PAPER ID-310908

Roll No: | | | | | | | | | | | | | |

BTECH
(SEM V) THEORY EXAMINATION 2024-25
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. **2 x 07 = 14**

Q no.	Question	CO	Level
a.	Differentiate between Predictive and Prescriptive Data Analytics.	1	K2
b.	Define the term data lake, data base and data warehouse.	1	K1
c.	Explain the concept of Outliers.	2	K2
d.	Describe the concept of Lasso Regression.	2	K2
e.	Differentiate between Steam Processing and Traditional Data Processing.	3	K2
f.	Write the two limitations of K-Mean.	4	K1
g.	Discuss the various categories of clustering techniques.	5	K2

SECTION B

2. Attempt any *three* of the following: **07 x 3 = 21**

a.	Explain the different categories of data analytics with examples.	1	K2
b.	Explore PCA. Given data = {4, 8, 13, 7; 11, 4, 5, 14}. Compute the principal component using PCA algorithm. Also use PCA to reduce dimension from 2 to 1.	2	K3
c.	Explore the term- Market Basket Analysis. Is it supervised or unsupervised? Determine how would a company use market basket analysis to improve its marketing strategies?	3	K3
d.	Differentiate between CLIQUE and ProCLUS clustering	4	K4
e.	Differentiate between NoSQL database and a Relational database. Identify when one should use a NoSQL database instead of a relational database with a suitable example.	5	K4

SECTION C

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

a.	Differentiate between Structured data, Semi-structured data and Unstructured Data.	1	K2
b.	Describe Big Data and its characteristics.	1	K2

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

a.	Differentiate between Neural Network and Artificial Neural Network.	2	K2
b.	A= {(10,0.2), (20,0.4), (25,0.7), (30,0.9), (40,1), (50,0.4} B= {(10,0.4), (20,0.1), (25,0.9), (30,0.2), (40,0.6), (50, 0.6)} Apply Union, Intersection, Complement, Bold Union and Bold Intersection operations on above listed Fuzzy Sets.	2	K3

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

a.	Explain and apply Flajolet-Martin algorithm on the following stream of data to identify unique elements in the stream. S=1,3,2,1,2,3,4,3,1,2,3,1 Given: h(x)=(6x+1) mod 5	3	K3

PAPER ID-310908

Roll No: | | | | | | | | | | | | | |

BTECH
(SEM V) THEORY EXAMINATION 2024-25
DATA ANALYTICS

TIME: 3 HRS **M.MARKS: 70**

b.	Discuss the Concept of filtering in Data Stream Processing. Explain Bloom Filtering in detail.	3	K2

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

a.	Cluster the following eight points (with (x, y) representing locations) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9). Initial cluster centers are A1(2, 10), A4(5, 8) and A7(1, 2). The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as P(a, b)= |x2 – x1| + |y2 – y1| Use K-Means Algorithm to find the three cluster centers after implanting all eight points.	4	K3
b.	The database has 6 transactions. Assume Support threshold=50%, Confidence= 60% **TID** / **Items Bought** 10 / Beer, Nuts, Diaper 20 / Beer, Coffee, Diaper 30 / Beer, Diaper, Eggs 40 / Nuts, Eggs, Milk 50 / Nuts, Coffee, Diaper, Eggs, Milk 60 / Beer, Nuts, Diaper i) Use Apriori algorithm to find all frequent itemsets. ii) Show all the strong association rules (with support and confidence)	4	K3

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

a.	Brief about the main components of MapReduce.	5	K2
b.	Draw the architecture of HIVE with its features.	5	K2