



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

BTECH
(SEM V) 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

Q no.	Question	CO	Level
a.	What is Data Analytics? डेटा एनालिटिक्स क्या है?	CO1	K1
b.	List any two characteristics of Big Data. बिग डेटा की कोई दो विशेषताएँ लिखिए।	CO1	K1
c.	What is regression analysis? प्रतिगमन विश्लेषण क्या है?	CO2	K2
d.	What is a data stream? डेटा स्ट्रीम क्या है?	CO3	K1
e.	Define frequent itemset. आवृत्त आइटमसेट को परिभाषित करें।	CO4	K1
f.	What is Hadoop? हडूप क्या है?	CO4	K1
g.	What is Exploratory Data Analysis (EDA)? अन्वेषणात्मक डेटा विश्लेषण (EDA) क्या है?	CO5	K2

SECTION B

2. Attempt any three of the following:

07 x 3 = 21

a.	Explain the phases of the Data Analytics Lifecycle with objectives of each phase. डेटा जीवन-चक्र के चरणों को उनके उद्देश्यों सहित समझाइए।	CO1	K2
b.	The following data represents study hours (X) and scores (Y): निम्नलिखित आँकड़े अध्ययन के घंटे (X) तथा प्राप्तांक (Y) को दर्शाते हैं। X: 2, 4, 6, 8, 10 Y: 50, 55, 65, 70, 78 (i) Find the linear regression equation. दिए गए डेटा के लिए रेखिक रिग्रेशन समीकरण ज्ञात करें (ii) Predict Y when X = 7. X = 7 पर Y का अनुमान लगाइए।	CO2	K3
c.	Explain Bayesian modeling and the role of prior and posterior probabilities. बेयेज़ियन मॉडलिंग तथा पूर्व व उत्तर प्रायिकताओं की भूमिका समझाइए।	CO2	K3
d.	A data stream generates 120,000 records per minute. If only 35% data is retained using a decaying window, calculate the number of records retained per minute and justify the use of windowing. यदि कोई डेटा धारा (Stream) प्रति मिनट 120,000 रिकॉर्ड उत्पन्न करती है तथा डिकेइंग विंडो का उपयोग कर केवल 35% डेटा संरक्षित किया जाता है, तो प्रति मिनट संरक्षित रिकॉर्ड की संख्या ज्ञात कीजिए तथा विंडोइंग तकनीक के उपयोग का औचित्य सिद्ध कीजिए।	CO3	K4



Roll No:

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

BTECH
(SEM V) THEORY EXAMINATION 2025-26
DATA ANALYTICS

TIME: 3 HRS

M.MARKS: 70

e.	Explain the importance of frequent itemset mining in market basket analysis. मार्केट बास्केट विश्लेषण में आवृत्त आइटमसेट माइनिंग का महत्व समझाइए।	CO4	K2
----	--	-----	----

SECTION C

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

a.	Explain Support Vector Machines (SVM) with the concept of margin and support vectors. मार्जिन एवं सपोर्ट वेक्टर के संदर्भ में सपोर्ट वेक्टर मशीन की व्याख्या करें।	CO2	K3
b.	A dataset has eigenvalues: 4.5, 2.8, 1.2, 0.4, 0.1. (i) Calculate percentage variance of each component. (ii) Determine the number of components required to retain 90% variance. किसी डेटा सेट के ईजेनवैल्यू क्रमशः 4.5, 2.8, 1.2, 0.4 तथा 0.1 हैं। (i) दिए गए ईजेनवैल्यू के आधार पर वैरिएंस प्रतिशत निकालिए। (ii) 90% वैरिएंस हेतु कंपोनेंट की संख्या बताइए।	CO2	K4

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

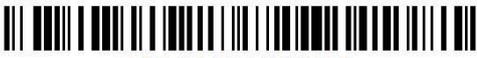
a.	A sliding window of size 200 slides by 50 elements. Calculate the number of windows formed after processing 1,000 elements and explain its impact on latency. यदि 200 आकार की स्लाइडिंग विंडो प्रत्येक 50 तत्वों पर आगे बढ़ती है, तो 1,000 तत्वों के संसाधन के बाद बनने वाली विंडो की संख्या ज्ञात कीजिए तथा विलंबता पर इसके प्रभाव की व्याख्या कीजिए।	CO3	K4
b.	Design a Real-Time Analytics Platform for social media sentiment analysis showing major components. सोशल मीडिया भाव-विश्लेषण हेतु प्रमुख घटकों सहित एक रियल-टाइम एनालिटिक्स प्लेटफॉर्म की संरचना (डिज़ाइन) प्रस्तुत कीजिए।	CO3	K5

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

a.	Given transactions: T1 {A, B, C}, T2 {A, B}, T3 {A, C}, T4 {B, C}, T5 {A, B, C} If minimum support = 60%, find all frequent item sets and show pruning. निम्नलिखित लेन-देन दिए गए हैं: T1 {A, B, C}, T2 {A, B}, T3 {A, C}, T4 {B, C}, T5 {A, B, C}। यदि न्यूनतम समर्थन 60% हो, तो सभी आवृत्त आइटमसेट ज्ञात कीजिए तथा प्रूनिंग प्रक्रिया को स्पष्ट कीजिए।	CO4	K4
b.	Compare hierarchical clustering and K-means clustering with suitable applications. हाइरार्किकल एवं के-मीन्स क्लस्टरिंग की अनुप्रयोग सहित तुलना करें।	CO4	K3

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

a.	Explain the working of MapReduce with intermediate key-value pairs. इंटरमीडिएट की-वैल्यू युग्म के संदर्भ में मैपरेड्यूस की कार्यप्रणाली की व्याख्या कीजिए।	CO4	K3
b.	Design a Hadoop-based framework for storing and processing large-scale data. बड़े पैमाने के डेटा के भंडारण एवं प्रसंस्करण हेतु हडूप आधारित फ्रेमवर्क की संरचना (डिज़ाइन) प्रस्तुत कीजिए।	CO4	K5



PAPER ID-311558

Printed Page: 3 of 3

Subject Code: BCS052

Roll No:

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

BTECH
(SEM V) THEORY EXAMINATION 2025-26
DATA ANALYTICS

TIME: 3 HRS**M.MARKS: 70****7. Attempt any one part of the following:****07 x 1 = 07**

a.	For the dataset {12, 15, 20, 22, 25, 30}, calculate mean and standard deviation and comment on data spread. दिए गए डेटा समुच्चय {12, 15, 20, 22, 25, 30} का माध्य एवं मानक विचलन ज्ञात कीजिए तथा डेटा के प्रसार पर टिप्पणी कीजिए।	CO5	K3
b.	Design an R-based workflow for analyzing unstructured text data. असंरचित पाठ्य डेटा के विश्लेषण हेतु R आधारित वर्कफ़्लो की संरचना (डिज़ाइन) प्रस्तुत कीजिए।	CO5	K6

QP26DP1_290

/ 12-Jan-2026 8:56:25 AM | 117.55.242.134