Name: Enrolment No:	UPES UNIVERSITY OF TOMORROW

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2022

Course: Machine learning
Program: BTech CSE
Course Code: CSAI 3013P

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

Instructions:

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Define Supervised Learning framework along with its applications.	4	CO1
Q2	Describe coefficient of determination in linear regression.	4	CO2
Q3	Discuss why we call weak learner in ensemble learning method?	4	CO4
Q4	Why Frequent Pattern Growth Algorithm is Fast ?	4	CO3
Q5	Explain Similarity metric, and Term weighting in information retrieval system.	4	CO4
SECTION B **(4Qx10M= 40 Marks)**			
Q6	Define K-mean clustering with the help of an example.	10	CO3
Q7	Given the training data in the table below (Buy Computer data), predict the class of the following new example using Naïve Bayes classification: age<=30, income=medium, student=yes, credit-rating=fair.	10	CO4

RID	age	income	student	credit_rating	Class: buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 ... 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 ... 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31 ... 40	medium	no	excellent	yes
13	31 ... 40	high	yes	fair	yes
14	>40	medium	no	excellent	no

OR

Explain the information retrieval of twitter services (only describe 2 services).

Q8	Consider the a database, D , consisting of 9 transactions. Suppose min. support count required is 2 (i.e. min_sup = 2/9 = 22 %). Use FP growth algorithm to solve the given example. **TID / List of Items table below**	10	CO3

TID	List of Items
T100	I1, I2, I5
T100	I2, I4
T100	I2, I3
T100	I1, I2, I4
T100	I1, I3
T100	I2, I3
T100	I1, I3
T100	I1, I2 ,I3, I5
T100	I1, I2, I3

Q9	Briefly explain linear regression along with its type.	10	CO2

<div align="center">

SECTION-C
(2Qx20M=40 Marks)

</div>

Q10	a) You are a social researcher interested in the relationship between income and happiness. You survey 10 people whose incomes range from 15k to 75k and ask them to rank their happiness on a scale from 1 to 10. Your independent variable (income) and dependent variable (happiness) are both quantitative, so implement regression analysis on it to see if there is a linear relationship between them or not. b) Explain the concept of line of best fit in linear regression.	10+5	CO2
Q11	Use single and complete link agglomerative clustering to group the data described by the following distance matrix. Show the dendrograms. |	A	B	C	D
---	---	---	---	---
A	0	1	4	5
B		0	2	6
C			0	3
D				0 | **OR** Suppose we have the following dataset that has various transactions, and from this dataset, we need to find the frequent itemsets and generate the association rules using the Apriori algorithm: | Transaction	List of items
---	---
T1	I1,I2,I3
T2	I2,I3,I4
T3	I4,I5
T4	I1,I2,I4
T5	I1,I2,I3,I5
T6	I1,I2,I3,I4 | Given: Support threshold=50% => 0.5*6= 3 => with min_sup=3	20	CO3