

--	--	--	--	--	--	--	--

St Aloysius (Deemed to be University)

Mangaluru

SCHOOL OF INFORMATION TECHNOLOGY (PG PROGRAMME)

Semester II – P. G. Examination – M.Sc. DATA SCIENCE

April/May - 2026

MULTIVARIATE TECHNIQUES FOR DATA SCIENCE

Time: 2½ Hours

Max. Marks: 60

Note: i) Answer all the questions

ii) Draw diagrams wherever necessary

PART - A

Answer any FIVE of the following.

(5×2=10)

1. What is meant by interval scale?
2. What is meant by multivariate analysis?
3. Define partial correlation. Write the formula to find partial correlation.
4. Define correlation. What is the range of Pearson correlation coefficient?
5. What is the purpose of multivariate data analysis?
6. Define statistical hypothesis.
7. What is MANOVA?

PART - B

Answer any SIX of the following.

(6×5=30)

8. Discuss any two techniques used for multivariate analysis.
9. How to assess the need for a transformation? Explain.
10. Explain in detail about the techniques for data entry.
11. Explain correlation analysis. Describe graphical representation of correlation and its interpretation.
12. Mention briefly the purpose and use of factor analysis.
13. Memory capacity of students was tested before and after giving the nourishing food (CHAVANPRAKASH), State whether CHAVANPRAKASH was effective or not from the following scores. [$t_{0.05}=2.26$]

Roll No:	1	2	3	4	5	6	7	8	9	10
Before	12	14	11	8	7	10	3	0	5	6
After	15	16	10	7	5	12	10	2	3	8

14. Marks obtained by six students in three subjects are given in the following data matrix, with rows as variables. Calculate mean vector and covariance matrix.

$$X = \begin{bmatrix} 41 & 37 & 43 & 65 & 71 & 58 \\ 68 & 45 & 39 & 53 & 60 & 55 \\ 58 & 27 & 79 & 81 & 80 & 70 \end{bmatrix}$$

Contd...2

15. An insurance agent has claimed that the average age of policy holders who insure through him is less than the average for all agents which is 30.5 years. A random sample of 100 policy holders who had insured through him gave the following age distribution.

Age last Birthday (years)	16 - 20	21 - 25	26 - 30	31 - 35	36 - 40	Total
Number of persons insured	12	22	20	30	16	100

Calculate the arithmetic mean and standard deviation of this distribution and use these values to test his claim at 5% level of significance.

PART - C

Answer any TWO of the following. (2×10=20)

16. Discuss Hotelling T^2 , distribution. Evaluate T^2 , for testing $H_0 : \mu_0 = [7, 11]$, using the data Specify the distribution of T^2

$$X = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix}$$

17. Following are the marks in MCA, entrance test and interview conducted by a company. All marks are out of 100. Find the regression line of marks in MCA on test and interview marks. Estimate the marks in MCA who has obtained 68 and 70 in Test and Interview respectively.

MCA	70	72	80	90	85
Test	65	64	70	71	69
Interview	62	60	76	74	72

18. Three different techniques namely medication, exercises and special diet are randomly assigned to (individuals diagnosed with high blood pressure) lower the blood pressure. After four weeks the reduction in each person’s blood pressure is recorded. Test at 5% level, whether there is significant difference in mean reduction of blood pressure among the three techniques.

$[F_{2,12}=3.8853]$.

Medication	10	12	9	15	13
Exercise	8	3	0	2	6
Diet	5	9	12	8	4

Mapping of Questions with Bloom’s Level and Course Outcomes (COs)

Part	A							B								C		
Q. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Blooms Level	1	1	2	1	1	1	1	2	2	2	2	1	3	3	3	3	3	3
Mapped CO(s)	1	5	4	2	2	3	4	5	1	1	2	5	3	4	3	4	2	3

IT3IPHC551

Reg. No.:

--	--	--	--	--	--	--	--

St Aloysius (Deemed to be University)

Mangaluru

SCHOOL OF INFORMATION TECHNOLOGY (PG PROGRAMME)

Semester II – P. G. Examination – M.Sc. DATA SCIENCE

April/May - 2026

MACHINE LEARNING ALGORITHMS

Time: 2½ Hours

Max. Marks: 60

Note: i) Answer all the questions

ii) Draw diagrams wherever necessary

PART - A

Answer any **FIVE** of the following.

(5×2=10)

1. What are the three basic parts of the Machine Learning process?
2. Why is data important in Machine Learning? Mention any two reasons.
3. What is Linear Regression?
4. Define Hierarchical Clustering.
5. What is meant by Reward in Reinforcement Learning?
6. What is meant by Resampling in machine learning?
7. What is a Recurrent Neural Network (RNN)?

PART - B

Answer any **SIX** of the following.

(6×5=30)

8. Discuss the benefits and challenges of Machine Learning.
9. Discuss the advantages and disadvantages of Pooling Layer.
10. Write a short note on ANN and CNN.
11. Differentiate the Frozen and Trainable Layers in Transfer Learning with context of the similarity and size of your target dataset.
12. Define Classification. Explain with its types.
13. Define Representation Learning. Explain why it's important.
14. Differentiate the Machine Learning and Deep Learning.
15. Define Autoencoders. Explain its different types.

PART - C

Answer any **TWO** of the following.

(2×10=20)

16. Abbreviate PCA. Explain how it can be used as Dimensionality reduction method.
17. Explain the need of Rule Extraction from Trees with its general steps.
18. Assessment process of a Classification Algorithm is necessary. Justify with the different reasons.

Mapping of Questions with Bloom's Level and Course Outcomes (COs)

Part	A							B								C		
Q. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Blooms Level	1	4	1	1	1	1	1	2	2	1	2	1	1	2	3	2	1	5
Mapped CO(s)	1	2	1	3	5	4	2	5	3	1	1	5	4	5	2	2	3	2

IT3IPHC552

Reg. No.:

--	--	--	--	--	--	--	--

St Aloysius (Deemed to be University)

Mangaluru

SCHOOL OF INFORMATION TECHNOLOGY (PG PROGRAMME)

Semester II – P. G. Examination – M.Sc. DATA SCIENCE

April/May - 2026

ENABLING TECHNOLOGIES FOR DATA SCIENCE

Time: 2½ Hours

Max. Marks: 60

Note: i) Answer all the questions

ii) Draw diagrams wherever necessary

PART - A

Answer any FIVE of the following.

(5×2=10)

1. Define Big Data.
2. What is unstructured data?
3. What is ETL?
4. Differentiate between OLAP and OLTP
5. Define data visualization.
6. Define Data Compression.
7. Differentiate between Prediction and Classification.

PART - B

Answer any SIX of the following.

(6×5=30)

8. What are the different types of Joins? Demonstrate the types of Joins in Pandas by creating two dictionaries containing Employee data.
9. What is the main purpose of Data Mining? Explain in detail.
10. Explain CAP and BASE Theorem.
11. Design MOLAP architecture and explain its advantages and disadvantages.
12. Detail the comprehensive steps involved in data pre-processing and explain their significance in transforming raw data into a suitable format for advanced analysis and machine learning applications.
13. What are the techniques to handle the Noisy data? Explain with example.
14. Discuss the concept of stratified sampling and quota sampling. Give examples.
15. What is a Pandas Series? What is the difference between a DataFrame and a Series in Pandas? Provide an example of creating a DataFrame using Pandas.

PART - C

Answer any TWO of the following.

(2×10=20)

16. Compare different classification methods such as Decision Tree, Naïve Bayes, and Support Vector Machine.
17. Illustrate ROLAP, MOLAP, and HOLAP architectures and also identify their advantages and disadvantages
18. Explain the concept of data cube and the following operations **i) Rollup**
ii) Drill down **iii) Slice & Dice** **iv) Pivot** by giving examples.

Contd...2

Mapping of Questions with Bloom’s Level and Course Outcomes (COs)

Part	A							B								C		
Q. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Blooms Level	2	1	1	3	2	2	3	1	1	2	5	4	2	3	2	3	4	3
Mapped CO(s)	1	1	2	2	4	3	5	4	2	1	2	3	3	3	4	5	2	2

IT3IPSC571a

Reg. No.:

--	--	--	--	--	--	--	--

St Aloysius (Deemed to be University)

Mangaluru

SCHOOL OF INFORMATION TECHNOLOGY (PG PROGRAMME)

Semester II – P. G. Examination – M.Sc. DATA SCIENCE

April/May - 2026

NATURAL LANGUAGE PROCESSING

Time: 2½ Hours

Max. Marks: 60

Note: i) Answer all the questions

ii) Draw diagrams wherever necessary

PART - A

Answer any FIVE of the following.

(5×2=10)

1. Mention two disciplines that contributed to the origin of NLP.
2. What is Lesk's Algorithm?
3. State the role of discourse markers in text processing.
4. Define content selection in NLG.
5. What is sentence ranking?
6. What is meant by free word order? Give one example.
7. What is the role of a discourse planner?

PART - B

Answer any SIX of the following.

(6×5=30)

8. Explain the relationship between language and knowledge representation.
9. Explain the principle of compositionality with example.
10. Explain number and gender agreement with examples.
11. Describe the transformation from conceptual representation to final output.
12. Differentiate between Information Retrieval and QA systems.
13. Analyze the role of POS tagging in Indian language processing.
14. Explain the concept of constituency using a parse tree example.
15. Explain different types of discourse relations with examples.

PART - C

Answer any TWO of the following.

(2×10=20)

16. Define POS tagging. Compare rule-based and statistical POS tagging approaches.
17. Consider the following Context-Free Grammar (CFG):
S → NP VP
NP → Det N
VP → V NP
Det → the
N → boy | ball
V → kicked
Using the above grammar, demonstrate how bottom-up (shift-reduce) parsing constructs the parse tree for the sentence: **"the boy kicked the ball"**

Contd...2

18. Given the following paragraph, identify all types of references and resolve them with proper justification. Also specify the type of reference phenomenon involved.

Paragraph:” Ravi bought a new laptop yesterday. The laptop was expensive, but he needed it for his project. Before he submitted it, he checked everything carefully. The presentation impressed the professor. It was well structured”.

Mapping of Questions with Bloom’s Level and Course Outcomes (COs)

Part	A							B								C		
Q. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Blooms Level	2	1	1	1	1	1	2	2	2	1	4	2	4	2	2	2	3	3
Mapped CO(s)	1	2	3	4	4	1	3	1	2	2	3	2	2	2	3	2	2	3

--	--	--	--	--	--	--	--

St Aloysius (Deemed to be University)**Mangaluru****School of Information Science & Technology (PG Programme)****Semester II – P.G. Examination****April -2026****OPEN ELECTIVE****ENTERPRISE INFORMATION SYSTEM****Time: 1½ Hours****Max Marks: 30****SECTION – A****Answer any FOUR of the following.****(4x6=24)**

1. Define MIS. Briefly explain the different types of Information Systems.
2. What are the expected features in a MIS for it to be effective?
3. What is Decision making? What are the differences between Problem Solving and Decision making?
4. A software company is facing repeated delays in completing client projects. The project manager collects information on task timelines, employee workload, and technical challenges. Several alternatives are discussed, including hiring temporary staff, extending deadlines, or reallocating tasks. Finally, the manager decides to redistribute tasks among existing team members. Identify actions corresponding to each phase of Simon's decision-making model.
5. What is a Group Decision Support system? Explain the different types of GDSS with their characteristic features.
6. Identify the type of decision and justify your answer
 - a. A university decides to open a new campus in another state.
 - b. The examination committee prepares the semester examination timetable
 - c. A company decides to enter the international market within the next three years.
 - d. A nurse decides the order in which patients will be attended during a shift.
 - e. A software company decides to shift from a monolithic system to microservices architecture.
 - f. A cashier decides to process refunds for returned items.

SECTION – B**Answer the Following: (Compulsory)****(6x1=6)****7. Case Study:****Walmart – Supply Chain Optimization with Retail Link**

Walmart's Retail Link system revolutionized its supply chain by enabling real-time data sharing with suppliers. This vendor-managed inventory system allowed suppliers to monitor sales data and manage inventory levels

Contd...2

proactively. The transparency and collaboration facilitated by Retail Link led to improved inventory turnover, reduced stockouts, and enhanced supplier relationships. Walmart's supply chain efficiency became a key competitive advantage in the retail industry.

Questions:

- a. What business need or problem led to the adoption of the enterprise system in this case?
- b. What were the key outcomes or results—both positive and negative—of the system implementation?

Mapping of Questions with Bloom’s Level and Course Outcomes (COs)

Part	A						B
Q. No.	1	2	3	4	5	6	7
Blooms Level	1	1	1	3	2	4	5
Mapped CO(s)	1	1	2	2	2	3	4