



Paper id: 250639

Roll No:

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

**BTECH**  
**(SEM VI) 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.**

**02 x 7 = 14**

Q no.	Question
a.	What are the main sources of data in data analytics?
b.	What is the purpose of the 'operationalization' phase?
c.	What is the purpose of support vector machines (SVM) in classification?
d.	What are fuzzy decision trees?
e.	Define stream computing and mention one of its key features.
f.	What are the advantages of K-means clustering?
g.	What is the purpose of Map Reduce in big data processing?

**SECTION B**

**2. Attempt any three of the following:**

**07 x 3 = 21**

a.	What are the different stages (phases) in a data analytics project? Explain each stage in simple words.
b.	Compare various types of support vector and kernel methods of data analysis.
c.	Discuss how mining data streams are applied in stock market predictions. What are the challenges and benefits of using stream data here?
d.	Describe the Apriori algorithm for mining frequent item sets. How does it work and what are its main steps?
e.	What is Pig and Hive? Compare their features and uses in data analysis.

**SECTION C**

**3. Attempt any one part of the following:**

**07 x 1 = 07**

a.	What is the difference between data analysis and data reporting? Give examples to show how each is used.
b.	What happens in the 'model planning' and 'model building' stages? How are these two different?

**4. Attempt any one part of the following:**

**07 x 1 = 07**

a.	What are neural networks? Explain how learning and generalization occur in neural networks. Include a simple diagram if possible.
b.	Given data = {1, 2, 3, 4; 4, 5, 6, 7}. Compute the principal components using the PCA algorithm.

**5. Attempt any one part of the following:**

**07 x 1 = 07**

a.	What is a Real-Time Analytics Platform (RTAP)? Discuss some applications of RTAP in industries today.
b.	Explain the concept of sampling in data streams. Why is it necessary, and how does it help in data analysis?

**6. Attempt any one part of the following:**

**07 x 1 = 07**

a.	Why is parallelism important in clustering large datasets? Describe some techniques used to achieve parallel clustering.
b.	What is Pro CLUS? How is it different from CLIQUE in handling clustering?



Paper id: 250639

Printed Page: 2 of 2  
Subject Code: BIT601

Roll No:

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

**BTECH**  
**(SEM VI) THEORY EXAMINATION 2024-25**  
**DATA ANALYTICS**

**TIME: 3 HRS**

**M.MARKS: 70**

7. Attempt any *one* part of the following:

07 x 1 = 07

a.	What is sharding in NoSQL databases? Why is it important for handling large volumes of data?	5	K2
b.	What is the Hadoop Distributed File System (HDFS)? How does it handle data storage and fault tolerance?	5	K3

QP25EP1\_143  
| 02-Jun-2025 1:29:58 PM | 122.185.51.242