

END SEMESTER EXAMINATION - OCTOBER 2025**SEMESTER 3 : INTEGRATED M.Sc. PROGRAMME COMPUTER SCIENCE - DATA SCIENCE****COURSE : 21UP3CRMCP07 : INTRODUCTION TO DATA SCIENCE***(For Regular - 2024 Admission and Improvement / Supplementary 2023/2022/2021 Admissions)*

Time : Three Hours

Max. Weightage: 30

PART A**Answer any 8 questions**

1. State any two methods that are used in recommendation systems.
2. Define the term 'population' with respect to bigdata.
3. State any two advantages of SVD in feature selection.
4. List any two feature selection methods.
5. List any two applications of data science on the mobile platform where a consumer is being datafied with their consent.
6. List any two commonly used metrics that are used to test the accuracy of a machine learning algorithm.
7. List any two cases where a heatmap can be employed to visualize data.
8. Define the term – bigdata.
9. Write the formula to calculate the Chi-Square score to conduct hypotheses test.
10. State the need of a machine learning algorithm in data science.

(1 x 8 = 8 weight)**PART B****Answer any 6 questions**

11. 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

12. State any four benefits of reducing the dimensions of a dataset.
13. Many statisticians and researchers have made remarkable contributions in the evolution of data science. Discuss the contributions made by John Tukey.
14. Discuss briefly about any two daily life applications where data science process is involved.
15. Differentiate between supervised and unsupervised learning with an example for each.
16. Explain bagging technique with an example.
17. List down the steps involved in implementing a Random Forest Algorithm.
18. Define a recommendation system in brief.

(2 x 6 = 12 Weight)

PART C
Answer any 2 questions

19. 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

20. Use SVD to decompose the matrix $\begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$ as the product of three matrices.

21. With the following data, find the feature that best splits the dataset. Form a decision tree based on the best splitting attribute as the root node. Use ID3 algorithm for building the tree.

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes

22. Make a detailed note on the tools used for exploratory data analysis with bivariate data.
(5 x 2 = 10 Weight)