

# Report

Submitted by:  
Uday Bhardwaj( 12241910)  
Vedansh Sharma(12242000)

## Problem Statement: CartPole Game using Reinforcement Learning

The problem statement is as follows:

A Pole is attached to a movable cart through one of its ends. The pole can not remain in a balanced position due to an object at its other end. The Target was to make the pole balance the pole as long as possible by moving the cart in a specified direction.

## Approach:

The problem statement was approached using the following methods:

### 1. Creating an environment for the CartPole:

We have used the **Gymnasium** library, which is a maintained fork of OpenAI's gym.

The library has provided the environment for the CartPole game. This enables us to apply our algorithm by understanding the observation space and action space for the environment.

The Observation space consists of 4 features:

- Position of the Cart
- Velocity of the Cart
- Angle of the Pole
- Angular Velocity of the Pole

The Action space consists of 2 options:

- Moving the cart towards right
- Moving the cart towards left

## 2. Converting the Continuous space to Discrete space of the Environment:

The Environment given by the Gymnasium provides continuous space.

This was converted into Discrete Space using the custom functions by providing a limit for the features such as Velocity, Angular Velocity.

The Position and angle have been converted into discrete values. This is explained in detail along with the code snippet in the presentation.

## 3. Applying the Q-learning algorithm:

We have applied a  $\epsilon$ -greedy Exploration policy for Q-learning algorithm. In this, we have used the formula given below:

$$Q(s,a) = Q(s,a) + \alpha( \text{reward} + \gamma \max_{a'}(Q(s',a')) - Q(s,a) )$$

The choice between Exploration and Exploitation is done with respect to the epsilon. The value of epsilon slowly decreases as the number of episodes is increased. This is achieved by applying a decay rate.

This ensures that Exploration is favored for smaller values of Episode number and Exploitation is favored for larger values of Episode number.