07-10-2025 10:56:39 AM

REINFORCEMENT LEARNING **Course Code : 316323**

Programme Name/s : **Artificial Intelligence and Machine Learning/ Data Sciences**
Programme Code : **AN/ DS**
Semester : **Sixth**
Course Title : **REINFORCEMENT LEARNING**
Course Code : **316323**

I. RATIONALE

Reinforcement learning enables machines to autonomously learn optimal actions through trial and error, making it vital for dynamic, real-world decision-making. It drives advancements in fields like robotics, gaming, and personalized systems. Students learn how agents make sequential decisions by interacting with environments to maximize cumulative rewards. They explore key concepts like Markov decision processes, Q-learning, policy gradients, and reinforcement learning techniques.

II. INDUSTRY / EMPLOYER EXPECTED OUTCOME

The aim of this course is to help the students to attain the following Industry Identified Outcomes through various teaching learning experiences: Develop Reinforcement Learning (RL) algorithms in real-world AI applications.

III. COURSE LEVEL LEARNING OUTCOMES (COS)

Students will be able to achieve & demonstrate the following COs on completion of course based learning

- CO1 - Explain fundamental concepts of Reinforcement Learning.
- CO2 - Apply Markov Decision Processes to solve RL Problems.
- CO3 - Apply the Bellman Equation and Dynamic Programming (DP) Methods to solve RL Problems.
- CO4 - Apply Monte Carlo (MC) Methods for solving RL Problems.
- CO5 - Select appropriate method between DP,MC,TD to solve RL problem.

IV. TEACHING-LEARNING & ASSESSMENT SCHEME

Course Code	Course Title	Abbr	Course Category/s	Learning Scheme					Credits	Assessment Scheme											
				Actual Contact Hrs./Week			SLH	NLH		Paper Duration	Theory				Based on LL & TL				Based on SL		Total Marks
															Practical						
				CL	TL	LL					FA-TH	SA-TH	Total		FA-PR		SA-PR		SLA		
											Max	Max	Max	Min	Max	Min	Max	Min	Max	Min	
316323	REINFORCEMENT LEARNING	RFL	DSE	3	-	2	1	6	3	3	30	70	100	40	25	10	25#	10	25	10	175

Total IKS Hrs for Sem. : Hrs

Abbreviations: CL- ClassRoom Learning , TL- Tutorial Learning, LL-Laboratory Learning, SLH-Self Learning Hours, NLH-Notional Learning Hours, FA - Formative Assessment, SA -Summative assessment, IKS - Indian Knowledge System, SLA - Self Learning Assessment

Legends: @ Internal Assessment, # External Assessment, *# On Line Examination , @$ Internal Online Examination

Note :

1. FA-TH represents average of two class tests of 30 marks each conducted during the semester.
2. If candidate is not securing minimum passing marks in FA-PR of any course then the candidate shall be declared as "Detained" in that semester.
3. If candidate is not securing minimum passing marks in SLA of any course then the candidate shall be declared as fail and will have to repeat and resubmit SLA work.
4. Notional Learning hours for the semester are (CL+LL+TL+SL)hrs.* 15 Weeks
5. 1 credit is equivalent to 30 Notional hrs.
6. * Self learning hours shall not be reflected in the Time Table.
7. * Self learning includes micro project / assignment / other activities.

REINFORCEMENT LEARNING **Course Code : 316323**

V. THEORY LEARNING OUTCOMES AND ALIGNED COURSE CONTENT

Sr.No	Theory Learning Outcomes (TLO's)aligned to CO's.	Learning content mapped with Theory Learning Outcomes (TLO's) and CO's.	Suggested Learning Pedagogies.
1	TLO 1.1 Differentiate between Supervised, Unsupervised, and Reinforcement Learning. TLO 1.2 Explain Agent Environment Interface. TLO 1.3 Describe the Key Components of Reinforcement Learning. TLO 1.4 Explain various applications of Reinforcement Learning. TLO 1.5 Compare Model-based and Model-free learning. TLO 1.6 Explain Different Types of Environments in Reinforcement Learning.	**Unit - I Introduction to Reinforcement Learning** 1.1 Basics of Machine Learning: Supervised Learning, Unsupervised Learning & Reinforcement Learning 1.2 The Agent Environment Interface 1.3 Components of RL: Agent, Environment, State, Action, Reward 1.4 Applications of RL (e.g., robotics, gaming, recommendation systems) 1.5 Fundamental concepts of RL: Action space, Policy, Episode , Horizon, Return and discount factor, The value function 1.6 Model-based and Model-free learning 1.7 Different types of environments :Deterministic and stochastic environments, Discrete and continuous environments, Episodic and non-episodic environments, Single and multi-agent environments	Lecture Using Chalk-Board Presentations Video Demonstrations
2	TLO 2.1 Solve Multi-Armed Bandit Problem. TLO 2.2 Explain the trade-off between exploration and exploitation. TLO 2.3 Explain Markov Property. TLO 2.4 Explain the concept of a Markov Chain and its components. TLO 2.5 Explain Markov Reward and Markov decision processes. TLO 2.6 Solve Decision-Making Problems using Markov Decision Processes (MDPs). TLO 2.7 Solve state transition problems using epsilon-greedy algorithm.	**Unit - II Markov Decision Processes (MDPs)** 2.1 Multi-armed Bandit Problem 2.2 Exploration vs. Exploitation 2.3 epsilon-greedy algorithm 2.4 Markov property 2.5 Markov Chain 2.6 Markov Reward Processes (MRPs) 2.7 Markov Decision Processes (MDPs)	Lecture Using Chalk-Board Demonstration Presentations
3	TLO 3.1 Find the solution for Model based problem the Bellman Equation for the Value Function. TLO 3.2 Solve the Bellman Equation for the Q Function. TLO 3.3 Find the solution for Model based problem using Dynamic programming Value iteration. TLO 3.4 Solve Dynamic programming Value Policy Iteration.	**Unit - III The Bellman Equation and Dynamic Programming(DP)** 3.1 Bellman equation of the value function 3.2 Bellman equation of the Q function 3.3 Dynamic programming: Value iteration 3.4 Dynamic programming: Policy Iteration	Lecture Using Chalk-Board Presentations Video Demonstrations Presentations

REINFORCEMENT LEARNING **Course Code : 316323**

Sr.No	Theory Learning Outcomes (TLO's)aligned to CO's.	Learning content mapped with Theory Learning Outcomes (TLO's) and CO's.	Suggested Learning Pedagogies.
4	TLO 4.1 Explain the Monte Carlo Method. TLO 4.2 Find the solution for Modeless problem using the Monte Carlo Prediction. TLO 4.3 Solve Modeless problem using the Monte Carlo Control.	**Unit - IV Monte Carlo(MC) Methods** 4.1 Introduction to Monte Carlo Method 4.2 Monte Carlo Prediction: First-visit Monte Carlo 4.3 Monte Carlo Prediction: Every-visit Monte Carlo 4.4 Monte Carlo Control : On-Policy control 4.5 Monte Carlo Control : Off-Policy control	Lecture Using Chalk-Board Presentations Presentations Video Demonstrations Flipped Classroom
5	TLO 5.1 Solve the Temporal Difference (TD) Learning . TLO 5.2 Solve the Q-Learning algorithm. TLO 5.3 Solve SARSA algorithm. TLO 5.4 Differentiate between DP,MC,TD.	**Unit - V Temporal Difference (TD) Learning** 5.1 Introduction to Temporal Difference (TD) Learning 5.2 TD prediction algorithm 5.3 SARSA (State-Action-Reward-State-Action) 5.4 Q-Learning 5.5 Comparing the DP, MC, and TD methods	Presentations Lecture Using Chalk-Board Video Demonstrations Presentations

VI. LABORATORY LEARNING OUTCOME AND ALIGNED PRACTICAL / TUTORIAL EXPERIENCES.

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 1.1 Set up RL environment.	1	*Setting Up Python and RL Libraries	2	CO1
LLO 2.1 Explore OpenAI Gym to be used for RL Problems.	2	*Exploring OpenAI Gym Environments : a) List available environments using Load b) Render a basic environment and observe working of the environment	4	CO1
LLO 3.1 Implement Markov Chain.	3	Simulating a Markov Chain in Python	2	CO2
LLO 4.1 Implement the Markov Property in a simple Gym environment.	4	*Implementing the Markov Property in OpenAI Gym	2	CO2
LLO 5.1 Implement Bellman Expectation Equation.	5	*Implementing Bellman Expectation Equation in OpenAI Gym	2	CO3
LLO 6.1 Implement Dynamic programming's value iteration.	6	Write a python program using OpenAI Gym to solve the Frozen Lake problem using Value Iteration	2	CO3
LLO 7.1 Implement Dynamic programming Policy Iteration.	7	*Implement Python program using OpenAI Gym to solve the Frozen Lake problem using Dynamic programming Policy Iteration	2	CO3
LLO 8.1 Implement First-visit MC prediction algorithm.	8	Implement Python program using OpenAI Gym to implement First-visit MC prediction	2	CO4
LLO 9.1 Implement Every-visit Monte Carlo prediction algorithm.	9	*Implement First-visit MC prediction using python and OpenAI Gym	2	CO4
LLO 10.1 Implement TD prediction algorithm.	10	Using TD prediction algorithm solve the problem of Predicting the value of states in the Frozen Lake environment	2	CO5
LLO 11.1 Implement SARSA algorithm.	11	Using SARSA algorithm solve the problem of Predicting the value of states in the Frozen Lake environment	4	CO5

REINFORCEMENT LEARNING **Course Code : 316323**

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 12.1 Implement Q-Learning algorithm.	12	*Using Q-Learning algorithm solve the problem of Predicting the value of states in the Frozen Lake environment	4	CO5

Note : Out of above suggestive LLOs -

- '*' Marked Practicals (LLOs) Are mandatory.
- Minimum 80% of above list of lab experiment are to be performed.
- Judicial mix of LLOs are to be performed to achieve desired outcomes.

VII. SUGGESTED MICRO PROJECT / ASSIGNMENT/ ACTIVITIES FOR SPECIFIC LEARNING / SKILLS DEVELOPMENT (SELF LEARNING)

Other

- Complete any course related to the Reinforcement Learning on MOOCS such as NPTEL, Coursera, Infosys Springboard etc.

Assignment

- Solve assignments covering all COs given by the course teacher.
- Assignments on Monte Carlo simulations to assess the risk and uncertainty in investment projects.
- Derive the Bellman Equation for Value Function.
- Solve the Bellman Equation for a Simple MDP.

Micro project

- Mountain Car Problem: In this OpenAI Gym environment, an agent learns to drive a car up a hill.
- Utilize OpenAI Gym to create an environment where an agent learns to balance a pole on a moving cart. Implement algorithms like Q-learning or policy gradients to train the agent.
- Tic-Tac-Toe Game: Develop an RL agent that learns to play Tic-Tac-Toe, focusing on state representation and reward shaping.

Note :

- Above is just a suggestive list of microprojects and assignments; faculty must prepare their own bank of microprojects, assignments, and activities in a similar way.
- The faculty must allocate judicial mix of tasks, considering the weaknesses and / strengths of the student in acquiring the desired skills.
- If a microproject is assigned, it is expected to be completed as a group activity.
- SLA marks shall be awarded as per the continuous assessment record.
- For courses with no SLA component the list of suggestive microprojects / assignments/ activities are optional, faculty may encourage students to perform these tasks for enhanced learning experiences.
- If the course does not have associated SLA component, above suggestive listings is applicable to Tutorials and maybe considered for FA-PR evaluations.

VIII. LABORATORY EQUIPMENT / INSTRUMENTS / TOOLS / SOFTWARE REQUIRED

Sr.No	Equipment Name with Broad Specifications	Relevant LLO Number
1	Computer system - (Computer System which is available in lab with minimum 4GB RAM)	All
2	Python 3.7 or higher	All

IX. SUGGESTED WEIGHTAGE TO LEARNING EFFORTS & ASSESSMENT PURPOSE (Specification

REINFORCEMENT LEARNING **Course Code : 316323**

Table)

Sr.No	Unit	Unit Title	Aligned COs	Learning Hours	R-Level	U-Level	A-Level	Total Marks
1	I	Introduction to Reinforcement Learning	CO1	12	4	10	6	20
2	II	Markov Decision Processes (MDPs)	CO2	9	4	4	6	14
3	III	The Bellman Equation and Dynamic Programming(DP)	CO3	8	4	4	4	12
4	IV	Monte Carlo(MC) Methods	CO4	8	2	6	4	12
5	V	Temporal Difference (TD) Learning	CO5	8	2	6	4	12
Grand Total				**45**	**16**	**30**	**24**	**70**

X. ASSESSMENT METHODOLOGIES/TOOLS

Formative assessment (Assessment for Learning)

- Two unit tests of 30 marks each conducted during the semester. Continuous assessment based on process and product related performance indicators. Each practical will be assessed considering 60% weightage to process, 40% weightage to product. A continuous assessment based term work.

Summative Assessment (Assessment of Learning)

- End semester examination, Lab performance, Viva voce.

XI. SUGGESTED COS - POS MATRIX FORM

Course Outcomes (COs)	Programme Outcomes (POs)							Programme Specific Outcomes* (PSOs)		
	PO-1 Basic and Discipline Specific Knowledge	PO-2 Problem Analysis	PO-3 Design/ Development of Solutions	PO-4 Engineering Tools	PO-5 Engineering Practices for Society, Sustainability and Environment	PO-6 Project Management	PO-7 Life Long Learning	PSO-1	PSO-2	PSO-3
CO1	3			1			1			
CO2	3	2	2	2	1		1			
CO3	2	2	2	3	1		1			
CO4	2	2	2	3	1	2	1			
CO5	2	2	2	3	1	2	1			

Legends :- High:03, Medium:02,Low:01, No Mapping: -
*PSOs are to be formulated at institute level

XII. SUGGESTED LEARNING MATERIALS / BOOKS

Sr.No	Author	Title	Publisher with ISBN Number
1	Sudharsan Ravichandiran	Deep Reinforcement Learning with Python	Packt Publishing ISBN : 978-1-83921-068-6, Second Edition 2020
2	Richard S. Sutton and Andrew G. Barto	Reinforcement Learning: An Introduction Second edition	The MIT Press. ISBN: 978-0-26203-924-6
3	Phil Winder	Reinforcement Learning Industrial Applications of Intelligent Agents	O'Reilly Media, Inc. ISBN-978-1-09811-483-1, First edition 2020
4	Miguel Morales	Grokking Deep Reinforcement Learning	Manning Publisher ISBN: 978-161729545 4

REINFORCEMENT LEARNING **Course Code : 316323**

Sr.No	Author	Title	Publisher with ISBN Number
5	Keng, Wah Loon, Graesser, Laura	Foundations of Deep Reinforcement Learning: Theory and Practice in Python	Addison-Wesley Professional ISBN: 978-0135172384

XIII . LEARNING WEBSITES & PORTALS

Sr.No	Link / Portal	Description
1	https://neptune.ai/blog/best-reinforcement-learning-tutorials-examples-projects-and-courses	Best Reinforcement Learning Tutorials, Examples, Projects, and Courses
2	https://www.geeksforgeeks.org/what-is-reinforcement-learning/	Reinforcement Learning tutorials
3	https://www.youtube.com/playlist?list=PLz2x4RAIbeXkTJFEipkD_ds3z0qx8_5D7	Reinforcement Learning Tutorials - NPTEL
4	https://swayam.gov.in/search_courses?searchText=RL	SWAYAM
5	https://infyspringboard.onwingspan.com/web/en/login	Infosys Springboard
6	https://deeplizard.com/learn/video/nyjbcRQ-uQ8	Reinforcement Learning Series Introduction
7	https://gordicaleksa.medium.com/how-to-get-started-with-reinforcement-learning-rl-4922fafeaf8c	How to get started with Reinforcement Learning (RL)
8	https://www.youtube.com/playlist?list=PLzvYlJMoZ02Dxtwe-MmH4nOB5jYlMGBjr	Reinforcement Learning By the Book

Note :

- Teachers are requested to check the creative common license status/financial implications of the suggested online educational resources before use by the students

MSBTE Approval Dt. 04/09/2025 **Semester - 6, K Scheme**