24BDAH101 Reg No : .....

CHOICE BASED CREDIT SYSTEM

M.Sc. FIRST SEMESTER DEGREE EXAMINATION JANUARY 2025

BIG DATA ANALYTICS

Data Warehousing and Data Mining

Duration:3 Hours Max Marks:70

PART A

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

a) Explain the role of statistical models that are widely used to model data and data classes.

b) Explain the general optimization techniques for efficient computation of data cubes.

c) Describe the concept of Outliers with suitable illustrations.

d) Explain Cyberwarfare and how single-point failures contribute to catastrophic failures in systems

e) Elaborate on Data Cube–Based Mining of Quantitative Associations.

PART B

II. Answer any FIVE questions selecting at least one question from each unit: (5×10= 50)

UNIT-I

2. Elaborate on the major issues encountered in Data Mining research.

3 Explain in detail the three areas of basic statistical descriptions.

UNIT-II

4. Explain the data warehousing multitiered architecture with an illustration

5. Explain in detail the process of Data Cleaning.

UNIT-III

6. Elaborate Naive Bayesian Classification with an illustration.

7. Explain Pruning Pattern Space with Pattern Pruning Constraints.

UNIT-IV

8 Explain Analytic Application Specification and the Development of Lifecycle Analytic Applications Track

9. Explain with examples the various dimension details and techniques used in a policy transaction system.

.....

24BDAH102 Reg No : .....

CHOICE BASED CREDIT SYSTEM

M.Sc FIRST SEMESTER DEGREE EXAMINATION JANUARY 2025

BIG DATA ANALYTICS

Python Programming for Data Analytics

Duration:3 Hours **Max Marks:70**

PART A

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

a) Explain any five features of Python.

b) Explain Multiple Indexed Series in Pandas.

c) Explain any two tools used in time series with suitable examples

d) Explain the concept of histogram and density plots with suitable illustrations.

e) Explain briefly any three regular expression methods with suitable python code.

PART B

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

(5×10= 50)

UNIT-I

2. a) Explain the following control statements: If, if else and else if with syntax and examples. (6)

b) Describe the term function. Write a function to check if a number is even or odd. (4)

3. (a) Write a Python program to perform the following:

(i) read a text file called file1.txt with the following numbers:

10,21,32,43,54,65,76,65,30,69,78

(ii) compute the mean

(iii) compute the median

(iv) compute the mode of these numbers

(v) write the output into a file called file2.txt

(b) Write Python program to count the number of words in a given text string. (5+5)

UNIT-II

4. (a) Explain Pandas index object with suitable examples (5)

(b) Demonstrate how to create a Pandas series object as a generalized NumPy array. (5)

5. a) Discuss the ufunc equivalent to arithmetic operators in Numpy. (8)

b) Explain ravel function with an example. (2)

UNIT-III

6. (a) Explain multilevel pivot tables with suitable python code. (5)
   (b) Describe the concept of cross tabulations. Explain with an example. (5)

7. (a) Explain various optimized groupby methods with suitable examples. (5)
   (b) Discuss data aggregation with suitable example. (5)

UNIT-IV

8. (a) Explain any five map-specific methods used to plot data on maps. (5)
   (b) Describe Three-dimensional plotting in Matplotlib. (5)

9. (a) Describe any five commands used to create text in matplotlib plots. (5)
   (b) Describe visualization with Seaborn in detail. (5)

****

24BDAH103 Reg No : .....

CHOICE BASED CREDIT SYSTEM

M.Sc. FIRST SEMESTER DEGREE EXAMINATION JANUARY 2025

BIG DATA ANALYTICS

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) Solve the following set of equations by Cramer's method

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

$x + 7y = 6z - 18$

$3y + 6z = 9$

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) Write a note on test for normality and explain it using box-plot.

e) How do we draw Box plot and Histogram for non categorical data using ggplot2 ?

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

a) $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ b) $A = \begin{bmatrix} 2 & -4 & 1 \\ 2 & 6 & 0 \\ 1 & 5 & 0 \end{bmatrix}$

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

UNIT-II

4. 1) Write a note on Structured and Unstructured data. (4)

   2) Explain NOIR classification with examples. (6)

5. 1) State Bayes Theorem (2)

   2) Three Companies produces electronic components to supply to the market. Company A produces 20%, 50% of the components are produced in Company B and 30% in Company C. 2% of the components produced in Company A, 1% of the components produced in Company B and 3% of the components produced in Company C are defective. A component is selected at random in the market and found to be defective. What is the probability that this component was produced by Company B? (8)

UNIT-III

6. Kamal milk products company , produces ghee in sachets of 500 ml each. Ghee is filled in sachets by a machine for which the standard deviation of filing is 5 ml. 3 different possible situations are:

   (i) It is required to verify whether the machine is filling 500 ml ghee on an average,then the machine is set properly.

   (ii) There is a complaint from customer that the sachets have less than 500 ml ghee.

   (iii) The management requires that, on an average the fillings should not exceed 500 ml Suppose one of the above situation arises and we are required to reach a conclusion. Among the filled sachets 72 sachets are randomly picked and its contents are measured, Mean of these measurements are found to be 501.4 ml. What is your conclusion?

7. 1) Write down the p.m.f , mean and variance of Binomial distribution. (3)

   2) The probability that a bomb dropped on a bridge hits it is 0.4. Eight bombs are dropped on the bridge. Two bomb hits are enough to destroy the bridge. Find the probability that

   (i) all the bombs hit the bridge.

   (ii) two bombs hit the bridge

   (iii) the bridge is destroyed. (7)

UNIT-IV

8. What are the principles to be followed in data visualization? Explain.

9. Briefly explain different scale functions.

****

24BDAS104 Reg No : .....

CHOICE BASED CREDIT SYSTEM

M.Sc. FIRST SEMESTER DEGREE EXAMINATION JANUARY 2025

BIG DATA ANALYTICS

Optimization Techniques

Duration:3 Hours **Max Marks:70**

PART A

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

a) Explain any five types of mathematical models in OR

b) Mr. Wagner retired from Government service and received his retirement benefits amounting to Rs.45 lakhs. He plans to invest this amount in various investment channels, so as to maximize return on investment. The investment alternatives, period of investment and estimated risk involved in each alternative (on a 5-point scale) are given in the table. His requirements are:

1. Average risk should not be more than 4.
2. Funds should not be locked up for more than 6 years.
3. He should necessarily invest at least 30% of the fund in Real Estate.

Formulate Mr. Wagner's problem as LP model.

Investment Choice	Return (%)	Period (in Years)	Risk (1 lowest, 5 highest)
Mutual Funds	20	5	1
Company deposits	9.5	2	3
Equity shares	50	4	5
Gold	70	6	1
Real Estate	40	5	2

c) Explain local and global optimum in non linear programming.

d) Explain the steps involved in carrying out Monte Carlo simulation.

e) Briefly explain the mathematical formulation of Simplex Table.

PART B

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

(5×10= 50)

UNIT-I

2. Explain the Scientific Method and its phases in OR.

3. Describe in detail the scope of Operations Research.

UNIT-II

4. Obtain the dual problem of the following primal of LP problem

Minimize $Z = x + 2y$

Subjected to constraints

$$2x + 4y < 160$$
$$x + y = 30$$
$$x > 10$$

and $x, y > 0$

5. Solve the below LPP model under Graphical solution:

Max $Z = 50x_1 + 80x_2$

Subject to: $x_1 \leq 80$

$$x_2 \leq 60$$
$$5x_1 + 6x_2 \leq 600$$
$$x_1 + 2x_2 \leq 160$$
$$x_1, x_2 \geq 0$$

UNIT-III

6. Explain in detail the classification of non-linear programming.

7. Explain in detail the concept of Quadratic Programming.

UNIT-IV

8. Explain the advantages and limitations of decision tree approach.

9. Explain the concepts of Waiting time and Idle time costs with illustration.

****