25UAMT304	FUNDAMENTALS OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE	Category	L	T	P	R	Credit	Year of Introduction
		ESC	3	1	0	0	4	2025

Preamble: Artificial Intelligence (AI) and Data Science are transformative technologies that are reshaping industries, economies, and everyday life across the globe. From intelligent virtual assistants and recommendation systems to autonomous vehicles and predictive healthcare analytics, AI and Data Science are driving innovation and enabling data-driven decision-making in unprecedented ways. The course provide students with a foundational understanding of the principles, methodologies, and applications of intelligent systems and data analytics.

Prerequisite: Nil

Course Outcomes: After the completion of the course, the student will be able to:

CO	Course Outcomes	Knowledge Level
CO1	Explain the concept of machine learning including neural networks and different learning techniques for engineering applications.	**K2** Understand
CO2	Apply mathematical concepts such as matrix operations, probability And regression to solve engineering problems.	**K3** Apply
CO3	Demonstrate knowledge of essential skills and tools required for Data Science, including data storage techniques and dataset creation.	**K3** Apply
CO4	Apply appropriate data pre-processing techniques to prepare datasets for analysis and modeling.	**K3** Apply

Mapping of course outcomes with program outcomes and program specific outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11
CO1	2	2	2	-	-	-	-	-	-	-	-
CO2	2	2	2	-	-	-	-	-	-	-	-
CO3	2	3	3	2	-	-	-	-	-	-	-
CO4	3	3	3	2	-	-	-	-	-	-	-
Avg	2.2	2.5	2.5	2	-	-	-	-	-	-	-

Mapping of course outcomes with Knowledge and Attitude Profile (WK)

	WK1	WK2	WK3	WK4	WK5	WK6	WK7	WK8	WK9
CO1	*	*	*	*	*				
CO2	*	*	*	*					
CO3		*	*	*	*				
CO4	*	*		*	*				

WKs Mapped with the course

WK1	WK2	WK3	WK4	WK5	WK6	WK7	WK8	WK9
*	*	*	*	*				

Mark Distribution:

Total Marks	CIE Marks	ESE Marks	ESE Duration
100	**40**	**60**	**2.5 hrs**

Continuous Internal Evaluation Pattern:

Attendance : 5 marks
Continuous Assessment Test : 20 marks
Continuous Assessment Assignment : 15 marks

SYLLABUS

Module 1 (14 hrs)

Basics of Machine Learning -types of Machine Learning systems-challenges in ML- Supervised learning model example- regression models- Classification model example- Logistic regression- unsupervised model example- K-means clustering. Artificial Neural Network- Perceptron- Universal Approximation Theorem.

Module 2 (14 hrs)

Role of linear algebra in Data representation and analysis – Matrix decomposition- Singular Value Decomposition (SVD)- Spectral decomposition. Basics of probability-random variables and statistical measures - rules in probability-Bayes theorem and its applications-Correlation analysis-linear correlation - regression analysis- linear regression.

Module 3 (14 hrs)

Data Science: Steps in doing Data Science - Data Science relation to other fields- Data Science and Information Science- Computational Thinking - Skills and tools needed to do Data Science - Storing data - Combining bytes into larger structures - Creating data sets - Identifying data problem - Understanding data sources - Exploring data models- Introduction to Big Data.

Module 4 (14 hrs)

Structured and unstructured data - Challenges with unstructured data - Data collection: Open data - multimodal data - Data Pre-processing: Data Cleaning - Data Integration, Data Transformation - Data Reduction - Data Discretization. Applications of Machine Learning in Data Science- modelling process- demonstration of ML. Benefits of data science-use of statistics and Machine Learning in Data Science- data science process applications in data science.

Text Books:

1. Gilbert Strang(2023), Introduction to Linear Algebra, Wellesley-Cambridge Press, 6th edition.
2. Gupta.S.C and V.K.Kapoor(2020) Fundamentals of mathematical statistics, Sultan Chand & Sons ,9th edition
3. Cielen, Davy and Arno Meysman(2016), Introducing data science: bigdata, machine learning, and more, using Python tools, Simon and Schuster, 1st edition

Reference Books:

1. Avrim Blum, John Hopcroft, and Ravi Kannan(2020) , Foundations of Data Science, Cambridge University Press, 1st edition
2. Carlos Fernandez(2017), Probability and Statistics for Data Science , Center for Data Science in NYU, 1st edition.
3. Kotu Vijay, and Bala Deshpande(2018) ,Data science: concepts and practice, Morgan 2nd edition

END SEMESTER EXAMINATION – MODEL QUESTION PAPER

Course Code:	**25UAMT304**
Course Name:	**Fundamentals of Artificial Intelligence & Data Science**
Max. Marks: 60	**Duration: 2.5 Hours**

Part A (Answer all questions. Each question carries 3 marks)				
Qn No.	**Question**	**Marks**	**CO**	**BL**
1	Define machine learning and explain the difference between supervised and unsupervised learning with suitable examples.	3	CO1	L2
2	Explain the concept of Logistic Regression and state how it differs from Linear Regression.	3	CO1	L2
3	Define Singular Value Decomposition (SVD) and state its significance in data representation.	3	CO2	L2
4	State Bayes' theorem and explain its role in probability-based classification.	3	CO2	L2
5	List the essential skills and tools required to perform Data Science and explain any two in brief.	3	CO3	L2

6	Explain the concept of Computational Thinking and its relevance to Data Science.	3	CO3	L2
7	Define data pre-processing and explain the need for data cleaning in the data analysis pipeline.	3	CO4	L2
8	Distinguish between structured and unstructured data with suitable examples.	3	CO4	L2

Part B (Answer any one full question from each Module. Each carries 9 marks)				
Module I				
Qn No.	**Question**	**Marks**	**CO**	**BL**
9	Explain the different types of Machine Learning systems in detail. Discuss the challenges involved in Machine Learning and illustrate how a supervised learning regression model works with an example.	9	CO1	L2
OR				
10	Describe the architecture and working principle of an Artificial Neural Network. Explain the Perceptron model and state the Universal Approximation Theorem with its significance in neural network design.	9	CO1	L2
Module II				
Qn No.	**Question**	**Marks**	**CO**	**BL**
11	A retail company has collected customer data represented as a matrix A consisting of 4 customers (rows) and 3 features: monthly spend, number of visits, and loyalty score (columns). The company wants to reduce the dimensionality of this data for visualization while preserving maximum variance. (a) Explain how Singular Value Decomposition (SVD) of matrix A can be used to achieve this dimensionality reduction. Outline the steps involved, including how the singular values indicate which components to retain. (b) A discrete random variable X represents the number of visits per month by a customer, with the following probability distribution: X = 1, 2, 3, 4 with P(X) = 0.1, 0.3, 0.4, 0.2 respectively. Calculate the mean (expected value) and variance of X, and interpret what these values tell the company about customer visiting behaviour.	9	CO2	L3
OR				
12	A medical testing centre uses a screening test for a disease that is	9	CO2	L3

	present in 2% of the population. The test correctly identifies the disease in 95% of people who actually have it (true positive rate), and incorrectly indicates the disease in 4% of people who do not have it (false positive rate). A randomly selected person tests positive. (a) Apply Bayes' theorem to calculate the probability that this person actually has the disease. Comment on what this result implies about the reliability of the test in practice. (b) The centre also recorded the age (X, in years) and blood pressure (Y, in mmHg) of 6 patients: X: 25, 32, 40, 48, 55, 60 Y: 118, 122, 130, 138, 145, 150 Calculate Pearson's correlation coefficient between age and blood pressure, and use simple linear regression to predict the blood pressure of a 50-year-old patient.			
Module III				
Qn No.	**Question**	**Marks**	**CO**	**BL**
13.a	A retail startup wants to build a system that predicts which products a customer will buy next, using past purchase records, website browsing logs, and social media reviews. Apply the steps involved in doing Data Science to outline how the team should proceed from problem identification to deployment, and justify how this process relates Data Science to other fields such as Information Science.	5	CO3	L3
13.b	A hospital wants to create a single dataset by combining numeric patient vitals, text-based doctor's notes, and X-ray images for analysis. Apply appropriate data storage techniques to combine these different types of bytes into larger structures, and identify the skills and tools the data science team would need to create this dataset.	4	CO3	L3
OR				
14.a	A telecom company is experiencing high customer churn but does not know the underlying cause. Apply the process of identifying the data problem and understanding the relevant data sources, such as call records, billing history, and complaint logs, that the company should investigate. Demonstrate with a suitable example how exploring different data models could help uncover patterns behind the churn.	5	CO3	L3
14.b	A social media platform generates millions of posts every day and wants to identify trending topics in real time. Apply the concept of Computational Thinking to decompose this problem into manageable parts, and demonstrate the skills and tools needed to handle the Big Data characteristics, namely volume, velocity, and variety, involved in solving it.	4	CO3	L3

Module IV				
Qn No.	**Question**	**Marks**	**CO**	**BL**
15.a	Apply data pre-processing techniques—including Data Cleaning, Data Integration, Data Transformation, and Data Reduction—to prepare a given dataset for machine learning modeling. Explain Data Discretization with examples.	6	CO4	L3
15.b	Analyse how Machine Learning is applied in Data Science by describing the modelling process and the benefits of using statistics and Machine Learning in Data Science applications.	3	CO4	L3
OR				
16.a	Analyse the challenges involved in handling unstructured data. with the methods of data collection including open data and multimodal data, and apply appropriate pre-processing strategies for such data.	6	CO4	L3
16.b	Apply the data science process with real-world applications. Demonstrate how the use of statistics and Machine Learning contributes to data science process outcomes.	3	CO4	L3