

Name:

Enrolment No:



UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2020

Course: Data Mining & Prediction Modeling

Semester: III

Program: B.Tech CSE BAO

Time : 03 hrs.

Course Code: CSBA3001

Max. Marks: 100

Instructions:

1. In Section A, you have to write one word/one sentence answers, no explanation, no calculation is required to be furnished.
2. Section B and C are the sections in which you will be writing the answers on A4 sheets and after clicking the picture, upload as per the direction.

SECTION A

S. No.		Marks	CO
Q 1	Name FIVE different methods through which the missing values in data sets are filled-in.	5	CO1
Q 2	Name THREE data mining task which are <i>predictive</i> in nature.	5	CO2
Q 3	Name FIVE measures which are generally used to evaluate the performance of a classifier or classification method.	5	CO3
Q 4	Name THREE validation methods of a model?	5	CO4
Q 5	Name FIVE steps in KDD process.	5	CO1
Q 6	a) Approximately how much data falls in the range [(Mean-3*Std. Dev.) - (Mean+3*Std.Dev.)] to qualify it as a NORMAL distributed data. b) Let, in a given data set, the mean μ is 40 and standard deviation σ is 16, what will be the z-score for the value 85? c) What will be the Supremum distance between two data points (5,7,10) and (6,8,2)? d) If we do the partitioning of dataset, and pick up the proportional volume from each partition, which type of sampling this is called? e) Name THREE data visualization techniques.	5	CO2

SECTION B

Q 7	For a given Symptoms and Diagnosis dataset – Classify whether patient has flu or not for the input (<i>No, No, Mild, yes,?</i>) (Hint: Use Naïve Bayes Classifier)	10	CO3																																													
	<table border="1"> <thead> <tr> <th>CHILLS</th> <th>RUNNY NOSE</th> <th>HEADACHE</th> <th>FEVER</th> <th>FLU</th> </tr> </thead> <tbody> <tr> <td>YES</td> <td>NO</td> <td>MILD</td> <td>YES</td> <td>NO</td> </tr> <tr> <td>YES</td> <td>YES</td> <td>NO</td> <td>NO</td> <td>YES</td> </tr> <tr> <td>YES</td> <td>NO</td> <td>STRONG</td> <td>YES</td> <td>YES</td> </tr> <tr> <td>NO</td> <td>YES</td> <td>MILD</td> <td>YES</td> <td>YES</td> </tr> <tr> <td>NO</td> <td>NO</td> <td>NO</td> <td>NO</td> <td>NO</td> </tr> <tr> <td>NO</td> <td>YES</td> <td>STRONG</td> <td>YES</td> <td>YES</td> </tr> <tr> <td>NO</td> <td>YES</td> <td>STRONG</td> <td>NO</td> <td>NO</td> </tr> <tr> <td>YES</td> <td>YES</td> <td>MILD</td> <td>YES</td> <td>YES</td> </tr> </tbody> </table>			CHILLS	RUNNY NOSE	HEADACHE	FEVER	FLU	YES	NO	MILD	YES	NO	YES	YES	NO	NO	YES	YES	NO	STRONG	YES	YES	NO	YES	MILD	YES	YES	NO	NO	NO	NO	NO	NO	YES	STRONG	YES	YES	NO	YES	STRONG	NO	NO	YES	YES	MILD	YES	YES
	CHILLS			RUNNY NOSE	HEADACHE	FEVER	FLU																																									
	YES			NO	MILD	YES	NO																																									
	YES			YES	NO	NO	YES																																									
	YES			NO	STRONG	YES	YES																																									
	NO			YES	MILD	YES	YES																																									
	NO			NO	NO	NO	NO																																									
NO	YES	STRONG	YES	YES																																												
NO	YES	STRONG	NO	NO																																												
YES	YES	MILD	YES	YES																																												
Q 8	Discuss the advantages and disadvantages of using sampling to reduce the number of data objects that need to be displayed. Would simple random sampling (without replacement) be a good approach to sampling? Why or Why not?	10	CO2																																													
Q 9	What are ensemble classifiers? Discuss Bagging and Boosting in nutshell. What is the underlying difference between these two methods?	10	CO4																																													

