



Roll No:

BTECH
(SEM. VI) THEORY EXAMINATION 2025-26
DATA ANALYTICS

TIME: 3 HRS

M.MARKS: 70

Note: Attempt all Sections. In case of any missing data; choose suitably.

SECTION A

1. Attempt all questions in brief.

02 x 7 = 14

Qno.	Question	CO	Level
a.	Explain structured vs semi-structured data.	1	K2
b.	List the phases of the Data Analytics Lifecycle.	1	K1
c.	Define multivariate analysis.	2	K1
d.	Define Bayesian inference.	2	K1
e.	What is stream data? Give one example.	3	K1
f.	What is the Apriori algorithm?	4	K2
g.	What are NoSQL databases?	5	K2

SECTION B

2. Attempt any three of the following:

07 x 3 = 21

Qno.	Question	CO	Level
a.	Explain the Data Analytics Lifecycle in detail with a neat diagram.	1	K2
b.	Explain the linear regression model along with its assumptions. Using the least squares method, fit a regression line for the following data and obtain the equation of the line ($y = mx + c$). Data: (1,2), (2,4), (3,5), (4,4), (5,5)	2	K3
c.	Discuss the concept of filtering data stream processing. Explain Bloom filtering in detail.	3	K2
d.	Provide an in-depth comparison between the CLIQUE and ProCLUS clustering algorithms. How do these methods handle challenges such as noise, outliers, and varying cluster shapes?	4	K3
e.	Discuss the components of Hadoop ecosystem: Apache Hive, Apache Pig, Apache HBase.	5	K2

SECTION C

3. Attempt any one part of the following:

07 x 1 = 07

Qno.	Question	CO	Level
a.	What happens in the 'model planning' and 'model building' stages? How are these two different?	1	K2
b.	Discuss the evolution of Big Data and need for data analytics in modern organizations.	1	K2

4. Attempt any one part of the following:

07 x 1 = 07

Qno.	Question	CO	Level
a.	Explain neural networks including learning, generalization, and competitive learning.	2	K2
b.	Discuss Fuzzy membership function. Apply Union, Intersection, Complement, Bold Union and Bold Intersection operations on fuzzy sets A and B.	2	K3



Paper ID : 261318

Printed Page: 2

Subject Code:BIT601

Roll No:

BTECH
(SEM. VI) THEORY EXAMINATION 2025-26
DATA ANALYTICS

TIME: 3 HRS

M.MARKS: 70

Given: $A = \{(10,0.2), (20,0.4), (25,0.7), (30,0.9), (40,1.0), (50,0.4)\}$ $B = \{(10,0.4), (20,0.1), (25,0.9), (30,0.2), (40,0.6), (50,0.6)\}$		
---	--	--

5. Attempt any one part of the following:**07 x 1 = 07**

Qno.	Question	CO	Level
a.	Discuss Real-Time Analytics Platforms (RTAP) and explain applications like sentiment analysis and stock prediction.	3	K2
b.	Apply the Flajolet–Martin algorithm on the following data stream to estimate the number of unique elements: $S = \{1,3,2,1,2,3,4,3,1,2,3,1\}$ Given the hash function: $h(x) = (3x + 1) \bmod 8$ Compute the hash values for each element i. Represent the hash values in 3-bit binary form ii. Determine the maximum number of trailing zeros (R) iii. Estimate the number of distinct elements using 2^R	3	K3

6. Attempt any one part of the following:**07 x 1 = 07**

Qno.	Question	CO	Level
a.	Discuss market basket analysis and association rule mining.	4	K2
b.	Cluster the following eight points (with (x, y) representing locations) into three clusters: $A_1(2, 10), A_2(2, 5), A_3(8, 4), A_4(5, 8), A_5(7, 5), A_6(6, 4), A_7(1, 2), A_8(4, 9)$ - Initial cluster centers are: $A_1(2, 10), A_4(5, 8)$ and $A_7(1, 2)$. - The distance function between two points $a = (x_1, y_1)$ and $b = (x_2, y_2)$ is defined as $d(a, b) = x_2 - x_1 + y_2 - y_1$ Use K-Means Algorithm to find the three cluster centers after implanting all eight points.	4	K3

7. Attempt any one part of the following:**07 x 1 = 07**

Qno.	Question	CO	Level
a.	Explain MapReduce model with architecture and working in Apache Hadoop.	5	K2
b.	Explain Exploratory Data Analysis (EDA) in R with steps and sample commands.	5	K2