

St Aloysius (Deemed to be University)

Mangaluru

School of Information Science & Technology (PG Programme)

Semester II – P.G. Examination - M.Sc. Big Data Analytics

April -2026

MACHINE LEARNING - I

Time: 2½ Hours

Max Marks: 60

Note: i) Answer all the questions

ii) Draw diagrams wherever necessary

SECTION – A

Answer any **FIVE** of the following. (5x2=10)

1. Name the synonyms for axons and dendrons in ANN.
2. Define sigmoid function.
3. Define feed-forward neural network.
4. Why is SVM called a maximum margin classifier?
5. List the steps involved in the K-Means algorithm.
6. Define Machine Learning
7. Why data preprocessing is important.

SECTION - B

Answer any **SIX** of the following. (6x5=30)

8. Explain linear regression with the help of a suitable equation.
9. Explain how skewness affects machine learning models.
10. Briefly describe how do you decide the number of clusters in K-means clustering.
11. Explain accuracy, precision and recall as performance metrics.
12. Explain the concept of dimensionality reduction and its importance in machine learning.
13. Explain supervised and unsupervised learning with examples.
14. Derive logistic regression from logit function.
15. Explain the perceptron model with architecture and working.

SECTION-C

Answer any **TWO** of the following. (2x10=20)

16. Explain the gradient descent algorithm and its working steps.
17. Compare overfitting and underfitting, and explain techniques to limit overfitting.
18. Describe the feed-forward and backpropagation phases in neural network learning with suitable example.

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	1	2	1	1	2	2	2	2	2	2	2	4	2	2	2	1
Mapped CO(s)	3	3	3	4	5	1	1	1	4	5	4	5	1	2	3	2	2	3

St Aloysius (Deemed to be University)
Mangaluru

School of Information Science & Technology (PG Programme)
Semester II – P.G. Examination - M.Sc. Big Data Analytics
April -2026

ENABLING TECHNOLOGIES FOR DATA SCIENCE - I

Time: 2¹/₂ Hours

Max Marks: 60

Note: i) Answer all the questions
ii) Draw diagrams wherever necessary

SECTION – A

Answer any FIVE of the following. (5x2=10)

- 1. Define cluster analysis.
- 2. Define outlier in a dataset.
- 3. What is ROLAP?
- 4. What is HOLAP?
- 5. What is meant by missing data?
- 6. What is dimensionality reduction?
- 7. What is a hyperplane in SVM?

SECTION - B

Answer any SIX of the following. (6x5=30)

- 8. Describe support, confidence and lift in association rules.
- 9. Write short notes on high dimensionality and dynamic data as scalability issues.
- 10. Explain the KDD process with its major stages.
- 11. Explain the general principles of data warehousing.
- 12. Explain metadata and its types in data warehousing.
- 13. Explain Data Visualization and its best practices for effective Data Visualization.
- 14. Discuss the working of DBSCAN algorithm.
- 15. Explain the structure of a Decision Tree with diagram.

SECTION-C

Answer any TWO of the following. (10x2=20)

- 16. Explain Bayesian learning in detail. Derive Bayes’ theorem and discuss Naïve Bayes classifier.
- 17. Explain graph-theoretic clustering approach.
- 18. Explain the different types of data in detail with suitable examples.

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	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Mapped CO(s)	5	3	2	2	3	3	4	2	1	1	2	2	3	4	4	5	5	1

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

St Aloysius (Deemed to be University)**Mangaluru****School of Information Science & Technology (PG Programme)****Semester II – P.G. Examination - M.Sc. Big Data Analytics****April -2026****NATURAL LANGUAGE PROCESSING****Time: 2¹/₂ Hours****Max Marks: 60****Note: i) Answer all the questions****ii) Draw diagrams wherever necessary****SECTION – A****Answer any FIVE of the following.****(5x2=10)**

1. Give one example of semantic role assignment.
2. Define Recurrent Neural Network (RNN).
3. What is a grammar formalism?
4. Give two examples of named entities.
5. What is decoding in MT?
6. What is syntax in NLP?
7. identify the ambiguity in the sentence: "Time flies like an arrow."

SECTION - B**Answer any SIX of the following.****(6x5=30)**

8. Explain the role of language models in NLP applications.
9. Illustrate sequence labeling for relation extraction.
10. Discuss the advantages and limitations of shift-reduce parsing.
11. Apply a context-based method to resolve ambiguity in the sentence: "The crane is flying."
12. Explain how probability of a sentence is calculated in SMT.
13. Describe the limitations of early NLP systems.
14. Describe syntax-based MT approach.
15. Using the following grammar rules:

 $S \rightarrow NP VP$ $NP \rightarrow Det N$ $VP \rightarrow V NP$ $Det \rightarrow the \mid a$ $N \rightarrow teacher \mid lesson$ $V \rightarrow explains$

Derive the sentence below using Leftmost Derivation: **"The teacher explains the lesson."** Show all derivation steps clearly.

Contd...2

SECTION-C

Answer any TWO of the following. (2x10=20)

- 16. Evaluate different approaches used to handle ambiguity in NLP systems.
- 17. Explain the step-by-step process of Named Entity Recognition (NER). Illustrate each stage with a suitable example sentence and demonstrate how entities are identified and labeled.
- 18. Given a PCFG and a sentence with structural ambiguity, compute the probabilities of all possible parse trees and identify the most probable parse.

S→NP VP(1.0)
NP→Det N(0.6)
NP→NP PP(0.4)
VP→V NP(0.7)
VP→VP PP(0.3)
PP→P NP(1.0)
Det→the (0.6),a (0.4)
N→girl (0.3), boy (0.3), telescope (0.4)
V→saw (1.0)
P→with (1.0)

Sentence is : "The girl saw the boy with a telescope."

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	1	1	1	1	3	1	4	2	3	2	1	1	3	5	4	3
Mapped CO(s)	4	5	3	4	5	3	3	2	4	3	3	5	1	5	3	1	4	3

St Aloysius (Deemed to be University)
Mangaluru

School of Information Science & Technology (PG Programme)
Semester II – P.G. Examination - M.Sc. Big Data Analytics
April -2026

FOUNDATIONS OF DATA SCIENCE

Time: 2½ Hours

Max Marks: 60

Note: i) Answer all the questions
ii) Draw diagrams wherever necessary
SECTION – A

Answer any **FIVE** of the following. (5x2=10)

1. Give one application of MST.
2. What is random projection?
3. Where are random graphs used in real-world networks?
4. How is SVD used in recommender systems?
5. Give one real-world example of data streams.
6. Define degree of a vertex in a directed graph.
7. How is distance used in clustering algorithms?

SECTION - B

Answer any **SIX** of the following. (6x5=30)

8. Derive relation between random walks and coin tossing.
9. Analyze connectivity threshold in G(n,p).
10. Explain binomial degree distribution with an example.
11. Derive SVD for a simple 2x2 matrix.
12. Discuss dimensionality reduction using matrix techniques.
13. Prove that a complete graph with n vertices has n(n–1)/2 edges and discuss its real-world relevance.
14. Analyze variance and distribution behavior as dimension increases.
15. Discuss applications of random graphs in epidemiology and communication networks.

SECTION-C

Answer any **TWO** of the following. (2x10=20)

16. Explain BFS and DFS algorithms with pseudocode and complexity analysis.
17. Explain how high-dimensional data appears in text mining.
18. Derive power method and explain convergence conditions.

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	3	4	1	1	4	2	4	2	2	2	5	4	2	2	1	2
Mapped CO(s)	1	2	3	4	5	1	2	4	3	3	4	4	1	5	3	1	2	4

IT3FPOE589

Reg. No.:

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

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: 1¹/₂ 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

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

St Aloysius (Deemed to be University)

Mangaluru

School of Information Science & Technology (PG Programme)

Semester II – P.G. Examination - M.Sc. Big Data Analytics

April -2026

ADVANCED STATISTICAL METHODS

Time: 2½ Hours

Max Marks: 60

Note: i) Answer all the questions**ii) Draw diagrams wherever necessary****SECTION – A****Answer any FIVE of the following.****(5×2=10)**

1. Define an estimate. Give example
2. What is a sufficient estimator?
3. When is the Z-test used in hypothesis testing? Give example.
4. What do you mean by one sample test? Write the t-test statistic for one sample test.
5. Write the normal equations used to obtain the least squares estimates of β_0 and β_1 .
6. What is meant by the term BLUE in the Gauss–Markov theorem?
7. Define the adjusted coefficient of determination (Adjusted R^2).

SECTION - B**Answer any SIX of the following.****(6×5=30)**

8. Find the best values of a and b so that $y = a e^{bx}$ fits the data given in the table:

x	2	4	6	8	10
y	4.077	11.084	30.128	81.897	222.62

9. If $X_1, X_2, \dots, X_n$, for a random sample from a distribution for which the probability distribution function is as follows.
 $f(x, \theta) = (1+\theta)x^\theta$, where $0 < x < 1$ and $\theta > 0$, find MLE for the parameter θ
10. In a random selection of 64 road crossing in a town, the mean number of automobile accidents per year was found to be 4.2 and sample standard deviation was 0.8. Construct a 95% confidence interval for the mean number of automobile accidents per crossing per year.
11. Explain the following
 - a. Hypothesis testing
 - b. Null and alternate hypothesis

12. Following are the gain in weights (in grams) of fishes fed on two diets A and B. Assume normal populations.
- A: 25, 32, 30, 34, 24, 14, 32, 24, 30, 31, 35, 25
- B: 44, 34, 22, 10, 47, 31, 40, 30, 32, 35, 18, 21, 35, 29, 22
- Test if the variances of the two diets differ significantly as regards their effect on increase in weight. [$F_{0.05,}=2.21$]

13. The accompanying table presents the number of by Smoking and Disease status.

	Disease A	
Smoking	Yes	No
Yes	80	120
No	20	280

- a. What is the odd's ratio of getting a disease?
- b. What is the odd's ratio of not getting a disease?
14. Given the following data obtained from a completely randomized design with four treatments, analyse the data and draw conclusions about the equality of treatment effects. [$F_{3,10}=3.71$]

Treatments			
T ₁	T ₂	T ₃	T ₄
20.9	23.7	13.2	5.8
12.4	14.4	10.2	6.1
10.1	9.0	5.1	4.8
4.2			1.5

15. What is Multiple linear regression model? How is it different from simple linear regression model?

SECTION-C

Answer any **TWO** of the following. (10x2=20)

16. The following data corresponds to the actual and predicted value of the regression line Y on X. Test the following data for Autocorrelation using Durbin Watson's test statistic. Explain the procedure in detail.

Y	41	36	12	18	23	19	16	14
y [^]	35.4	37.4	17.6	12.0	24.4	15.9	22.5	21.4

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. Consider the following data which relates to the ICU patient’s vital status.
The coefficients of the model are as follows:

Intercept	BP	heartbeat
-3.84	0.03	-0.02
vital status	BP	heartbeat
1	80	96
0	142	88
0	112	80
0	100	70
1	128	90
0	142	103
0	110	154
0	110	132
0	104	66

Stating the model find the probabilities of prediction for the vital status (1: died, 0: lived) by using the coefficients of the logistic regression model.

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	2	1	1	3	3	3	2	3	3	3	2	3	3	3
Mapped CO(s)	1	2	2	2	3	4	2	3	1	2	2	2	5	3	4	3	4	5