ABSTRAK

Optimisasi merupakan prosedur untuk mencari penyelesaian terbaik dari sekian pilihan yang tersedia dengan mempertimbangkan batasan masalah dan fungsi tujuan yang telah ditentukan. Salah satu permasalahan optimisasi klasik adalah *Travelling Salesman Problem* (TSP), yang bertujuan mencari rute terpendek dengan mengunjungi setiap kota tepat satu kali sebelum kembali ke titik awal. Pendekatan eksak seperti *Branch and Bound* dan *Dynamic Programming* mampu menghasilkan solusi optimal, namun memiliki kompleksitas eksponensial yang membuatnya tidak praktis untuk jumlah kota yang besar. Penelitian ini mengimplementasikan algoritma *Q-Learning*, salah satu pendekatan *Reinforcement Learning*, untuk menyelesaikan TSP. Agen dimodelkan sebagai *Travelling Salesman* dalam kerangka *Markov Decision Process* dengan representasi *state* sederhana berupa posisi kota saat ini. Kinerja *Q-Learning* dibandingkan dengan algoritma *Held-Karp* dan *Branch and Bound* menggunakan dua metrik unjuk kerja, yaitu waktu komputasi dan kualitas solusi. Hasil simulasi menunjukkan bahwa *Q-Learning* mampu menyelesaikan TSP hingga 50 kota dengan waktu komputasi yang relatif stabil dan cenderung linear, sementara algoritma eksak mengalami lonjakan eksponensial dan tidak lagi praktis di atas 20-an kota. Dari sisi kualitas solusi, *Evaluation Average Q-Learning* berada sekitar 10–19% di atas solusi optimal. Namun pada kasus terbaik, selisihnya dapat menyempit hingga kurang dari 3%.

Kata Kunci: *Reinforcement Learning, Q-Learning, Travelling Salesman Problem*

ABSTRACT

Optimization is a procedure for finding the best possible solution among available alternatives, subject to defined problem constraints and objective functions. One classical optimization problem is the Travelling Salesman Problem (TSP), which seeks the shortest route that visits every city exactly once before returning to the starting point. Exact approaches such as Branch and Bound and Dynamic Programming yield optimal solutions but suffer from exponential complexity, rendering them impractical for large instances. This study implements Q-Learning, a Reinforcement Learning algorithm, to solve the TSP. The agent is modelled as a Travelling Salesman within a Markov Decision Process framework using a simplified state representation of the agent's current city position. Q-Learning performance is compared against the Held-Karp and Branch and Bound algorithms using two performance metrics: computation time and solution quality. Simulation results demonstrate that Q-Learning solves TSP instances of up to 50 cities with relatively stable, near-linear computation time, whereas exact algorithms exhibit exponential growth and become impractical beyond approximately 20 cities. In terms of solution quality, the Q-Learning Evaluation Average falls roughly 10–19% above the optimal solution. However, in the best-case episode, the gap narrows to under 3%.

Keywords: Reinforcement Learning, Q-Learning, Travelling Salesman Problem