P P SAVANI UNIVERSITY

Seventh Semester of B. Tech. Examination

December 2021

SEIT4013 Data Science

01.12.2021, Wednesday **Time: 09:00 a.m. To 11:30 a.m.** **Maximum Marks: 60**

Instructions:

1. The question paper comprises of two sections.
2. Section I and II must be attempted in separate answer sheets.
3. Make suitable assumptions and draw neat figures wherever required.
4. Use of scientific calculator is allowed.

SECTION – I

Q - 1	Answer the Following: (MCQ/Short Question/Fill in the Blanks)	**[05]**
(i)	API stands for _____	
(ii)	State Normal distribution formula.	
(iii)	What is significance level in hypothesis testing?	
(iv)	What is Q1 value for dataset containing following set of values? set=[90,9,12,13,14,15,16,17,23,90,99]	
(v)	Division operator can be applied on 'gender' column (True/False)	
Q - 2 (a)	What is Data Science? Explain the process of Data Science taking suitable example.	**[05]**
Q - 2 (b)	What is ANOVA? Explain different steps involved in ANOVA test by taking suitable example	**[05]**
	OR	
Q - 2 (b)	What do you mean by Inferential Statistics? Explain different steps involved in chi-square test by taking suitable example.	**[05]**
Q - 3 (a)	Explain different application areas of Data Science	**[05]**
Q - 3 (b)	List different probability distribution present in Data Science. Discuss normal distribution and its properties in detail.	**[05]**
	OR	
Q - 3 (b)	Explain in detail following terms a) p-value in hypothesis testing b) Data scales	**[05]**
Q - 4	Attempt any one	**[05]**
(i)	Differentiate between primary data source and secondary data source.	
(ii)	Differentiate between Type –I and Type –II error in hypothesis testing.	

SECTION – II

Q - 1	Answer the Following (Attempt any five)	**[05]**
(i)	PCA	
(ii)	Logistic Regression	
(iii)	Cost Function	
(iv)	Hyper plane	
(v)	Sensitivity and specificity	
(vi)	Confusion Matrix	
(vii)	Retinal Variables	
Q - 2 (a)	Explain the Validation process of the Simple Linear Regression Model.	**[05]**
Q - 2 (b)	Explain Decision tree algorithm with suitable example.	**[05]**
	OR	
Q - 2 (a)	Explain Random forest? How does it works?	**[05]**
Q - 2 (b)	Explain Naive Bayesian Classifier with example.	**[05]**
Q - 3 (a)	Explain any two methods for Data Visualization.	**[05]**

Q - 3 (b)	Explain Data Encoding with example.	**[05]**

OR

Q - 3 (a)	Explain pandas Library.	**[05]**
Q - 3 (b)	What is supervised and unsupervised learning? Differentiate between them.	**[05]**
Q - 4	Attempt any one	**[05]**
(i)	Explain various Application Development Methods used in Data Science.	
(ii)	Explain various Data Collection and Analysis Techniques.	

****