Analisis dan Diskusi

Pada percobaan ini digunakan algoritma *Deep Q-Network* (DQN) untuk mengendalikan *environment CartPole*. Jumlah episode pelatihan yang seharusnya 1000 dikurangi menjadi 100 agar proses lebih cepat, sehingga performa agen masih beragam dan belum stabil optimal. Pada 1000 episode, agen berpeluang besar mencapai strategi optimal sehingga nilai *reward* mendekati maksimum (500). Namun dengan 100 episode, agen belum sepenuhnya mengeksplorasi semua kemungkinan strategi, sehingga ada episode uji coba yang berhenti lebih cepat. Nilai *gamma* (γ) mempengaruhi fokus agen terhadap *reward* jangka panjang, nilai *epsilon* (ε) mengatur keseimbangan eksplorasi dan eksploitasi, sedangkan *learning rate* menentukan seberapa cepat bobot jaringan saraf diperbarui. Keseimbangan eksplorasi dan eksploitasi menjadi kunci agar agen tidak hanya mencoba strategi baru, tetapi juga mengeksploitasi strategi terbaik yang sudah dipelajari.

Berbeda dengan *supervised learning* yang menggunakan data berlabel, *reinforcement learning* mengandalkan *feedback* berupa *reward* dari interaksi dengan lingkungan. Tantangan utama adalah menyesuaikan parameter dan jumlah episode agar agen benar-benar mampu mencapai *reward* maksimum. Meskipun sederhana, percobaan CartPole ini mencerminkan potensi RL dalam sistem kendali nyata, seperti robotika, transportasi rel, kendaraan otonom, maupun optimasi proses industri.