

Name:	 UPES UNIVERSITY WITH A PURPOSE
Enrolment No:	

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2019

Course: Data Mining & Prediction Modeling
Program: B.Tech CSE BAO
Course Code: CSBA3001

Semester: V
Time : 03 hrs.
Max. Marks: 100

Instructions:

SECTION A

S. No.		Marks	CO
Q 1	In real world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem	4	CO1
Q 2	What are the major challenges of mining a huge amount of data (millions of tuples) in comparison with mining a small amount of data (hundred tuples)?	4	CO2
Q 3	Explain how the predictive accuracy of classification methods be estimated.	4	CO3
Q 4	What does it mean to deploy a machine learning model?	4	CO4
Q 5	Explain concept of CRISP-data mining with the help of suitable diagram.	4	CO1

SECTION B

Q 6	For a given Symptoms and Diagnosis dataset – Classify whether patient has flu or not for input given below <p style="text-align: center;"><i>Input: - (Yes, No, Mild, yes,?)</i></p> <table border="1" style="margin-left: auto; margin-right: auto;"> <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	10	CO3
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 7	Outliers are often discarded as noise. However, one person’s garbage could be another’s treasure. For example, exceptions in credit card transactions can help us detect the fraudulent use of credit cards. Using fraudulence detection as an example, propose a method that can be used to detect outliers.	10	CO2																																													
Q 8	Write an algorithm for k-nearest neighbor classification given k , the nearest number of neighbors, and n , the number of attributes describing each tuple. OR	10	CO3																																													

