Name:

Enrolment No:

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2022

Course: Data Analysis and Modelling Technique **Semester: 8th**
Program: B-Tech. CSE (AIML) **Time : 03 hrs.**
Course Code: CSBA4014 **Max. Marks: 100**

Instructions: Check the questions very minutely. Utilize your time according to the marks listed for every questions.

SECTION A
(5Qx4M=20Marks)

Each Question will carry 4 Marks. Explain max by 50-60 words wherever required. Attempt all questions from Sec A.

S. No.		Marks	CO
Q 1.	You got a dataset depicting the popularity of two graphic novels given by a critic which contains three variables. 1) Time of survey (in dd-mm-yy format) 2) Rating of 'Marvel' (in range between 0 to 10) 3) Rating of 'DC' (in range between 0 to 10) The data is collected every day since 1970. You need to graphically represent the data in a chart. What will you use? And why?	**04**	**CO1**
Q 2.	Three athletes A, B and C are participating in the Olympics. A is twice as likely to win as B and B is twice as likely to win as C. What are the probabilities of their winning?	**04**	**CO2**
Q 3.	How do you test a small sample hypothesis?	**04**	**CO3**
Q 4.	What is difference between simple linear and multiple linear regressions?	**04**	**CO4**
Q 5.	Differentiate between discrete and continuous random variable.	**04**	**CO1**

SECTION B
(4Qx10M= 40 Marks)

Each question will carry 10 marks. Write short / brief notes (Explain max by 100-150 words wherever required).

Q 6.	Analyzing the Mid-sem marks for students. The following data was observed.	**10**	**CO2**

Sl no	Total Students
0-10	5
10-20	3

	<table><tr><td>20-30</td><td>5</td></tr><tr><td>30-40</td><td>8</td></tr><tr><td>40-50</td><td>16</td></tr><tr><td>50-60</td><td>18</td></tr><tr><td>60-70</td><td>5</td></tr><tr><td>70-80</td><td>3</td></tr><tr><td>80-90</td><td>2</td></tr><tr><td>90-100</td><td>0</td></tr></table> **a)** Compute the Skewness present in the data? What can you conclude? **(5 marks)** **b)** Compute the kurtosis. What is the observation indicating? **(5 marks)**		
Q 7.	**a)** What do you understand by the term descriptive statistics. **(3 marks)** **b)** And provide an example of descriptive statistics? **(7 marks)**	**10**	**CO2**
Q 8.	**Attempt 8(a and b) or 8(c)** **a)** For the marks of 25 students studying BAO : 20, 21, 19, 18, 20, 20, 19, 18, 21, 19, 22, 21, 18, 19, 21, 22, 19, 18, 20, 19, 20, 22, 20, 21, 20.Discuss the discrete frequency distribution. **(4 marks)** **b)** Explain the Central Limit Theorem. **(6 marks)** **OR** **c)** A fellow researcher claims that at least 15% of smokers fail to eat any fruits and vegetables at least 3 days a week. You find this hard to believe and decide to check the validity of this statistic by taking a random (representative) sample of smokers. Do you have sufficient evidence to reject your colleague's claim if you discover that 17 of the 200 smokers in your sample eat no fruits and vegetables at least 3 days a week? **(10 marks)**	**10**	**CO1**
Q 9.	Write short note on: (**Attempt any two**) **(2*5)** **(i)** Z Test. **(ii)** T Test. **(iii)** Bayesian Network. **(iv)** Maximum likelihood estimation	**10**	**CO3**
SECTION-C **Each Question carries 20 Marks. Instruction: Write long answer. Explain max by 200 words wherever required. Make diagrams wherever needed. (2Qx20M=40 Marks)**			
Q 10.	**<u>Attempt 10(a) or 10(b)</u>** **a)** Explain the concept and working principle of the Monte Carlo simulation along with their advantages and disadvantages. **(20 marks)** **OR**	**20**	**CO4**

	b) Explain the basic concepts of Hidden Markov Model(HMM) including **i)** Markov chain, **ii)** definition of HMM, **iii)** HMM assumptions, **iv)** Computing Likelihood: The Forward Algorithm, **v)** Learning in HMM, **vi)** Advantages and Disadvantages of HMM). **(3+2+4+5+2+4)**		
Q 11.	**a)** Given the following statistics, what is the probability that a woman has cancer if she has a positive mammogram result? 1. 1% of women have cancer. 2. 90% of women who have cancer test positive on mammograms. 3. 8% of women will have false positives. **(8 marks)** **b)** How to find f test and t test p values? **(6 marks)** **c)** Difference between Bayesian Network and Markov model? **(6 marks)**	**20**	**CO3, CO4**