VIT
Vellore Institute of Technology

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2024-2025

SLOT: C1 + TC1

Programme Name & Branch : B.Tech. & Computer Science (Core and Specialization)
Course Code and Course Name : BCSE318L & Data Privacy
Faculty Name(s) : common to all
Class Number(s) : VL2024250501576, 1565, 1719, 1580, 1604, 1607
Date of Examination : 18.03.2025
Exam Duration : 90 minutes
Maximum Marks: 50

**General instruction(s):**

CO3: . Identify the list of threats on the various types of anonymized data.
CO4: . Classify and analyze the methods of test data generation with Privacy and utility

| Q. No | Question | M | CO | BL |
|---|---|---|---|---|

1.

| RACE | BIRTH | GENDER | ZIP | PROBLEM |
|---|---|---|---|---|
| Black | 1965 | M | 021401 | Short breath |
| Black | 1965 | M | 021402 | Chest Pain |
| Black | 1965 | F | 021301 | Hyper Tension |
| Black | 1965 | F | 021302 | Hyper Tension |
| Black | 1964 | F | 021304 | Obesity |
| Black | 1964 | F | 021305 | Chest Pain |
| White | 1964 | M | 021306 | Chest Pain |
| White | 1964 | M | 021307 | Obesity |
| White | 1967 | M | 021308 | Chest Pain |
| White | 1967 | M | 021309 | Chest Pain |

Discuss the strategies and algorithm for achieving k-Anonymity by using the above dataset as an example with a minimum value of 'k'. (6)

After finding K-anonymity value and anonymized table, if the values are given as shown below, can you identify the problem of Bob? Justify your answer pertaining to the attacks that are possible in K-anonymity Techniques. (4)

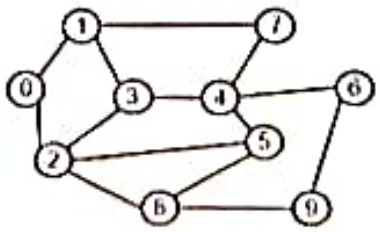
| Bob | |
|---|---|
| *Zipcode* | *Birth* |
| 021305 | 1964 |

10 | 2 | 4

2.

| Date | Open | High | Low | Close | Adj Close | Volume |
|---|---|---|---|---|---|---|
| 2/12/1980 | 0.128348 | 0.128906 | 0.128348 | 0.128348 | 0.100178 | 469033600 |
| 1/9/1981 | 0.142299 | 0.142857 | 0.142299 | 0.142299 | 0.111067 | 21504000 |
| 2/2/1982 | 0.083147 | 0.083147 | 0.082589 | 0.082589 | 0.064463 | 26633600 |

a) Identify the type of above mentioned dataset and explain the suitable data protection methods for the same. — 5 | 3 | 4




b) For the above Graph data, Compute Degree Centrality for each node and betweeness centrality for nodes 1 and 5. — 5

VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

REG.NO.:

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2024-2025

SLOT: C1 + TC1

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 3. | Consider the below data set and answer the following.<br>A. Explain the necessity of moving from k-anonymity to l-diversity along with its limitations. (3)<br>B. Create the diverse dataset using the above dataset as an example. (3)<br>C. Apply earth mover's algorithm on the same data set to find closeness values with respect to the salary attribute. (4) | | | | | | 10 | 2 |

| S.NO | Zip Code | Age | Nationality | Condition | Salary |
|---|---|---|---|---|---|
| 1. | 13053 | 28 | Russian | Heart Disease | 7K |
| 2. | 13068 | 29 | American | Heart Disease | 8K |
| 3. | 13068 | 21 | Japanese | Viral Infection | 9K |
| 4. | 13053 | 23 | American | Viral Infection | 11K |
| 5. | 14853 | 50 | Indian | Cancer | 12K |
| 6. | 14853 | 55 | Russian | Heart Disease | 10K |
| 7. | 14850 | 47 | American | Viral Infection | 5K |
| 8. | 14850 | 49 | American | Viral Infection | 3K |
| 9. | 13053 | 31 | American | Cancer | 2K |
| 10. | 13053 | 37 | Indian | Cancer | 4K |
| 11. | 13068 | 36 | Japanese | Cancer | 13K |
| 12. | 13068 | 35 | American | Cancer | 14K |

| | | | |
|---|---|---|---|
| 4. | Imagine a city government collects data from various sources to optimize services and improve quality of life. This dataset includes:<br>• **Transportation:** GPS coordinates of public transit vehicles, traffic flow data from sensors, ride-sharing data.<br>• **Public Safety:** Crime reports, 911 call logs, surveillance camera footage (anonymized, but with timestamp and location), social media activity.<br>• **Utilities:** Water and energy consumption data, sensor readings from infrastructure.<br>• **Health:** Aggregated (but not truly anonymized) data from public health clinics, air quality sensor readings, anonymized fitness tracker data.<br>• **Demographics:** Census data, voter registration records (with partial anonymization), social service records.<br>Explain the techniques available for identifying the threats to the above scenario at different levels for anonymized data. (6)<br>What information or knowledge does an adversary or attacker use to gain information about a record owner and his sensitive data? illustrate the above scenario with different dimensions of background and external knowledge of an adversary. (4) | 10 | 3 |
| 5. | Business applications and customers are in the same country, and business operations are in a different country. Data protection acts such as European Union (EU) Data Protection Act and Swiss Data Protection Act (FADP) enforce that customer-sensitive data should not cross the geographical boundaries of the country. In such case, how can business operations access customer data and assist their customers? Explain the suitable method and its types with a neat diagram and examples. List out the benefits of the chosen method compared to other methods. | 10 | 4 |