Reg. No ....   Name **23U306**

END SEMESTER EXAMINATION : NOVEMBER 2023
SEMESTER 3 : INTEGRATED M.Sc. PROGRAMME COMPUTER SCIENCE - DATA SCIENCE
COURSE : **21UP3CRMCP7 : INTRODUCTION TO DATA SCIENCE**
(For Regular 2022 Admission and Improvement/Supplementary 2021 Admission)

Time : Three Hours   Max. Weightage: 30

PART A
Answer any 8 questions

1. List any two statistical methods in python that can be applied to a dataset in data exploration.
2. Mention a real-world example of a population and its sample.
3. List any two examples of optimization algorithms.
4. Data Science evolved initially with the idea of mapping _____ with computer science.
5. List any two cases where a heatmap can be employed to visualize data.
6. List any two metrics that are used in calculating the accuracy of a recommendation system.
7. State an advantage of incorporating data science technology in government sector.
8. Mention an example of a system that employs SVD in its datasets.
9. Define sanity checking in exploration of data.
10. State the main challenge in using a machine learning algorithm.

(1 x 8 = 8 Weight)

PART B
Answer any 6 questions

11. Discuss the role of a data scientist in the data science process.
12. Many statisticians and researchers have made remarkable contributions in the evolution of data science. Discuss the contributions made by John Tukey.
13. Using a bipartite graph, explain how a real world recommendation engine works.
14. Differentiate between Decision Tree and Random Forest algorithms.
15. Define the concept – fitting a model. Also, state when does a model overfit.
16. With the following data, calculate the median and mode:

S	L
0	7 8
1	3 3 5 7
2	2 4 5 7 8
3	2 2 2 6
4	3

17. Discuss the various categories of data visualization.
18. Discuss the benefits and drawbacks of Random Forest algorithm.

(2 x 6 = 12 Weight)

PART C
Answer any 2 questions

19. Using a suitable statistical method, test whether the variables Educational Qualification and Marital Status are related to each other, based on the following data:

Qualification Marital Status	Middle School	High School	Bachelor's	Master's	Ph.D.	Total
Never Married	18	36	21	9	6	90
Married	12	36	45	36	21	150
Divorced	6	9	9	3	3	30
Widowed	3	9	9	6	3	30
Total	39	90	84	54	33	300

20. Make a detailed note on the tools used for exploratory data analysis with bivariate data.

21. Given below is the sample representing the relationship between GDP and vehicle sale in each year. Predict the sales in 2017 if the GDP rate was 8.12. Also, calculate the accuracy of the model.

Year	2011	2012	2013	2014	2015	2016
GDP	6.2	6.5	5.48	6.54	7.18	7.93
Vehicle Sales (in lakhs)	26.3	26.65	25.03	26.01	27.9	30.47

22. Decompose the representative matrix $\begin{bmatrix} -4 & -7 \\ 1 & 4 \end{bmatrix}$ into product of three matrices.

(5 x 2 = 10 Weight)