Name:

Enrolment No:

UPES
UNIVERSITY WITH A PURPOSE

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

Course: B. Tech CSE+AI/ML **Semester:** IV
Program: Introduction to Machine Learning **Time** : 03 hrs.
Course Code: CSAI2006 **Max. Marks:** 100

Instructions:

SECTION A

1. Each Question will carry 5 Marks

<table>
<tr><th>S. No.</th><th></th><th>Marks</th><th>CO</th></tr>
<tr><td>Q 1</td><td>Define Machine Learning. Write down five application of it.</td><td>5</td><td>CO1</td></tr>
<tr><td>Q 2</td><td>Discuss Similarity Metrics and Term Weighting.</td><td>5</td><td>CO5</td></tr>
<tr><td>Q 3</td><td>Discuss, how the statistics is important for machine learning. Define mean, median, and mode.</td><td>5</td><td>CO2</td></tr>
<tr><td>Q 4</td><td>
<table>
<tr><td>Time Point</td><td>Infosys</td><td>TCS</td></tr>
<tr><td>Jan 2021</td><td>6</td><td>20</td></tr>
<tr><td>Feb 2021</td><td>5</td><td>10</td></tr>
<tr><td>March 2021</td><td>4</td><td>14</td></tr>
<tr><td>April 2021</td><td>3</td><td>5</td></tr>
<tr><td>May 2021</td><td>2</td><td>5</td></tr>
</table>
It is given the average stock price of Infosys and TCS for five consecutive months. Find it either the stock price are independent to each other or not.
</td><td>5</td><td>CO1</td></tr>
<tr><td>Q 5</td><td>Discuss Vector Space Model in information retrieval system.</td><td>5</td><td>CO5</td></tr>
<tr><td>Q 6</td><td>Discuss R-Squared, Adjusted R-Squared, and Sum Squared Error.</td><td>5</td><td>CO2</td></tr>
</table>

SECTION B

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

<table>
<tr><td>Q 7</td><td>Explain and discuss the architecture and process of information retrieval system for Google Search Engine.</td><td>10</td><td>CO5</td></tr>
<tr><td>Q 8</td><td>Discuss and derive the mathematical proof of logistic regression model.</td><td>10</td><td>CO3</td></tr>
<tr><td>Q 9</td><td>
<table>
<tr><td></td><td>X</td><td>Y</td></tr>
<tr><td>A</td><td>2</td><td>6</td></tr>
<tr><td>B</td><td>3</td><td>4</td></tr>
<tr><td>C</td><td>3</td><td>8</td></tr>
<tr><td>D</td><td>4</td><td>7</td></tr>
<tr><td>E</td><td>6</td><td>2</td></tr>
</table>
</td><td>10</td><td>CO4</td></tr>
</table>

	F	6	4		
	G	7	3		
	H	7	4		
	I	8	5		
	J	7	6		
	The above table is containing the 10 different data points. Implement k-means and k-medoid clustering algorithm for k=2 on the given dataset and discuss which model is better with subjective basis.				
Q 10	**Transactional data of AllElectronics** *TID* — *List of item_IDs* T100 — I1, I2, I5 T200 — I2, I4 T300 — I2, I3 T400 — I1, I2, I4 T500 — I1, I3 T600 — I2, I3 T700 — I1, I3 T800 — I1, I2, I3, I5 T900 — I1, I2, I3 Evaluate the most frequent data item set of 3 items using FP-Growth algorithm for the above AllElectronics data set. Assume min_support=2			**10**	**CO3**
Q 11	Outlook / Temperature / Humidity / Windy / Class sunny / hot / high / false / N sunny / hot / high / true / N overcast / hot / high / false / P rain / mild / high / false / P rain / cool / normal / false / P rain / cool / normal / true / N overcast / cool / normal / true / P sunny / mild / high / false / N sunny / cool / normal / false / P rain / mild / normal / false / P sunny / mild / normal / true / P overcast / mild / high / true / P overcast / hot / normal / false / P rain / mild / high / true / N Discuss Bays' Theorem. Write down the steps of Bayesian Classification Algorithm. Based on above data try to predict the label of given instance using Bayesian Classification Algorithm: Outlook=Rain, Temperature= hot, Humidity=High, Windy= false, Class=? **OR** Explain KNN algorithm. Why it is also called Lazy Learner? What are the points to be subjected when choosing the value of k? For the below problem predict for the class of David using KNN and assume the value of k=3.			**10**	**CO4**

Customer	Age	Income (K)	No. of cards	Response
John	35	35	3	Yes
Rachel	22	50	2	No
Ruth	63	200	1	No
Tom	59	170	1	No
Neil	25	40	4	Yes
David	37	50	2	?

SECTION-C

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

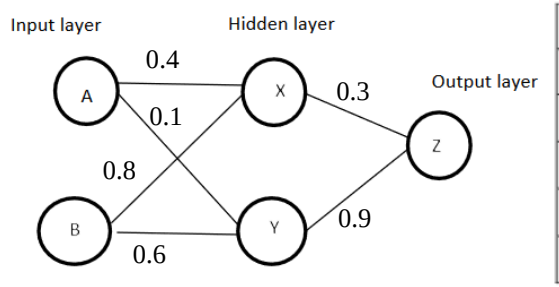
Q 12

(a) "The support vector machine is highly accurate classification method", justify the statement. Explain the SVM model with the support of equation of hyperplane. List down the kernels name with equation used in SVM model. (10)

(b) Discuss and explain DBSCAN algorithm with advantages and limitations. (10)

OR

.



Input		Output
A	B	Z
0	0	0
0	1	1
1	0	1
1	1	1

Learning rate=0.35

Biases are $\theta x = \theta y = \theta z = 0$. Neural Network of above diagram has two nodes (A,B) in the input layer, two nodes in the hidden layer (X,Y)and one node in the output layer (Z). The values given to weights are taken randomly and will be changed during back propagation iterations. Initial weights of the top input nodes taken at random are 0.4, 0.1 .Weights of bottom input node are 0.8 and 0.6. Weights of top hidden node is 0.3 and that of bottom hidden node is 0.9. Assume the number of iterations are two.

20 **CO4**