Name:

Enrolment No:

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2023

Course: BCA **Semester : 2**
Program: Machine Learning for Business **Time : 03 hrs.**
Course Code: CSAI2011 **Max. Marks: 100**

Instructions: Attempt all questions

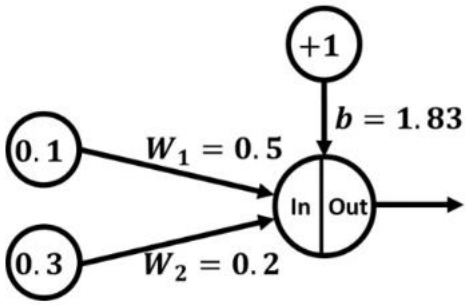
SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Elaborate the term "Statistics". What do you understand by term "population" & "sample"?	**4**	**CO1**
Q2.	Discuss all the types of sampling techniques.	**4**	**CO2**
Q3	Explain Five Number Summary concept with an example.	**4**	**CO3**
Q4	a) Differentiate between Type 1 and Type 2 error. b) Describe the concept of Spearman's Rank Correlation	**4**	**CO1**
Q5	Describe Supervised and Unsupervised Learning.	**4**	**CO4**

SECTION B
(4Qx10M= 40 Marks)

Q6	In an ODI Series, In 2021 series average score is 250 with standard deviation of 10. Rishabh Panth's average score of the series in 2021 was 240. Whereas in 2020, Series average score was 260 with a standard deviation of 12. In the same year Rishabh Path average score of the series was 245. Compare to both the series in which year Rishabh Panth performance was better?	**10**	**CO4**
Q7	On the quant test of an exam, the standard deviation is known to be 100. A sample of 25 test takers has a mean of 520 score. Construct a 95% Confidence Interval about the mean. OR For the given set of points identify clusters using complete link and average link using agglomerative clustering	**10**	**CO1**

A	A	B
P1	1	1
P2	1.5	1.5
P3	5	5

	P4 | 3 | 4 P5 | 4 | 4 P6 | 3 | 3.5		
Q8	Differentiate between the hyperplanes of Logistic Regression and SVM. Which algorithm to use when, Explain with an example.	**10**	**CO3**
Q9	Explain the concept of underfitting and overfitting in machine learning. Describe two techniques that can be used to prevent overfitting in a machine learning model, and provide a brief explanation of how each technique works	**10**	**CO2**
	SECTION-C **(2Qx20M=40 Marks)**		
Q10	Find the output in the following ANN having one neuron with two input layer and one activation function and bias value: 	**20**	**CO3, CO4**
Q11	The following data set contains factors that determine whether tennis is played or not. Using Naive Bayes classifier, find the play prediction for the day <Sunny, Cool, High, Strong>	**20**	**CO2**

DAY	OUTLOOK	TEMP	HUMIDITY	WIND	PLAY
Day 1	Sunny	Hot	High	Weak	NO
Day 2	Sunny	Hot	High	Strong	NO
Day 3	Overcast	Hot	High	Weak	YES
Day 4	Rain	Mild	High	Weak	YES
Day 5	Rain	Cool	Normal	Weak	YES
Day 6	Rain	Cool	Normal	Strong	NO
Day 7	Overcast	Cool	Normal	Strong	YES
Day 8	Sunny	Mild	High	Weak	NO
Day 9	Sunny	Cool	Normal	Weak	YES
Day 10	Rain	Mild	Normal	Weak	YES
Day 11	Sunny	Mild	Normal	Strong	YES
Day 12	Overcast	Mild	High	Strong	YES
Day 13	Overcast	Hot	Normal	Weak	YES
Day 14	Rain	Mild	High	Strong	NO

OR

Describe the steps involved in the k-nearest neighbors (KNN) algorithm for classification. Include details on how to determine the optimal value of k, and explain the strengths and weaknesses of the KNN algorithm. Explain Time and Space Complexity of KNN algorithm.