Reg. No ..... Name **24U306**

END SEMESTER EXAMINATION - OCTOBER 2024

SEMESTER 3 : INTEGRATED M.Sc. PROGRAMME COMPUTER SCIENCE - DATA SCIENCE

COURSE : **21UP3CRMCP07 : INTRODUCTION TO DATA SCIENCE**

(For Regular 2023 Admission and Improvement/Supplementary 2022/2021 Admissions)

Time : Three Hours Max. Weightage: 30

PART A

Answer any 8 questions

1. State any two constituent technologies of data science.
2. Name any two machine learning algorithms used for clustering.
3. State the main challenge in using a machine learning algorithm.
4. Mention a real-world example of a population and its sample.
5. A decision tree algorithm supports _____ learning.
6. Define the term – bigdata.
7. Define the term 'coverage' with respect to a recommendation system.
8. List any two methods that are used in dimensionality reduction.
9. List any two feature selection methods.
10. Define data munging.

(1 x 8 = 8 Weight)

PART B

Answer any 6 questions

11. Given the following data, calculate the mean of the data and form the stem-and-leaf plot: [72, 49, 62, 58, 73, 55, 78, 83, 57, 63, 73, 73, 75, 85, 85, 64, 61, 67, 75, 91]
12. Discuss the importance of performing data exploration in the data science process, prior to creating a model.
13. Define Eigen Values and Eigen Vectors.
14. Define collaborative filtering and explain its classifications.
15. Differentiate between descriptive statistics and inferential statistics in data analysis.
16. List any four reasons signifying the importance of feature selection.
17. Prepare short notes on Random Forest algorithm.
18. Brief the relevance of statistical analysis and define the term 'sample' in statistical analysis.

(2 x 6 = 12 Weight)

PART C

Answer any 2 questions

19. Given below is the sample selected from a dataset income.csv that shows the relationship between income and happiness quotients. Predict the happiness quotient if the income quotient is 6.93.

Income	3.86	4.97	4.92	3.21	7.19	3.72	4.67	4.49	3.12
Happiness	2.31	3.43	4.59	2.79	5.59	2.45	3.19	1.90	2.94

20. With the following data, find the feature that best splits the dataset and form a decision tree based on the best splitting attribute as the root node.

Weekend	Weather	Parents	Money	Decision
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay In
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

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

22. The following data shows the sales and advertising expenses of an Italian clothing company. If the amount spent on advertising in year 2010 is 62 million Euro, predict the Sales expenses in the year.

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
Sales (Million Euro)	651	762	856	1063	1190	1298	1421	1440	1518
Advertising (Million Euro	23	26	30	34	43	48	52	57	58

(5 x 2 = 10 Weight)