

CHOICE BASED CREDIT SYSTEM

M.Sc. BIG DATA ANALYTICS FIRST SEMESTER DEGREE EXAMINATION

NOVEMBER 2025

Data Warehousing and Data Mining

Duration:3 Hours

Max Marks:70

PART A

I. Answer any FOUR of the following

(4×5= 20)

- a) Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8): (a) Compute the Euclidean distance between the two objects. (b) Compute the Manhattan distance between the two objects. (c) Compute the Minkowski distance between the two objects, using $q = 3$.
- b) Explain OLAP-style analysis that can be fused with data mining techniques.
- c) Explain Supervised methods that can be used to build outlier detection models.
- d) Explain Cyberwarfare and how single-point failures contribute to catastrophic failures in systems.
- e) Differentiate between star schema and Snowflake schema.

PART B

II. Answer any FIVE questions selecting at least one question from each unit:

(5×10= 50)

UNIT-I

2. Explain in detail the three areas of basic statistical descriptions.
3. Describe the steps involved in data mining when viewed as a knowledge discovery process with an illustration.

UNIT-II

4. Explain general optimization techniques for efficient computation of data cubes.
5. Explain in detail the process of Data Cleaning.

UNIT-III

6. Elaborate on mining closed frequent itemsets with suitable example.
7. Explain briefly the pattern mining method with respect to 1. Pattern Diversity 2. Different kinds of data and applications involved.

UNIT-IV

8. Explain in detail the various tasks involved in Product Selection and Installation.
9. Describe the Policy/Claims Consolidated Schema and explain the use of a factless fact table for an accident events.

CHOICE BASED CREDIT SYSTEM

M.Sc. BIG DATA ANALYTICS FIRST SEMESTER DEGREE EXAMINATION

NOVEMBER 2025

Python Programming for Data Analytics

Duration:3 Hours

Max Marks:70

PART A**I. Answer any FOUR of the following (4×5= 20)**

- a) Consider the csv file with the following data: a,b,c,d,Good 1,2,3,4,Hello 5,6,7,8, World 9,10,11,12,food.
(i) Write the command to read the csv file.
(ii) Write the command to make "a,b,c,d,Good" as column names.
(iii) Write a command to make index column as "Good".
- b) With an example illustrate merging on index.
- c) With example explain grouping by index levels.
- d) Describe the different ways of map projections with suitable code.
- e) Demonstrate how to create a Pandas series object as a generalized NumPy array.

PART B**II. Answer any FIVE questions selecting at least one question from each unit:****(5×10= 50)****UNIT-I**

2. a) Explain Lambda function with an example. (5)
b) With an example explain how to pass a function as a parameter to another function. (5)
3. a) Explain the following control statements: if, if else and else if with syntax and example (6)
b) Describe the term function. Write a function to calculate the square of a number. (4)

UNIT-II

4. (a) Explain Numpy Fancy Indexing with an example. (6)
(b) Explain any 2 NumPy Array creation functions with examples. (4)
5. (a) Explain the feature of reindexing in Pandas with a suitable example. (5)
(b) Explain on the Ufuncs- Index Alignment. (5)

6. Enumerate and explain various options of pivot tables.
7. (a) Explain the concept of shifting in time series with suitable examples. (5)
(b) Describe rolling windows concept. (5)

UNIT-IV

8. (a) Explain the concept of subplots in Matplotlib. (5)
(b) Describe the usage of matplotlib as a visualization tool. (5)
9. With an example explain the method of plotting multiple lines with different line styles.

CHOICE BASED CREDIT SYSTEM

M.Sc. BIG DATA ANALYTICS FIRST SEMESTER DEGREE EXAMINATION

NOVEMBER 2025

Quantitative Techniques for Data Science & Data Visualization

Duration:3 Hours

Max Marks:70

PART A

I. Answer any FOUR of the following

(4×5= 20)

a) Find the inverse of a matrix $A = \begin{bmatrix} 1 & 4 & -2 \\ -2 & -5 & 4 \\ 1 & -2 & 1 \end{bmatrix}$.

b) What is kurtosis? Explain various types of kurtosis with the help of graph.

c) 1) State multiplication theorem of probability. (2)

2) An equipment has 3 parts A,B,C with probability of failure of these parts during usage are 0.03,0.02 and 0.05 respectively. Find the probability that the equipment fails during usage. (3)

d) Briefly explain Critical region, Level of significance, Power of the test, Size of the test, p-value, and Most powerful test.

e) Explain five mapping components to aesthetic attributes.

PART B

II. Answer any FIVE questions selecting at least one question from each unit:

(5×10= 50)

UNIT-I

2. Find the rank of the following matrix (4)

a) $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

b) Solve the following set of equations using Cramers Rule (6)

$$2x + 5y + z + 1 = 0$$

$$x + 7y = 6z - 18$$

$$3y + 6z = 9$$

3. What is a set? Explain different types of sets with an example.

UNIT-II

4. Perform descriptive analysis for the following data related to the number of customers who prefers a specific product in different localities and represent the data using a suitable graphical presentation

54, 49, 48, 49, 50, 54, 52, 48, 47, 55, 48, 53, 54, 53, 54

5. 1) State Bayes Theorem

(2)

2) Three factories produce light bulbs to supply to the market. Factory A produces 20%, 50% of the tools are produced in factory B and 30% in factory C. 2% of the bulbs produced in factory A, 1% of the bulbs produced in factory B and 3% of the bulbs produced in factory C are defective. A bulb is selected at random in the market and found to be defective. What is the probability that this bulb was produced by factory B?

(8)

UNIT-III

6. During the first 13 weeks of television season, Saturday evening 8 PM to 9 PM audience proportion were recorded as ABC 29%, CBS 28%, NBC 25% and independents 18%. A sample of 300 homes two weeks after a Saturday night schedule revision yielded the following viewing audience data. ABC 95 homes, CBS 70 homes, NBC 89 homes, and independents 46 homes. Test at 5% level of significance that whether the viewing proportion audience changed ($TV=7.815$)

7. The average number of acres burned by forest and range fires in a large New Mexico county is 4,300 acres per year, with a standard deviation of 750 acres. The distribution of the number of acres burned is normal. What is the probability that between 2,500 and 4,200 acres will be burned in any given year? What number of burnt acres corresponds to the 38th percentile?

UNIT-IV

8. What are the different types of data visualization available? Explain the various types along with the tools used in visualization..
9. By drawing 500 random numbers from an Exponential distribution with the parameter 4 how do you draw a histogram and obtain a Normal curve for it using ggplot2?

24BDAS104

Reg No :

CHOICE BASED CREDIT SYSTEM

M.Sc. BIG DATA ANALYTICS FIRST SEMESTER DEGREE EXAMINATION

NOVEMBER 2025

Optimization Techniques

Duration:3 Hours

Max Marks:70

PART A

I. Answer any FOUR of the following

(4×5= 20)

- a) Describe the various objectives of OR in detail.
- b) Explain in detail the concept of Linear Programming Model.
- c) Solve the LPP using Simplex method
Max $Z = 3x+2y$
Subject to
 $x+y \leq 4$
 $x \leq 2$
 $y \leq 3$
 $x,y \geq 0$
- d) Explain Quadratic programming.
- e) Explain the elements of a queuing structure.

PART B

II. Answer any FIVE questions selecting at least one question from each unit:

(5×10= 50)

UNIT-I

- 2. Write in detail the scope of Operations Research.
- 3. How judgement phase helps in identifying right problem? Explain the process with the steps.

UNIT-II

- 4. Solve the following assignment problem using Hungarian Method.

	J1	J2	J3	J4	J5
M1	9	11	14	11	7
M2	6	15	13	13	10
M3	12	13	6	8	8
M4	11	9	10	12	9
M5	7	12	14	10	14

5. Determine the optimal feasible solution to the following Transportation problem using North West Corner Method and MODI Method.

	A	B	C	Supply
P	2	2	4	5
Q	3	3	1	8
R	5	4	7	7
S	1	6	2	14
Demand	7	9	18	

UNIT-III

6. Explain concave and convex functions in non linear programming with illustration.
7. What is separable programming? Explain in detail.

UNIT-IV

8. Explain the advantages and limitations of decision tree approach.
9. Describe the need and applications of Simulation Techniques. Explain their advantages and limitations.
