UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
Online End Semester Examination, May 2021

Course: Data Mining and Prediction Modeling	Semester: VI
Program: B. Tech. (ECRA+BFSI+CSF+IoT+GG+OG)	Time : 03 hours
Course Code: CSBA3001P	Max. Marks: 100

SECTION A

1. Each Question will carry 5 Marks
2. Instruction: Complete the statement / Select the correct answer(s)

Q1	Which Algorithm supports both symbolic and numeric outputs? A. Quest and Decision List B. C&R Tree and CHAID C. Quest and C&R Tree D. Decision list and CHAID	**CO4**
Q2	Supervised learning is basically a synonym for A. Regression B. Classification C. Clustering D. Active Learning	**CO1**
Q3	In which mining, the goal is to find temporal associations between events. A. Sequential Pattern B. Association Rule C. Co relation D. Causal	**CO2**
Q4	The degree to which the model meets the business objectives is associated in _____ phase of CRISP DM. A. Assess Model B. Evaluation C. Modeling D. Deployment	**CO1**
Q5	Both the data mining process, as a whole and all phases, separately are highly iterative. A. True B. False	**CO1**

Q6	The full form of expand PMML is :
	
	    A. Procedural Model Mark Up Language
	    B. Predictive Model Mark Up Language
	    C. Probability Model Mark Up Language
	    D. Programmable Model Mark Up Language
		CO4

SECTION B

1. Each question will carry 10 marks
2. Instruction: Write short / brief notes

Q7	• Discuss the concept for classification and regression for predictive analysis with proper example. (6) • Also explain the data mining as a step in the process of knowledge discovery (KDD). (4)	**CO1**
Q8	• Define the term Data Visualization. (4) • Explain the 3 types of data visualization techniques in details.(6)	**CO2**
Q9	• What are Bayesian classifiers and explain Naïve Bayesian classification algorithm in detail. (5) • Explain the terminology Clustering and explain the K-means method for clustering with the help of proper algorithm & diagram. (5)	**CO3**
Q10	• What are the metrics for evaluating classifier performance? Discuss the term Confusion Matrix with proper example.(7) • Discuss the terminology Gain Ratio and Gini Index. (3)	**CO4**
Q11	• What are outliers, discuss the types of outliers and explain the challenges of outliers.(10) **OR** • Discuss the applications of Data mining along with challenges in it.And state the differences between different learning techniques. (5+5=10)	**CO2**

Section C

1. Each Question carries 20 Marks.
2. Instruction: Write long answer.

Q12	Explain the structure of biological neuron and ANN. In which manner biological neurons are relevant to ANN? Explain. And in a neural network the input(s) are given by I1= 0.2, I2=0.1, I3=0.4 and associated weights are respectively W1=0.1, W2=0.4,	**CO3**

	W3=0.5, calculate the net input. And define market basket analysis. (5+5+5+5)=20 **OR** How are decision trees used for classification and why decision tree classifiers are so popular, explain it with proper diagrammatic example, and explain basic algorithm for inducing a Decision tree from training tuples? (4+6+10)=20	**CO3**