

Paper Id:

2	3	2	1	7	3
---	---	---	---	---	---

Roll No.

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

B.Tech.
(SEM V) THEORY EXAMINATION 2022-23
INTRODUCTION TO DATA ANALYTICS AND VISUALIZATION

Time: 3 Hours**Total Marks: 100****Note:** Attempt all Sections. If you require any missing data, then choose suitably.**SECTION A****1. Attempt all questions in brief. 2*10 = 20**

- (a) List the features of big data analytics platform.
- (b) Why there is need of data analytics?
- (c) What are the components of time series?
- (d) List various types of regression analysis techniques.
- (e) What is sentiment analysis?
- (f) What are problem in Flajolet-Martin (FM) algorithm?
- (g) What is Market Basket Analysis?
- (h) Explain hierarchical method of clustering.
- (i) Define sharding and database shard.
- (j) List any two visualization tools.

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

- (a) Differentiate between structured, semi-structured and unstructured data.
- (b) Describe grid-based fuzzy model.
- (c) Explain Alon-Matias-Szegedy Algorithm for second moments on the given stream {a, b, c, b, d, a, c, d, a, b, d, c, a, a, b}.
- (d) Explain market basket analysis with an example.
- (e) Discuss various classification of visualization techniques.

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

- (a) Discuss all the modern tools of data analytic.
- (b) Describe all the phases of data analytics life cycle.

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

- (a) Describe PCA algorithm. Compute the principal component using PCA algorithm on given data as:
 CLASS 1: $X = 2, 3, 4$
 $Y = 1, 5, 3$
 CLASS 2:
 $X = 5, 6, 7$
 $Y = 6, 7, 8$
- (b) Illustrate Bayesian network with an example.

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

- (a) Describe briefly the architecture of Data Stream Management System (DSMS).
- (b) Explain any one algorithm for counting oneness in a window.

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

- (a) For the given data, find the association rule using apriori algorithm.

TID	ITEMSETS
T1	A, B
T2	B, D
T3	B, C
T4	A, B, D
T5	A, C
T6	B, C
T7	A, C
T8	A, B, C, E
T9	A, B, C

Given: Minimum Support= 2, Minimum Confidence= 50%

- (b) What do you mean by k-means clustering? How does the k-means algorithm work? Write k-means algorithm for partitioning.

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

- (a) Define HDFS. Discuss the HDFS architecture and HDFS commands in brief.
- (b) Differentiate between Map Reduce and Apache Pig.