PAPER ID-310861

Subject Code: KCS051

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

BTECH
(SEM V) 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**

Q no.	Question	Marks
a.	How have advancements in technology contributed to the scalability of analytics?	2
b.	What are the sources of data in the context of data analytics?	2
c.	Elaborate on the mathematical foundations of support vector machines.	2
d.	Discuss the advantages of using Bayesian methods in real-world applications.	2
e.	Elaborate on the methods used for filtering streams in real-time analytics.	2
f.	What considerations should be taken into account when implementing sampling techniques for stream data?	2
g.	How do stream-based algorithms differ from batch processing algorithms in the context of frequent itemset mining?	2
h.	What are the challenges associated with implementing Apriori in scenarios with limited available memory?	2
i.	Explain the role of Hive in the Hadoop ecosystem.	2
j.	How does the MapReduce framework facilitate distributed processing?	2

SECTION B

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

a.	Describe the characteristics of data that are relevant in the field of data analytics. How do these characteristics impact the analysis process?	10
b.	Explain the concept of Bayesian networks and their applications in modeling probabilistic relationships among variables. Discuss how Bayesian networks can be constructed from data and used for reasoning under uncertainty.	10
c.	Provide an overview of Real-time Analytics Platforms applications, emphasizing their role in processing continuous data streams. How do these platforms support the development of real-time analytics solutions?	10
d.	Compare the strengths and weaknesses of hierarchical clustering and K-means clustering. Under what circumstances would one technique be preferred over the other, and why?	10
e.	Discuss the concept of sharding in the context of NoSQL databases. How does sharding contribute to scalability, and what challenges does it address?	10

SECTION C

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

a.	Explain the concept of generalization in neural networks. How does it relate to the trade-off between bias and variance, and what strategies can be employed to enhance generalization performance?	10
b.	Provide a detailed explanation of how fuzzy logic is used to extract models from data. Discuss the advantages of fuzzy modeling in capturing uncertainty and handling imprecise information in comparison to traditional crisp models.	10

PAPER ID-310861

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

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

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

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

a.	In the context of stream data, explain different approaches for counting distinct elements. How do these methods address challenges associated with continuously changing data?	10
b.	Describe the concept of counting uniqueness in a window in the context of stream processing. How does this relate to measuring the frequency and uniqueness of elements within a specified time frame?	10

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

a.	Provide an in-depth comparison between the CLIQUE and ProCLUS clustering algorithms. How do these methods handle challenges such as noise, outliers, and varying cluster shapes?	10
b.	Explore the challenges and considerations when performing clustering in non-Euclidean spaces. How do distance metrics and similarity measures differ in non-Euclidean environments, and what impact does this have on clustering outcomes?	10

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

a.	How do interactive techniques contribute to the exploration and analysis of large datasets? Provide examples of systems or tools that leverage interactive approaches effectively.	10
b.	Discuss the role of NoSQL databases in handling unstructured data. Provide examples of scenarios where NoSQL databases outperform traditional relational databases.	10

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

a.	Differentiate between analysis and reporting in the context of data analytics. How do these two aspects contribute to the overall understanding of data?	10
b.	Explore modern data analytic tools and their functionalities. How have these tools transformed the landscape of data analytics?	10