Name:
Enrolment No:

UPES
UNIVERSITY WITH A PURPOSE

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2021

Course: Machine Learning **Semester: 7th**
Program: B. Tech-CSE-GG **Max. Marks: 100**
Course Code: CSAI2001P **Duration: 03 hrs.**

SECTION A (5Qx 4M = 20 Marks)

Note: This section have 5 Questions (short answer type) of 4 marks each.
All the questions shall be compulsory.

S. No.		Marks	CO
Q 1	Differentiate between Supervised learning, unsupervised learning and reinforcement learning.	4	CO1
Q2	State the difference between Classification and Clustering?	4	CO4
Q3	Contrast the difference between linear regression and logistic regression.	4	CO3
Q4	Discuss the Density-Based Clustering.	4	CO4
Q5	Describe the characteristics of Backpropagation in Multilayer Feed-Forward Neural Networks	4	CO1

SECTION B (4Qx 10M = 40 Marks)

Note: This section have 4 Questions of 10 marks each. Q6 - Q8 are compulsory and Q9 has internal choice to attempt any one.

Q 6	a) How could you divide the 'training Set' and 'test Set' in a Machine Learning Model? How much data will you allocate for training, validation, and test Sets? b) Explain why k-fold cross validation does not work well with time series model? What can you do about it?	5+5	CO1
Q 7	Describe why Nave Bayes classifier is so powerful for text classification and why Normalization is required in machine learning?	10	CO3
Q8	In which algorithm, Ginni index is used. Explain the algorithm in detail with suitable example and also discuss that why the decision tree suffer often with overfitting problem?	10	CO3
Q9	a) Summarize the goal of SVM and also discuss how to select the margin for the given data for the sales (in million dollars) of Car of an Automobile Company for 6 consecutive years.	10	CO2

Year	2013	2014	2015	2016	2017	2018
Sales	110	100	250	275	230	300

Based on the above data, predict the sales for next three consecutive years.

OR

A data set is given to you about utilities fraud detection. You have built a classifier model and achieved a performance score of 98.5%. Is this a good model? If yes, justify. If not, what can you do about it?

SECTION-C **(2Qx 20M= 40 Marks)**

Note: Q11 has internal choice to attempt any one.

Q10	Discuss Entropy in ID3 algorithm with an example 5. Compare Entropy and Information Gain in ID3 with an example.	**20**	**CO4**
Q11	Use the k-means algorithm and Euclidean distance to cluster the following 8 examples into 3 clusters: A1=(2,10),A2=(2,5),A3=(8,4),A4=(5,8),A5=(7,5),A6=(6,4),A7=(1,2),A8=(4,9) A1=(2,10),A2=(2,5),A3=(8,4),A4=(5,8),A5=(7,5),A6=(6,4),A7=(1,2),A8=(4,9). Suppose that the initial seeds (centers of each cluster) are A1A1, A4A4 and A7A7. Run the k-means algorithm for 1 epoch only. At the end of this epoch show: **a)** The new clusters (i.e. the examples belonging to each cluster) **b)** The centers of the new clusters **c)** Draw a 10 by 10 space with all the 8 points and show the clusters after the first epoch and the new centroids. **d)** How many more iterations are needed to converge? Draw the result for each epoch. **OR** In which approach, multiple models or 'weak learners' are trained to rectify the same problem and integrated to gain desired results. Weak models combined rightly give accurate models. Explain it with its different types.	**20**	**CO5**