Enrolment No: _____

Name of Student: _____

Department/ School: _____

SUPPLEMENTARY EXAMINATION ODD SEMESTER 2022-23

COURSE CODE: CSET211	**MAX. DURATION**	3 Hours
COURSE TITLE: STATISTICAL MACHINE LEARNING		
COURSE CREDIT: 4(3-0-2)	**TOTAL MARKS**	100

GENERAL INSTRUCTIONS: -

1. Do not write anything on the question paper except **name, enrolment number** and **department/school.**
2. All Questions are Compulsory.
3. Use of Scientific Calculator is allowed.

Note: If require any missing data; then choose suitably

1) Discuss Bayes' theorem and its application in the Naive Bayes Classifier.

[10 Marks]

2) Explain how reinforcement learning differs from unsupervised learning and its concepts with proper example.

[10 Marks]

3) Examine the non-parametric nature of the KNN algorithm and whether feature scaling is necessary. Provide reasoning for your conclusions.

[10 Marks]

4) Detail the process for determining the principal components of a given data set.

[10 Marks]

5) Identify and explain two scenarios in which K-means clustering may not produce optimal results.

[10 Marks]

6) Draw the decision tree for the following training data:

[10 Marks]

Gender	Car ownership	Travel cost	Income Level	Transportation (Class)
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

7) Examine the advantages and limitations of the Random Forest algorithm.

[10 marks]

8) Describe the k-NN algorithm, including its method of operation, the distance metric used and the procedure for selecting the k nearest neighbours.

[10 marks]

9) Explain the kernel trick, including its use in addressing non-linear classification issues with SVM.

[10 marks]

10) Provide a summary report on implementing a non-linear SVM classifier.

[10 marks]