


Name: Enrolment No:		
UNIVERSITY OF PETROLEUM AND ENERGY STUDIES Online End Semester Examination, May 2021		
Course: Data Mining and Prediction Modeling Program: B. Tech. (BFSI+CSF+IoT+GG+OG) Course Code: CSBA3001P	Semester: VI Time : 03 hours Max. Marks: 100	
SECTION A 1. Each Question will carry 5 Marks 2. Instruction: Complete the statement / Select the correct answer(s)		
Q1	Which chat measure the effectiveness of a predictive model calculated as the ration between the results obtained with and without the predictive model? A. Gain B. ROI C. Lift D. Profit	CO4
Q2	Which dataset is used to train or build a model? A. Training Set. B. Validation Set. C. Test Set D. None of these.	CO1
Q3	Outlier detection is also known as anomaly detection. A. True B. False	CO2
Q4	_____ studies the collection, analysis, interpretation or explanation, and presentation of data. A. Models B. Machine Learning C. Statistics D. Information retrieval	CO1
Q5	Both the data mining process, as a whole and all phases, separately are highly iterative. A. True B. False	CO1

Q6	<p>_____ is a measure of how well the model co relates an outcome with the attributes in the data that has been provided.</p> <p>A. Accuracy B. Reliability C. Usefulness D. Robust</p>	CO4
<p>SECTION B</p> <p>1. Each question will carry 10 marks 2. Instruction: Write short / brief notes</p>		
Q7	<ul style="list-style-type: none"> • Discuss the concept for classification and data characterization for predictive analysis with proper example. (6) • Describe four challenges to data mining regarding data mining methodology and user interaction issues. (4) 	CO1
Q8	<ul style="list-style-type: none"> • Define the term segmentation and explain how to deal with missing values. (6) • Explain the types of data visualization techniques in details.(4) 	CO2
Q9	<ul style="list-style-type: none"> • How are decision trees used for classification and why decision tree classifiers are so popular? (4) • Explain the terminology discriminant analysis and explain the SVM with the help of proper algorithm & diagram. (6) 	CO3
Q10	<ul style="list-style-type: none"> • Describe the rule induction using CHAID.(6) • Discuss the terminology Gain Ratio and RoI. (4) 	CO4
Q11	<ul style="list-style-type: none"> • In data mining how the data partitioning is done, explain the concept and also explain the challenges of outliers. (5+5=10) <p style="text-align: center;">OR</p> <ul style="list-style-type: none"> • Discuss the steps involved in data understanding and data preparation with the help of suitable diagram. (5+5=10) 	CO2
<p>Section C</p> <p>1. Each Question carries 20 Marks. 2. Instruction: Write long answer. 3. Answer any one question either 12(A) or 12(B).Marks of clearly written for your reference. In 12(A) total 4 questions are there to be answered and in 12(B) total two questions are there to be answered.</p>		
Q12	<p>A. “Which groups or sets of items are customers likely to purchase on a given trip to the store?” to answer the question which type of analysis to be performed and why, do explain it with your justification and diagram wherever needed. (7)</p>	CO3

Draw and explain the structure of ANN and state in which manner biological neurons are relevant to ANN. (5)

Only write the formula for calculating net input in Neural Network. (2)

And explain how association rule mining can be viewed and is it a two-step process? Justify. (6)

OR

B. Explain basic algorithm for inducing a Decision tree from training tuples?

And also explain the K-Means clustering algorithm. (10+10)=20

CO3