



23002945

QP CODE: 23002945

Reg No :

Name :

M Sc DEGREE (CSS) EXAMINATION, MARCH 2023

Third Semester

Faculty of Science

M Sc Artificial Intelligence

CORE - AI010302 - INTRODUCTION TO DATA ANALYTICS

2020 ADMISSION ONWARDS

A7108A32

Time: 3 Hours

Weightage: 30

Part A (Short Answer Questions)

*Answer any **eight** questions.*

Weight 1 each.

1. What are the six steps of Data Science process?
2. What is the meaning of Parameter tuning?
3. Discuss about R2 Statistics.
4. Discuss about mean squared error
5. Discuss about ROC curves.
6. Explain about Linear discriminant functions
7. Discuss about clustering and classification.
8. What is ment by PCA?
9. Discuss about support vector machine.
10. Discuss about the application of Decision tree.

(8×1=8 weightage)





Part B (Short Essay/Problems)

*Answer any **six** questions.*

Weight 2 each.

11. What are the different methods available for cleaning the collected data?
12. What are the functionalities of data mining?
13. Explain about different types of correlation coefficients.
14. Explain about feature selection method in regression
15. Explain Lazy Learners mechanism taking k-Nearest-Neighbor classifiers as example.
16. Explain about regression tree.
17. Compare Canonical correlation analysis and Principal component analysis
18. How bagging improves the accuracy of classification?

(6×2=12 weightage)

Part C (Essay Type Questions)

*Answer any **two** questions.*

Weight 5 each.

19. a) Discuss about the application area of data science b) Explain about different facts of data.
20. Explain the terms a) Multiple Linear Regression b) Polynomial regression c) Regularization
21. Use Bayesian classification to predict class label using the training data given below. The tuple to be classified is $X=(\text{Age}=\text{Youth}, \text{Income}=\text{Medium}, \text{Student}=\text{Yes}, \text{Credit-rating}=\text{fair})$
Class-Labeled Training Tuples from the AllElectronics Customer Database





RID	Age	Income	Student	Credit-rating	Class: buys_computer
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle_aged	High	No	Fair	yes
4	Senior	Medium	No	Fair	yes
5	Senior	Low	yes	Fair	yes
6	Senior	Low	yes	Excellent	no
7	Middle_aged	Low	yes	Excellent	yes
8	Youth	Medium	No	Fair	no
9	Youth	Low	yes	Fair	yes
10	Senior	Medium	yes	Fair	yes
11	Youth	Medium	yes	Excellent	yes
12	Middle_aged	Medium	No	Excellent	yes
13	Middle_aged	High	yes	Fair	yes
14	Senior	Medium	No	Excellent	no

22. Given the following data, use PCA to reduce dimension

X	2	1	0	-1
Y	4	3	1	0.5

(2×5=10 weightage)

