

## 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.

SE	CT	PΤ	M	M	Λ
	v	LL	v.		$\boldsymbol{\Gamma}$

S. No.		Marks	CO
Q 1	Name <b>FIVE</b> different methods through which the missing values in data sets are filled-in.	5	CO1
Q 2	Name <b>THREE</b> data mining task which are <i>predictive</i> in nature.	5	CO2
Q 3	Name <b>FIVE</b> measures which are generally used to evaluate the performance of a classifier or classification method.	5	CO3
Q 4	Name <b>THREE</b> validation methods of a model?	5	CO4
Q 5	Name FIVE steps in KDD process.	5	CO1
Q 6	<ul> <li>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.</li> <li>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?</li> <li>c) What will be the Supremum distance between two data points (5,7,10) and (6,8,2)?</li> <li>d) If we do the partitioning of dataset, and pick up the proportional volume from each partition, which type of sampling this is called?</li> <li>e) Name THREE data visualization techniques.</li> </ul>	5	CO2

## SECTION B

	For a given Symptoms and Diagnosis dataset – Classify whether patient has flu or no								
	for the input (No, No, Mild, yes,?) (Hint: Use Naïve Bayes Classifier)								
		CHILLS	RUNNY NOSE	HEADACHE	FEVER	FLU			
		YES	NO	MILD	YES	NO			
		YES	YES	NO	NO	YES			
Q 7		YES	NO	STRONG	YES	YES		10	CO3
		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 advanta	ges and disadvanta	ages of using san	npling to re	duce the nur	nber of		

<b>V</b>	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	10	COA

underlying difference between these two methods?

	Explain the terr	ns: a) Model e	OR valuation b) I	Model Validati	on		
	c) Model Deployment d) Model Performance Write a short note on the following. Both short notes should essentially include the						
Q 10	explanations with a) Linear r b) Logistic	ith examples regression regression ote on the followith examples	OR wing. Both short no ayesian Belief Netv	tes should esse		10	CO3
Q11	Explain CRISP	data mining an	d its various phases	s in detail.		10	CO1
	I						
	<del>_</del>		SECTIO				
	Create a comple on the parameter		of the following da	ta set using C	4.5 algorithm (based		
	OUTLOOK	TEMP	HUMIDITY	WIND	DECISION		
	Sunny	Hot	High	Weak	No		
	Sunny	Hot	High	Strong	No		
	Overcast	Hot	High	Weak	Yes		
	Rain	Mild	High	Weak	Yes		
	Rain	Cool	Normal	Weak	Yes		
Q 12	Rain	Cool	Normal	Strong	No	20	CO3
	Overcast	Cool	Normal	Strong	Yes	20	COS
	Sunny	Mild	High	Weak	No		
	Sunny	Cool	Normal	Weak	Yes		
	Rain	Mild	Normal	Weak	Yes		
	Sunny	Mild	Normal	Strong	Yes		
	Overcast	Mild	High	Strong	Yes		
	Overcast	Hot	Normal	Weak	Yes		
	Rain	Mild	High	Strong	No		
			OR				
	a) Find out all f	requent item set	ts in the follwing da	ta set using $A_{I}$	priori Algorithm with		
	$min\_sup = 30\%$	=		0.1	<b>O</b>		
	b) Find all asso	ciation rules wi	th more than 80% c	confidence.			
		T 7T	ransaction ID   Items B	tought			
		1	ransaction ID I I I I I I I I I I I I I I I I I I				
			0024 $\{a, b, c, c\}$	e}			
			$ \begin{array}{c c} 0012 & \{a, b, d, \\ 0031 & \{a, c, d, \\ \end{array} $			20	CO3
	$ \begin{array}{c c} 0031 & \{a, c, d, e\} \\ 0015 & \{b, c, e\} \end{array} $						
Q 12	$0022 \qquad \{b,d,e\}$						
Q 12	$ \begin{array}{c c} 0029 & \{c,d\} \\ 0040 & \{a,b,c\} \end{array} $						
	$0033 \qquad \{a,d,e\}$						
	$0038 \qquad \qquad \{a,b,e\}$						