(Academic Year – 2024-25 Sem-II)

UNIT TEST - I

Subject: DSBDA
Class: TE IT
Date: 10 /02 / 2025 Day: Monday
Div.: IX, X, XI
Subject Code: 314452
[Max. Marks: 30]
Duration: 1 Hour

Instructions to the candidate:

1. All questions are compulsory
2. Draw a neat diagram wherever necessary.

Que. No.	Questions	Max Marks	CO Mapped	Bloom's Learning Level
1-a	Explain the concept of Big Data. What are the four V's (Volume, Velocity, Variety, and Veracity) associated with Big Data, and how do they each impact data storage, processing, and analytics?	05	CO-I	L2
1-b	A social media platform generates approximately 250 million posts per day, and the average size of each post is 2 KB. Calculate the total data generated per day in gigabytes (GB). (Assume 1 GB = 1,000,000 KB for simplicity.)	05	CO-I	L3
1-c	Elaborate on the need to re-engineer a traditional data warehouse in the era of Big Data. What are the key considerations and benefits of modernizing a data warehouse?	05	CO-I	L2
2-a	A Bloom filter uses m = 1000 bits and k = 3 hash functions. If n = 50 elements are inserted, derive the approximate false positive probability using the standard Bloom filter formula.	05	CO-II	L4
2-b	There are three stores in a small town. On any given week 200 people visit store A, 120 visits store B and180 visits store C. In the following week from store A 80% will go to same store, from store B 70% will go to same store and from store C 60% will go to same store. People who don't go back to store A 10% go to B and 10% go to C, from B 20% to A and 10% to C and from store C 30% to B and 10% to A. Find the number of customers going to Store A, B and C in the next week and week after that. **Draw suitable FSM and Transition Matrix.**	05	CO-II	L3
2-c	Given the dataset: {3, 5, 7, 7, 9, 10, 12, 12, 12, 15}, compute the mean, median, mode, variance, and standard deviation.	05	CO-II	L2