

# CBCS SCHEME

USN 

--	--	--	--	--	--	--	--	--	--

22SCS/SAM/SCN/SAD/VSA/VSC/VCS12

## First Semester M.Tech Degree Examination, June/July 2023 Fundamentals of Data Science

Time: 3 hrs.

Max. Marks: 100

*Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.  
2. M: Marks, L: Bloom's level, C: Course outcomes.*

Module – 1			M	L	C
Q.1	a.	What is data science and Big data? Explain need for data science.	10	L1	CO1
	b.	Explain the following terms : i) Data integration ii) Data reduction iii) Data transformation iv) Data discretization v) Datafication.	10	L1	CO1
<b>OR</b>					
Q.2	a.	With a neat diagram explain data science process.	10	L2	CO1
	b.	Briefly explain how the class and objects are defined in R.	10	L2	CO2
<b>Module – 2</b>					
Q.3	a.	State and explain the steps involved in K-means clustering.	10	L2	CO2
	b.	Explain EDA (Exploratory Data Analysis) with the help of example.	10	L2	CO2
<b>OR</b>					
Q.4	a.	What is linear regression algorithm? What are the basic assumptions of the linear regression algorithm?	10	L1	CO2
	b.	Define KNN. Explain similarity or distance matrices in detail.	10	L2	CO2
<b>Module – 3</b>					
Q.5	a.	How to build and apply Naïve Bayes classification for spam filtering.	10	L2	CO3
	b.	Which algorithms are best to use for spam filtering? How should they be implemented?	10	L1	CO3
<b>OR</b>					
Q.6	a.	Explain the purpose of Laplace smoothing.	10	L2	CO3
	b.	Explain the three core problems of data scientists at M6D.	10	L2	CO3
<b>Module – 4</b>					
Q.7	a.	What is the difference between wrapper and embedded feature selection?	10	L1	CO3
	b.	What is Entropy? Explain Decision Tree Algorithm.	10	L2	CO3
<b>OR</b>					
Q.8	a.	Explain Random forests algorithm with a sample code.	10	L2	CO3
	b.	Explain principal component analysis (PCA) with its importance and limitations.	10	L2	CO3
<b>Module – 5</b>					
Q.9	a.	What is Map Reduce? Explain the working of Map Reduce.	10	L2	CO4
	b.	Define Data Visualization. Illustrate how data visualization is better than the traditional text based data methods.	10	L2	CO4
<b>OR</b>					
Q.10	a.	What are Social Networks Graphs? How does clustering of social networks graph works?	10	L1	CO4
	b.	Explain Girvan – Newman Algorithm for Community Detection.	10	L2	CO4

\*\*\*\*\*

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.  
2. Any revealing of identification, appeal to evaluator and /or equations written eg. 42+8 = 50, will be treated as malpractice.