COURSE NAME: ARTIFICIAL INTELLIGENCE

COURSE CODE: BTCS602-18

ASSIGNMENT No: 3

DATE OF ALLOTMENT:

DATE OF SUBMISSION:

COURSE INSTRUCTOR:

SUBMITTED BY:

REGISTRATION NO:

CT Group of Institutions, Shahpur			
Assignment No: 3			
Program:	**B Tech CSE**	**Semester:**	**6th**
Name of Subject:	**Artificial Intelligence**	**Subject Code:**	**BTCS 602-18**
Name of Topic:	**Reinforcement Learning**	**Maximum Marks:**	**10**
Date of Allotment:	**22/03/25**	**Date of Submission:**	**26/03/25**

1. Define **Direct Utility Estimation**, **ADP**, and **TD Learning** in the context of passive RL.

2. The immediate rewards received at each location over three trips are:

Museum (M) → Park (P) → Restaurant (R)

Trip	Museum (M)	Park (P)	Restaurant (R)
1	5	6	8
2	4	7	9
3	6	5	7

Given a **discount factor $\gamma = 0.9$**, compute the utility values for each location using the following methods:

(a) Compute the utility for each place using the simple average reward formula:

(b) Use the **Bellman equation** iteratively to estimate the utility values:

(c) Use the **TD update rule** to compute the updated utility values after one trip using a **learning rate $\alpha = 0.5$**.

(d) Compare the computed utility values from all three methods and discuss how they differ. Which method provides the best estimate in the long run?

Prepared by:	Verified by:
Name of Course Instructor: Er. Sukhdayal Dhanday	**Name of HOI: Dr. Sandeep Ranjan**
Signature with Date:	**Signature with Date:**