**Abstract**

This report investigates the application of Monte Carlo, Value Iteration, and Q-Learning algorithms to solve a Markov Decision Process (MDP) that simulates decision-making in a student's life. The study aims to evaluate the efficiency, convergence, and effectiveness of these reinforcement learning techniques in managing real-life decision processes characterized by stochastic transitions and rewards. The findings demonstrate varying degrees of success across the algorithms, providing insights into their applicability for different types of decision-making scenarios.

**Objective:** The primary objective is to implement and analyze three reinforcement learning algorithms—Monte Carlo, Value Iteration, and Q-Learning—to determine the optimal decision-making policy in the context of a student's life.

**Methodology**

**MDP Model Description**

* **States**: Rested, Tired, Homework Done, Homework Undone, 8pm (terminal state).
* **Actions**: Party (P), Rest (R), Study (S).
* **Rewards**: Defined for actions in specific states (e.g., partying may yield a positive reward for social satisfaction but a penalty for reduced study time).

**Algorithms Implemented**

* **Monte Carlo Method**:
  + Implemented a first-visit Monte Carlo simulation with an equiprobable policy over 50 episodes. The state values were updated based on the rewards accumulated from each episode.
* **Value Iteration**:
  + Conducted value iteration using a discount factor of 0.99 until the maximum change in value estimates was less than 0.001. The policy was then derived from these value estimates.
* **Q-Learning**:
  + Utilized a model-free Q-learning approach with an initial learning rate of 0.2, employing an epsilon-greedy policy for action selection. The learning rate decayed by a factor of 0.995 per episode until the Q-values converged.

**Results**

**Simulation Outcomes**

* **Monte Carlo**:
  + The Monte Carlo method provided a policy that balanced the immediate rewards of actions, leading to a moderate understanding of state values.
* **Value Iteration**:
  + Value iteration rapidly converged to an optimal policy, demonstrating high efficiency in this context with clear and stable state value estimations.
* **Q-Learning**:
  + Q-learning adapted effectively over iterations, showing gradual improvement and convergence to an optimal policy, albeit slower than value iteration.

**Performance Analysis**

* **Computational Efficiency**: Value Iteration was the most computationally efficient.
* **Convergence**: Value Iteration and Q-Learning showed definitive convergence, with the former being faster.
* **Policy Reliability**: Policies derived from Value Iteration and Q-Learning were more reliable and consistent compared to Monte Carlo.

**Discussion**

**Interpretation of Results**

The study indicates that while Monte Carlo methods are useful for estimating policy outcomes without requiring prior knowledge of state transitions, methods like Value Iteration and Q-Learning provide more precise and reliable policy recommendations in structured environments.

**Limitations and Assumptions**

* The simulation assumes perfect model knowledge for Value Iteration and deterministic transitions for simplification, which might not hold in more complex real-life scenarios.

**Conclusion**

The comparative analysis reveals that Value Iteration and Q-Learning are more suited for scenarios where accurate and reliable decision-making is critical. Future research could explore hybrid models combining the exploratory benefits of Monte Carlo with the stability of Q-learning or Value Iteration.

**References**

**Appendices**

* **Code Snippets**: Included in the supplementary materials.
* **Additional Data**: Charts and tables representing the evolution of state values and policy efficacy over iterations.