



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. भविष्य कहनेवाला और आदेशात्मक डेटा विश्लेषण के बीच अंतर करें।	1	K2
b.	Define the term data lake, data base and data warehouse. डेटा लेक, डेटा बेस और डेटा वेयरहाउस शब्द को परिभाषित करें।	1	K1
c.	Explain the concept of Outliers. आउटलायर्स की अवधारणा को समझाइए।	2	K2
d.	Describe the concept of Lasso Regression. लासो प्रतिगमन की अवधारणा का वर्णन कीजिए।	2	K2
e.	Differentiate between Stream Processing and Traditional Data Processing. स्ट्रीम प्रोसेसिंग और पारंपरिक डेटा प्रोसेसिंग के बीच अंतर करें।	3	K2
f.	Write the two limitations of K-Mean. K-माध्य की दो सीमाएँ लिखिए।	4	K1
g.	Discuss the various categories of clustering techniques. क्लस्टरिंग तकनीकों की विभिन्न श्रेणियों पर चर्चा करें।	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. डेटा एनालिटिक्स की विभिन्न श्रेणियों को उदाहरणों के साथ समझाए।	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. PCA एक्सप्लोर करें. दी गयी जानकारी = {4, 8, 13, 7; 11, 4, 5, 14} पीसीए एल्गोरिथम का उपयोग करके प्रमुख घटक की गणना करें। आयाम को 2 से 1 तक कम करने के लिए पीसीए का भी उपयोग करें।	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? शब्द का अन्वेषण करें- मार्केट बास्केट विश्लेषण। क्या यह पर्यवेक्षित या असुरक्षित है? निर्धारित करें कि कोई कंपनी अपनी मार्केटिंग रणनीतियों को बेहतर बनाने के लिए मार्केट बास्केट विश्लेषण का उपयोग कैसे करेगी?	3	K3
d.	Differentiate between CLIQUE and ProCLUS clustering. CLIQUE और ProCLUS क्लस्टरिंग के बीच अंतर.	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. NoSQL डेटाबेस और एक रिलेशनल डेटाबेस के बीच अंतर करें। पहचानें कि किसी को एक उपयुक्त उदाहरण के साथ एक रिलेशनल डेटाबेस के बजाय नोएसक्यूएल डेटाबेस का उपयोग कब करना चाहिए।	5	K4



Roll No:

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

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

TIME: 3 HRS

M.MARKS: 70

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. संरचित डेटा, अर्ध-संरचित डेटा और असंरचित डेटा के बीच अंतर करें।	1	K2
b.	Describe Big Data and its characteristics. बिग डेटा और इसकी विशेषताओं का वर्णन करें।	1	K2
4. Attempt any one part of the following:		07 x 1 = 07	
a.	Differentiate between Neural Network and Artificial Neural Network. तंत्रिका नेटवर्क और कृत्रिम तंत्रिका नेटवर्क के बीच अंतर स्पष्ट कीजिए।	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. $E = \{(10,0.2), (20,0.4), (25,0.7), (30,0.9), (40,1), (50,0.4)\}$ $B_i = \{(10,0.4), (20,0.1), (25,0.9), (30,0.2), (40,0.6), (50, 0.6)\}$ उपरोक्त सूचीबद्ध फजी सेट पर यूनियन, इंटरसेक्शन, कॉम्प्लीमेंट, बोल्ड यूनियन और बोल्ड इंटरसेक्शन ऑपरेशन लागू करें।	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: $h(x) = (6x+1) \bmod 5$ स्ट्रीम में अद्वितीय तत्वों की पहचान करने के लिए डेटा की निम्नलिखित धारा पर फ्लैजोलेट-मार्टिन एल्गोरिथ्म को समझाएं और लागू करें। $एस = 1,3,2,1,2,3,4,3,1,2,3,1$ दिया गया है: $h(x) = (6x+1) \bmod 5$	3	K3
b.	Discuss the Concept of filtering in Data Stream Processing. Explain Bloom Filtering in detail. डेटा स्ट्रीम प्रोसेसिंग में फ़िल्टरिंग की अवधारणा पर चर्चा करें। ब्लूम फ़िल्टरिंग को विस्तार से समझाएं।	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: $A_1(2, 10)$, $A_2(2, 5)$, $A_3(8, 4)$, $A_4(5, 8)$, $A_5(7, 5)$, $A_6(6, 4)$, $A_7(1, 2)$, $A_8(4, 9)$. 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 $P(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



Roll No:

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

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

TIME: 3 HRS

M.MARKS: 70

	<p>निम्नलिखित आठ बिंदुओं (स्थानों का प्रतिनिधित्व करने वाले (x, y) के साथ) को तीन समूहों में क्लस्टर करें: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)। प्रारंभिक क्लस्टर केंद्र A1(2, 10), A4(5, 8) और A7(1, 2) हैं। दो बिंदुओं $a = (x_1, y_1)$ और $b = (x_2, y_2)$ के बीच दूरी फ़ंक्शन को किस रूप में परिभाषित किया गया है?</p> $\rho(a, b) = x_2 - x_1 + y_2 - y_1 $ <p>सभी आठ बिंदुओं को प्रत्यारोपित करने के बाद तीन क्लस्टर केंद्रों को खोजने के लिए के-मीन्स एल्गोरिथम का उपयोग करें।</p>																														
b.	<p>The database has 6 transactions. Assume Support threshold=50%, Confidence= 60%</p> <table border="1"> <thead> <tr> <th>TID</th> <th>Items Bought</th> </tr> </thead> <tbody> <tr> <td>10</td> <td>Beer, Nuts, Diaper</td> </tr> <tr> <td>20</td> <td>Beer, Coffee, Diaper</td> </tr> <tr> <td>30</td> <td>Beer, Diaper, Eggs</td> </tr> <tr> <td>40</td> <td>Nuts, Eggs, Milk</td> </tr> <tr> <td>50</td> <td>Nuts, Coffee, Diaper, Eggs, Milk</td> </tr> <tr> <td>60</td> <td>Beer, Nuts, Diaper</td> </tr> </tbody> </table> <p>i) Use Apriori algorithm to find all frequent itemsets. ii) Show all the strong association rules (with support and confidence)</p> <p>डेटाबेस में 6 लेनदेन हैं। समर्थन सीमा मान लें = 50%, आत्मविश्वास = 60%</p> <table border="1"> <thead> <tr> <th>TID</th> <th>आइटम खरीद</th> </tr> </thead> <tbody> <tr> <td>10</td> <td>बीयर, नट, डायपर</td> </tr> <tr> <td>20</td> <td>बीयर, काफी, डायपर</td> </tr> <tr> <td>30</td> <td>बीयर, डायपर, अंडे</td> </tr> <tr> <td>40</td> <td>नट्स, अंडे, दूध</td> </tr> <tr> <td>50</td> <td>नट्स, काफी, डायपर, अंडे, दूध</td> </tr> <tr> <td>60</td> <td>बीयर, नट, डायपर</td> </tr> </tbody> </table> <p>i) सभी लगातार आइटम खोजने के लिए Apriori एल्गोरिथम का उपयोग करें। ii) सभी मजबूत एसोसिएशन नियम दिखाएं (समर्थन और आत्मविश्वास के साथ)</p>	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	TID	आइटम खरीद	10	बीयर, नट, डायपर	20	बीयर, काफी, डायपर	30	बीयर, डायपर, अंडे	40	नट्स, अंडे, दूध	50	नट्स, काफी, डायपर, अंडे, दूध	60	बीयर, नट, डायपर	4	K3
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																														
TID	आइटम खरीद																														
10	बीयर, नट, डायपर																														
20	बीयर, काफी, डायपर																														
30	बीयर, डायपर, अंडे																														
40	नट्स, अंडे, दूध																														
50	नट्स, काफी, डायपर, अंडे, दूध																														
60	बीयर, नट, डायपर																														

7. Attempt any one part of the following:

07 x 1 = 07

a.	Brief about the main components of MapReduce. MapReduce के मुख्य घटकों के बारे में संक्षिप्त करें।	5	K2
b.	Draw the architecture of HIVE with its features. इसकी विशेषताओं के साथ HIVE की वास्तुकला को ड्रा करें।	5	K2