



Registration Number:

Date & session:

ST. JOSEPH'S UNIVERSITY, BENGALURU -27
M. Sc (STATISTICS) – III SEMESTER
SEMESTER EXAMINATION: OCTOBER 2023
(Examination conducted in November/December 2022)
ST 9220 – Data Mining and Machine Learning

Time: 2 Hours

Max Marks: 50

This paper contains ONE printed page and ONE part

I Answer any FIVE of the following

10 X 5 = 50

1. A) What are the various challenges associated with Machine Learning, illustrate with examples. (4)
B) Differentiate between training, testing and validation set. (3)
C) Define Bias, Variance and also comment upon Bias Variance Trade off. (3)
2. A) What are the different types of missing values, illustrate with a suitable example. (5)
B) Differentiate between One hot encoding and Label Encoding (3)
C) What is data preprocessing and why is it needed in machine learning. (2)
3. A) Define and explain the following performance metrics for regression: Mean Squared Error (MSE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC). How are these metrics useful for model evaluation and selection? (4)
B) Define confusion matrix and explain various components of the same (4)
C) What does one mean by i) Transformation, ii) interaction (2)
4. A) What is curse of dimensionality and describe any 2 methods to resolve it. (5)
B) What is cross validation, describe any two types of cross validation (5)
5. A) Give a brief overview of Logistic regression. (5)
B) Give a brief overview of SVM in classification and also provide suitable applications in real life scenario. (5)
6. A) Explain the procedure of K means Clustering. (5)
B) What are ensemble methods? Briefly explain any 2 of them. (5)
7. A) What are the various parts of a decision tree? Describe the 3 ways to identify the best split of a decision tree. (5)
B) Give a brief overview of Hierarchical Clustering and Gaussian Mixture Models. (5)