Basavarajeswari Group of Institutions

2022 SCHEME

BALLARI INSTITUTE OF TECHNOLOGY & MANAGEMENT

(Autonomous Institute under Visvesvaraya Technological University, Belagavi)

USN | | | | | | | | | | |

Course Code | 2 | 3 | M | C | A | 4 | 1 | 1 |

Fourth Semester MCA Degree Examinations, September 2025

INTRODUCTION TO DATA SCIENCE

Duration: 3 hrs **Max. Marks: 100**

Note: ***1. Answer any* FIVE *full questions, choosing* ONE *full question from each module.***
2. Missing data, if any, may be suitably assumed

Q. No		***Question***	***Marks***	*(RBTL:CO: PI)*
		MODULE – 1		
1.	**a.**	Discuss the importance of data visualization in the field of data science. Explain any two common types of visualization.	**10**	**(2:1:2.2.1)**
	b.	Define linear algebra and summarize the role of linear algebra in data science.	**10**	**(2:1:3.2.2)**
		(OR)		
2.	**a.**	Describe the operations and use cases of matrices in data science and why they are important.	**10**	**(2:1:2.1.3)**
	b.	What do you understand by the central limit theorem? Describe how it works.	**10**	**(2:1:2.2.1)**
		MODULE – 2		
3.	**a.**	Explain statistical hypothesis testing in the context of data science using an example.	**10**	**(2:2:2.1.3)**
	b.	Analyze the problem of p-hacking in statistical research. How can it be identified and prevented?	**10**	**(3:2:2.4.1)**
		(OR)		
4.	**a.**	How does gradient descent help in optimizing a machine learning model? Explain with its working process.	**10**	**(3:2:2.1.3)**
	b.	Define web scraping. Describe how web scraping is done and discuss its use cases.	**10**	**(2:2:1.2.1)**
		MODULE – 3		
5.	**a.**	Compare and contrast feature extraction and feature selection in terms of various aspects.	**10**	**(4:3:2.3.2)**
	b.	Discuss K-Nearest Neighbour (KNN). Explain its working with an example.	**10**	**(2:3:2.2.1)**
		(OR)		
6.	**a.**	What are the features and working principles of a more sophisticated spam filter? Discuss the email classification example.	**10**	**(2:3:2.3.2)**
	b.	Give a brief overview of simple linear regression and describe the further assumptions associated with the least squares approach.	**10**	**(2:3:2.1.3)**

Note: (**RBTL - Revised Bloom's Taxonomy Level: CO - Course Outcome: PI - Performance Indicator)**

MODULE – 4

7.	**a.**	Explain how random forests improve upon decision trees. What role does randomness play in their effectiveness?	**10**	**(2:4:2.2.1)**
	b.	What is a perceptron, and how does it function as the building block for neural networks?	**10**	**(2:4:2.3.1)**

(OR)

8.	**a.**	Discuss the purpose of using activation functions in neural networks. Give examples of commonly used ones.	**10**	**(2:4:2.1.3)**
	b.	What is a tensor, and why is it the fundamental data structure in deep learning frameworks like PyTorch or Tensor Flow?	**10**	**(2:4:2.2.1)**

MODULE – 5

9.	**a.**	How does eigenvector centrality differ from degree centrality? What does it imply if a node has high eigenvector centrality?	**10**	**(3:5:2.4.2)**
	b.	Describe directed graph in the context of network analysis and explain why it matters?	**10**	**(2:5:2.1.2)**

(OR)

10.	**a.**	What are the advantages and limitations of manual curation in recommender systems?	**10**	**(2:5:2.3.2)**
	b.	Briefly describe user-based and item-based collaborative filtering techniques used in recommender systems.	**10**	**(2:5:2.3.1)**

** ** **

Note: (**RBTL - Revised Bloom's Taxonomy Level: CO - Course Outcome: PI - Performance Indicator)**