

Registration Number:

Date & session:

ST.JOSEPH'S UNIVERSITY, BENGALURU -27 M.Sc. – III SEMESTER SEMESTER EXAMINATION: OCTOBER 2022 (Examination conducted in December 2022) ST 9220 –DATA MINING AND MACHINE LEARNING

Time: 2 ½ Hours Max Marks: 70

This paper contains TWO printed pages and TWO parts.
Usage of Scientific Calculator is allowed.

PART-A

Answer any 6 from the following.

3*6 = 18

- 1. Define the three types of machine learning with examples.
- 2. What are Training, Testing and Validation sets? How is it useful?
- 3. What is numpy, scipy and matplotlib. Illustrate one code from each.
- 4. Differentiate between lable encoding and one-hot encoding with an example.
- 5. Illustrate k fold cross validation.
- 6. Define Logistic regression. How is it useful?
- 7. Differentiate between Tree based Regression model and Linear Regression model.
- 8. Draw a decision tree, label and describe the parts of the same.

PART-B

Answer any FOUR from the following

13*4 = 52

- 9. a) Explain Bias Variance Trade off with an example.
 - b) What is a balanced and imbalanced dataset? Explain with an example
 - c) Outline the algorithm of K means Clustering.

(5+3+5)

- 10. a) What is interaction effect? Explain the procedure to obtain interaction plots and hence interpret it.
 - b) Why does one need logistic regression? Explain with an example.

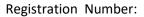
(8+5)

- 11. a) Describe polynomial regression and regression splines.
 - b) Describe the algorithm of adaboost with the help of an example.

(5+8)

- 12. a) What is a weak learner? Where is it used?
 - b) Give a brief overview of Maximal Margin Classifier with the mathematical model.
 - c) Differentiate between Tree based and Linear Regression model with an example.

(3+7+3)





Date & Session

13. Identify the best split for your decision tree from the data below.

student

Credit rating

	(13)
uys	
ouse	
0	
р	
es	
es	
es	
0	
es	
0	
es	
es	

				i iouse
<25	High	No	Fair	No
<25	High	No	Excellent	Np
25-35	High	No	Fair	Yes
>35	Medium	No	Fair	Yes
>35	Low	Yes	Fair	Yes
>35	Low	Yes	Excellent	No
25-35	Low	Yes	Excellent	Yes
<25	Medium	No	Fair	No
<25	Low	Yes	Fair	Yes
>35	Medium	Yes	Fair	Yes
<25	Medium	Yes	Excellent	Yes
25-35	Medium	No	Excellent	Yes
25-35	High	Yes	Fair	Yes
>35	Medium	No	Excellent	No

- 14. a) Describe any three challenges pertaining to machine learning.
 - b) Write the algorithm for gradient boosting.

income

age

c) Write the algorithm for Hierarchical Clustering.

(3+5+5)