by selecting the action with the highest Q-value for each state in the Q-table.

Screenshots and live testing you can see in video.