SHAMBHUNATH INSTITUTE OF ENGINEERING AND TECHNOLOGY

Subject Code: BCS052 **Subject: DATA ANALYTICS**

Course: B.Tech **SEMESTER: Vth**

FIRST SESSIONAL EXAMINATION, ODD SEMESTER, (2024-2025)

Branch: COMPUTER SCIENCE & ENGINEERING

Time – 2 Hr. **Maximum Marks–45**

NOTE:(Attempt All Sections)

1. Attempt any FIVE of the following.

QN	QUESTION	Marks	CO	BL
a	What is the role of sampling data in a stream?	2	CO3	L2
b	Give the full form of RTAP and discuss its application.	2	CO3	L1
c	What is Bernoulli Sampling.	2	CO3	L2
d	Compare DSMS with DBMS	2	CO3	L1
e	What is DSMS.	2	CO3	L1
f	Write short notes on Sentiment Analysis.	2	CO3	L2

2. AttemptAny ONE of the following.

QN	QUESTION	Marks	CO	BL
a	Explain the architecture of data stream model.	5	CO3	L2
b	Discuss the case study of stock market predictions / Real Time Sentiment Analysis in detail.	5	CO3	L2
c	Explain Datar-Gionis-Indyk-Motwani (DGIM) algorithm for counting oneness in a window.	5	CO3	L2

3. Attempt Any FIVE of the following.

QN	QUESTION	Marks	CO	BL
a	Write down the name of various algorithm for finding frequent itemset.	2	CO4	L2
b	Why PCY algorithm is preferred over Apriori algorithm.	2	CO4	L2
c	Write down a different hash based techniques for improving efficiency of Apriori based mining.	2	CO4	L2
d	Illustrate the K-means algorithm in detail with its disadvantages.	2	CO4	L4
e	Differentiate between CLIQUE and ProCLUS clustering.	2	CO4	L3

f	Explain the principle behind Hierarchical clustering technique.	2	CO4	L2

4. Attempt Any ONE of the following.

QN	QUESTION	Marks	CO	BL
a	What are the different approaches in clustering?	5	CO4	L3
b	Write short note on generating association rules from frequent item sets.	5	CO4	L2
c	A database has 5 transactions. Let minimum support=60% and minimum confidence=80%.	5	CO4	L4

TID	Items Bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

Find all frequent itemsets using Apriori algorithm.

5. Attempt any FIVE of the following.

QN	QUESTION	Marks	CO	BL
a	Differentiate between Pig and Map reduce.	2	CO5	L2
b	Differentiate between data visualization and data analytics.	2	CO5	L3
c	Write down the benefit and drawback of SHARDING.	2	CO5	L2
d	Write down the component of H base.	2	CO5	L2
e	How RDBS is different from NoSQL?	2	CO5	L2
f	List five R functions used in descriptive statistics.	2	CO5	L2

6. AttemptAny ONE of the following.

QN	QUESTION	Marks	CO	BL
a	Draw and discuss the architecture of Hive in detail with its condition and cases.	5	CO5	L2
b	Write R function to check whether the given number is prime or not.	5	CO5	L3
c	Explain the architecture of HDFS and write three commands for Hadoop.	5	CO5	L2