



Roll No:

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

BTECH
(SEM VI) THEORY EXAMINATION 2023-24
DATA ANALYTICS

TIME: 3 HRS**M.MARKS: 100**

Note: 1. Attempt all Sections. If require any missing data; then choose suitably.

SECTION A**1. Attempt all questions in brief.****2 x 10 = 20**

Qno.	Question	Marks	CO
a.	List the steps involved in the analytic process.	2	1
b.	Differentiate data analysis from data reporting.	2	1
c.	Explain the concept of Bayesian inference.	2	2
d.	Define the concept of the 'kernel trick'.	2	2
e.	List the uses of sliding windows in stream processing.	2	3
f.	Explain the concept of a decaying window.	2	3
g.	Mention the importance of frequent itemsets in market basket analysis.	2	4
h.	Explain the significance of support and confidence in the context of frequent itemsets.	2	4
i.	Define Hadoop Distributed File System.	2	5
j.	List some visual data analysis techniques.	2	5

SECTION B**2. Attempt any three of the following:****10 x 3 = 30**

a.	Discuss the key characteristics of data that impact data analytics.	10	1
b.	Differentiate linear and logistic regression models, and in what scenarios would each be most appropriate.	10	2
c.	Discuss differences between traditional batch data processing and stream data processing, and the main challenges associated with stream processing.	10	3
d.	Mention the limitations of the Apriori algorithm, and how they are addressed through optimization techniques.	10	4
e.	Explain the MapReduce programming model and its role in processing large datasets.	10	5

SECTION C**3. Attempt any one part of the following:****10 x 1 = 10**

a.	Discuss the various phases of the data analytics lifecycle and activities involved in each phase.	10	1
b.	Explain modern data analytics tools in detail.	10	1

4. Attempt any one part of the following:**10 x 1 = 10**

a.	Describe the process of conducting a principal component analysis (PCA) and explain how it can be used to reduce the dimensionality of a dataset.	10	2
b.	Explain how support vector machines (SVMs) utilize kernel functions to handle non-linearly separable data.	10	2



Roll No:

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

BTECH
(SEM VI) THEORY EXAMINATION 2023-24
DATA ANALYTICS

TIME: 3 HRS

M.MARKS: 100

5. Attempt any one part of the following:**10 x 1 = 10**

a.	Discuss a case study that demonstrates the use of stream processing in stock market prediction, including the techniques used and the results achieved.	10	3
b.	Explain Real-time Analytics Platforms (RTAP), and how do they enable real-time decision-making and analytics.	10	3

6. Attempt any one part of the following:**10 x 1 = 10**

a.	Describe the K-means clustering algorithm and the process of initializing centroids.	10	4
b.	Explain how ProCLUS (Projected Clustering) algorithm differs from CLIQUE in its approach to high-dimensional clustering.	10	4

7. Attempt any one part of the following:**10 x 1 = 10**

a.	Compare and contrast NoSQL databases with traditional SQL databases. List four main types of NoSQL databases.	10	5
b.	Describe Apache Hive and its use in querying and managing large datasets stored in Hadoop.	10	5