MACHINE LEARNING ANALYSIS REPORT

TEAM NO: 15

SRN1: PES2UG23CS334

Medha Raj

SRN2: PES2UG23CS375

Navyata Venkatesh

SRN3: PES2UG23CS378

Neha Harish

KEY OBSERVATIONS:

The most challenging part was integrating the probabilistic reasoning of the HMM with the decision-making process of the RL agent. Handling varying word lengths and designing an HMM that could generalize across all patterns required careful preprocessing and state definition. Another challenge was designing a balanced reward system that encouraged both exploration and accuracy while avoiding repeated guesses.

STRATEGIES:

The HMM was designed with (position, previous_letter) as hidden states and letters (A–Z) as emissions to capture sequential and contextual dependencies such as “QU” or “ING.” Separate HMMs for each word length were trained to model structural differences without padding, and Laplace smoothing ensured nonzero probabilities for unseen combinations. A pattern-matching component was also integrated to refine predictions by combining corpus-based candidate frequencies with HMM probabilities. The RL agent used six normalized features (word length, revealed ratio, hidden ratio, guessed count, lives remaining, and pattern difficulty) as the state representation, allowing generalization through linear function approximation rather than tabular Q-learning. Actions corresponded to unguessed letters, and learning employed Q-learning with experience replay and an ε-greedy exploration strategy. The reward function awarded +2 for each correctly revealed letter, −5 for wrong guesses, +20 for completing the word, and −10 for losing, promoting accuracy and efficiency while slightly discouraging risky exploration.

EXPLORATION:

The exploration–exploitation balance was handled using an ε-greedy strategy with exponential decay. At the start, the agent explored with 10% random guesses ($\varepsilon=0.1$) to learn different possibilities, and this value gradually decreased using $\varepsilon = \max(0.01, 0.1 \times 0.995^{\text{episode}})$ so that the agent relied more on what it had learned over time. A small 1% exploration rate was kept to ensure occasional random guesses even after training. During exploitation, the agent chose the letter with the highest combined score from HMM probabilities and Q-values. This approach was simple and effective, though the decay happened slightly too fast, reducing exploration early and sometimes causing the agent to settle for less optimal strategies.

FUTURE IMPROVEMENTS:

If given another week, the agent could be improved in several ways. The state representation could be expanded to include the full HMM probability distribution and one-hot encodings of revealed letters, giving the model richer contextual awareness. The HMM–RL weighting could be made adaptive, giving more importance to HMM predictions when pattern matches are strong and relying more on RL when uncertainty is high. The reward function could be reshaped to be less punitive and more informative, rewarding letters that provide greater information gain or reveal vowels early. The exploration schedule could decay more slowly and sample letters proportionally to their HMM probabilities instead of uniformly, encouraging smarter exploration. Finally, using a Deep Q-Network (DQN) instead of linear Q-learning could help capture non-linear relationships in state features.