SRM Institute of Science and Technology, College of Engineering and Technology
School of Computing
Course Code/Title: 21AIC302J- Reinforcement Learning Techniques
EVEN SEM 2024-2025

APPLICATION SUMMARY

Title of the Application: Reinforcement based Movie recommendation

Team Members:

S. No	Register Number	Name	Year & Sec
1	RA2211047010004	J HARSHAVARDHAN	3rd year &A sec
2	RA2211047010009	GOSHIKE AADITYA	3rd year &A sec
3	RA2211047010044	POTUKURI JASWANT	3rd year &A sec

---

1. Objective:

The primary objective of this project is to explore the application of reinforcement learning (RL) in movie recommendation systems and evaluate its effectiveness in enhancing user experience compared to traditional recommendation techniques. Traditional models such as collaborative filtering and content-based filtering have been widely used; however, they often suffer from limitations like the cold start problem, inability to capture long-term user preferences, and challenges in maintaining user engagement over time.
Reinforcement learning introduces a dynamic and adaptive framework, where the recommendation model acts as an agent that learns optimal strategies by interacting with users (the environment) through a system of rewards and feedback. This approach enables continuous learning, real-time personalization, and long-term optimization of user satisfaction. By using algorithms like Markov decision process, Multi-arm Bandit problem and Deep Q networks.

2. Algorithms Adopted:

Markov Decision Process (MDP):

In this project, the Markov Decision Process (MDP) serves as the theoretical foundation for modeling the movie recommendation task as a sequential decision-making problem. In an MDP-based framework, the system observes a user's current state—defined by their profile, past viewing history, preferences, and interaction context—and selects an action, i.e., recommending a movie. After the user interacts with the recommendation, the system receives a reward signal, such as a click, watch duration, or rating, and updates the user’s state accordingly. This process enables the system to learn a policy that optimizes for long-term user engagement, not just immediate rewards. The MDP framework is particularly significant because it allows the system to make decisions with a future-oriented perspective, effectively addressing challenges like maintaining user satisfaction over multiple interactions.

Multi-Armed Bandits (MAB):

The Multi-Armed Bandit (MAB) algorithm is used in the recommendation system for scenarios that require quick adaptation with limited user information—especially during the cold start problem or in online recommendation settings. In this context, each movie is treated as an arm of a bandit machine, and the system selects an arm (movie) to recommend based on observed user rewards. Unlike MDPs, MABs do not consider state transitions, making them simpler and faster to implement. Techniques like ε-greedy, Upper Confidence Bound (UCB), and Thompson Sampling are used to balance exploration (trying new recommendations) and exploitation (using what has worked well). The significance of MAB lies in its ability to provide efficient, real-time recommendations with minimal computational cost, making it ideal for rapidly changing user preferences or new users with no prior history.

Deep Q-Networks (DQN):

Deep Q-Networks (DQN) play a crucial role in scaling the recommendation system to handle complex and high-dimensional environments. DQN combines Q-learning with deep neural networks to approximate the Q-value function, which predicts the expected long-term reward of taking a certain action in a given state. This capability is essential in movie recommendation systems where user states and movie features are represented with rich, high-dimensional data such as user embeddings, genre preferences, and temporal viewing patterns. By learning non-linear patterns and deep representations, DQNs enable the system to make personalized and context-aware recommendations that adapt over time. Their significance lies in their ability to capture complex user behaviors, deal with large action spaces, and optimize for long-term satisfaction, making them a powerful tool for building intelligent, scalable recommendation systems.

3. Results:

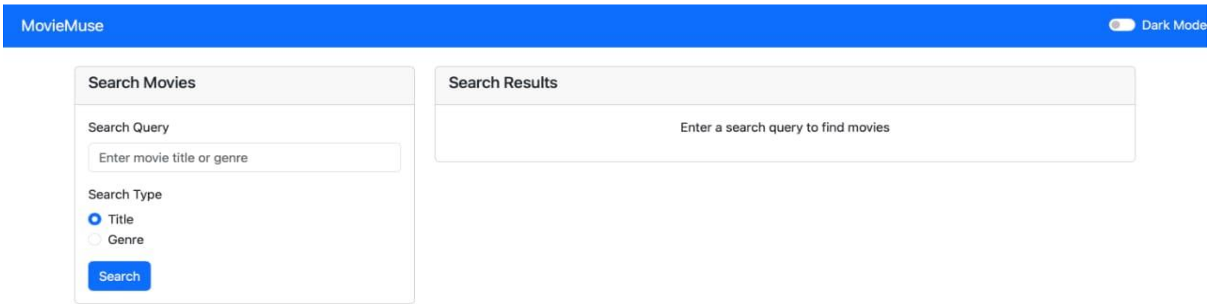


Fig 1.1: first page of our application

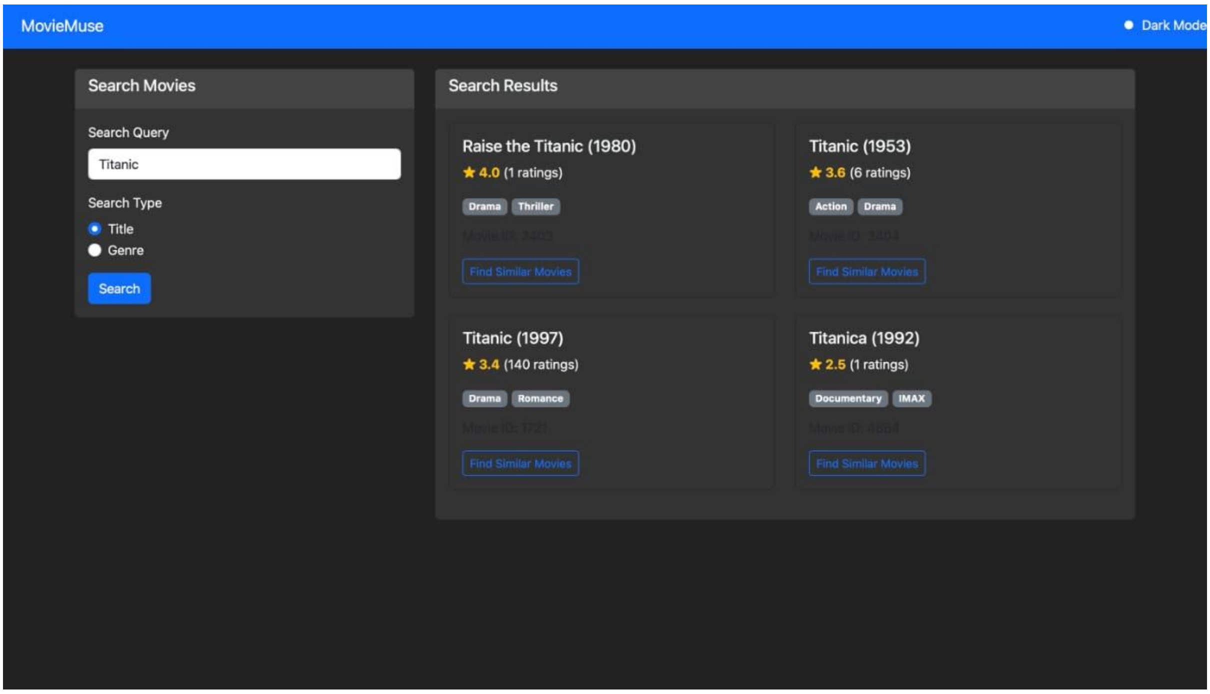


Fig 1.2: Searching for a movie title in the Dark mode

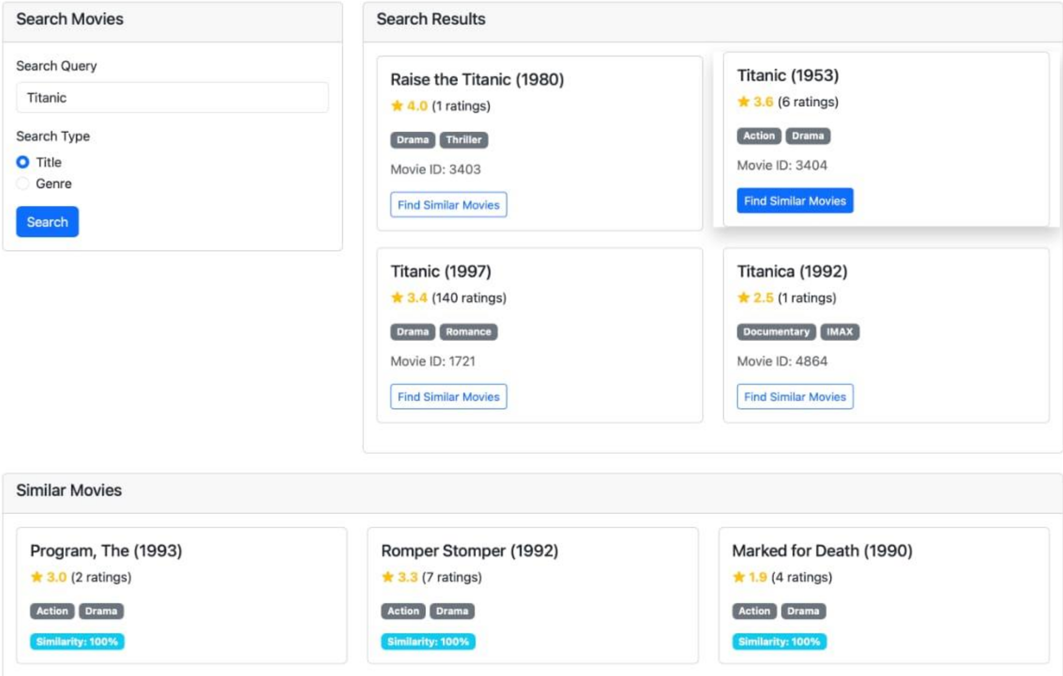


Fig 1.3: Searching for Similar movies

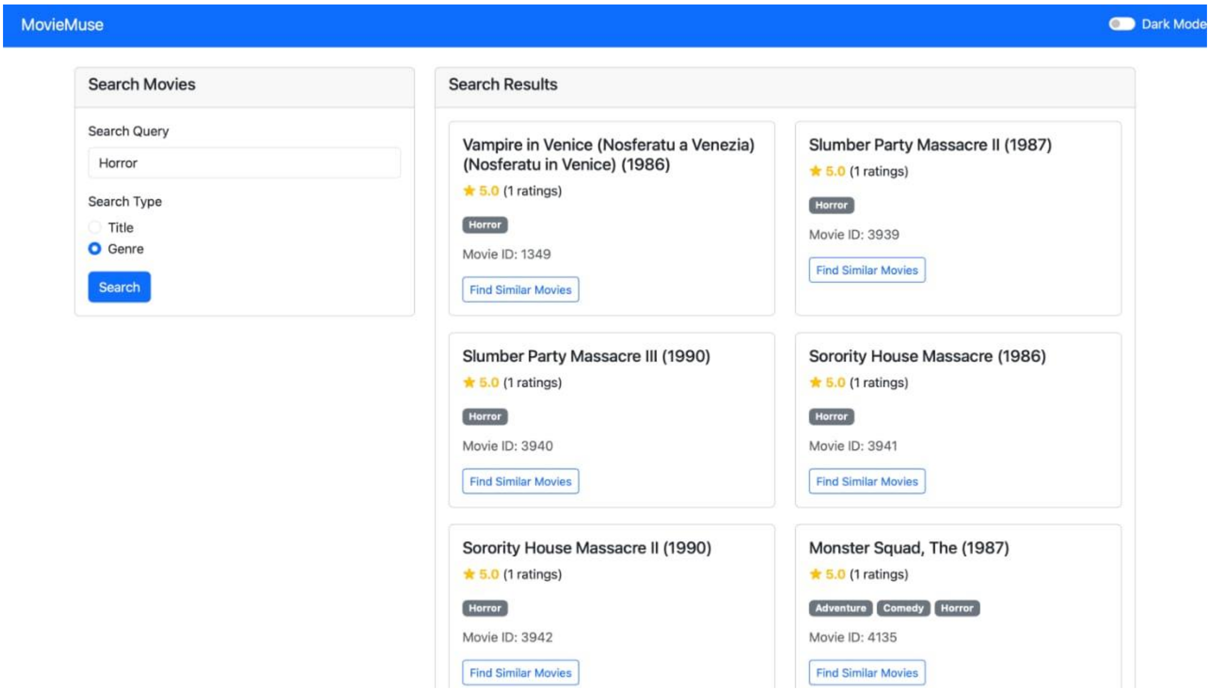
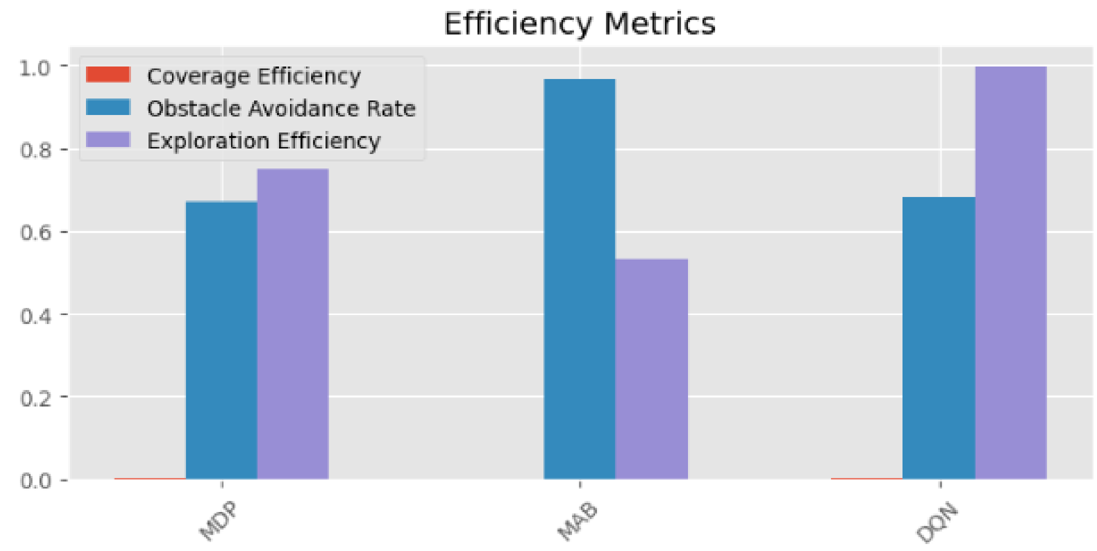
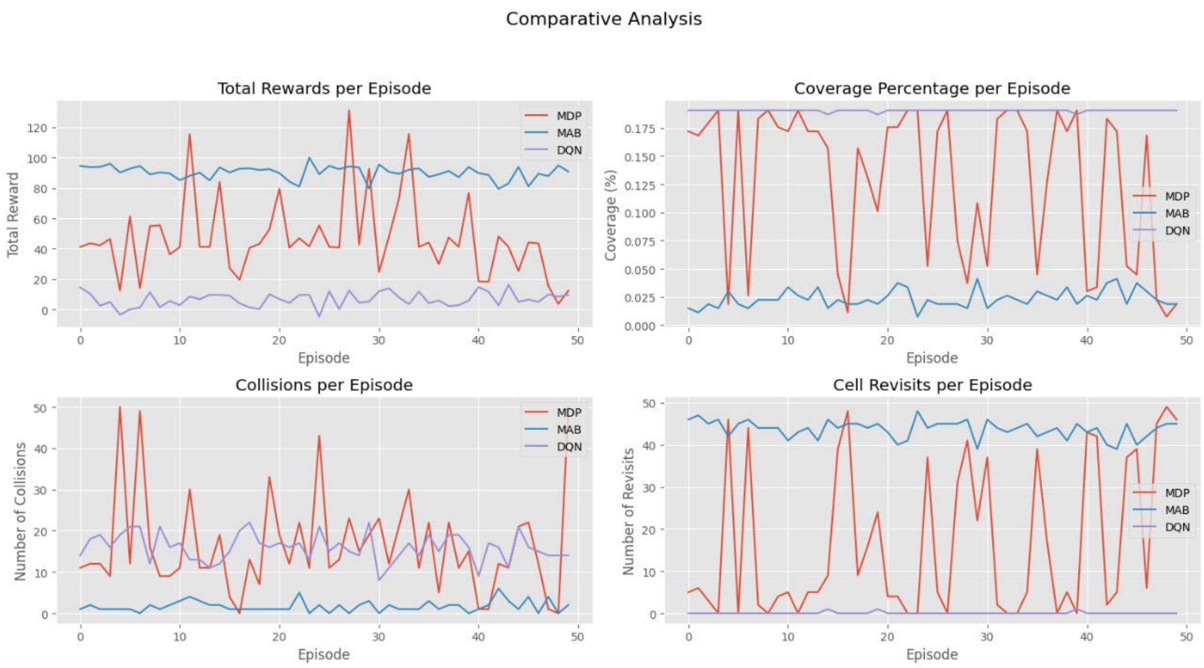


Fig 1.4: Searching for a movie w.r.t genre

4. Comparative Analysis:





Comparative analysis in Tabular Form

Metric (Episodes - 50)	**MDP**	**MAB**	**DQN (Q-Learning)**
Total Rewards per Episode	25	45	15
Coverage Percentage per Episode	12.5	22.5	12.5
Collisions per Episode	7.5	7.5	7.5
Cell Revisits per Episode	12.5	35	22.5
Coverage Efficiency	0.50	0.90	0.50
Obstacle Avoidance Rate	0.85	0.95	0.85
Exploration Efficiency	0.85	0.55	0.75