



School of Computer Science and Engineering

Winter Semester 2023-2024

Continuous Assessment Test – 2

SLOT – G1+TG1

Programme Name & Branch : B.Tech. CSE Spec. in Information Security

Course Name & code: Data Privacy – BCSE318L

Class Number (s): VL2023240500995 and VL2023240500997

Faculty Name (s): Dr. Geetha Mary A and Dr. Jasmin T. Jose

Exam Duration: 90 Min.

Maximum Marks: 50

Q.No.	Question	Max Marks	CO	BL																																													
1.	Apply Samariti's Algorithm over the below mentioned data set and convert it into a 2-anonymous data set. <table><tr><th>ID</th><th>Name</th><th>Age</th><th>Location</th><th>Disease</th></tr><tr><td>1</td><td>Tim</td><td>10</td><td>Mysore</td><td>Cancer</td></tr><tr><td>2</td><td>John</td><td>17</td><td>Bangalore</td><td>Pneumonia</td></tr><tr><td>3</td><td>Sam</td><td>28</td><td>Chennai</td><td>Osteoporosis</td></tr><tr><td>4</td><td>Jane</td><td>18</td><td>Mysore</td><td>Sinusitis</td></tr><tr><td>5</td><td>Lincoln</td><td>48</td><td>Trivandrum</td><td>Viral Infection</td></tr><tr><td>6</td><td>Angelina</td><td>30</td><td>Vellore</td><td>Bronchitis</td></tr><tr><td>7</td><td>Ted</td><td>29</td><td>Chennai</td><td>Flu</td></tr><tr><td>8</td><td>Steve</td><td>40</td><td>Cochin</td><td>Gastritis</td></tr></table> Consider the below domain generalization hierarchy for the attributes age and location. Age Hierarchy: - Level 3: 5-25, 26-50 - Level 2: 5-25, 26-35, 36-50 - Level 1: 5, 6, 7, ..., 25; 26, 27, ..., 35; 36, 37, ..., 50 Location Hierarchy: - Level 3: South India - Level 2: Tamil Nadu, Karnataka, Kerala - Level 1: Chennai, Trichy, Bangalore, Mysore, Cochin, Trivandrum Solution: Step by step anonymization according to the lattice has to be done. Final 2-Anonymous dataset is given below.	ID	Name	Age	Location	Disease	1	Tim	10	Mysore	Cancer	2	John	17	Bangalore	Pneumonia	3	Sam	28	Chennai	Osteoporosis	4	Jane	18	Mysore	Sinusitis	5	Lincoln	48	Trivandrum	Viral Infection	6	Angelina	30	Vellore	Bronchitis	7	Ted	29	Chennai	Flu	8	Steve	40	Cochin	Gastritis	10	CO2	BL3
ID	Name	Age	Location	Disease																																													
1	Tim	10	Mysore	Cancer																																													
2	John	17	Bangalore	Pneumonia																																													
3	Sam	28	Chennai	Osteoporosis																																													
4	Jane	18	Mysore	Sinusitis																																													
5	Lincoln	48	Trivandrum	Viral Infection																																													
6	Angelina	30	Vellore	Bronchitis																																													
7	Ted	29	Chennai	Flu																																													
8	Steve	40	Cochin	Gastritis																																													

	<table><tr><td>Age</td><td>Location</td><td>Disease</td></tr><tr><td>5-25</td><td>Karnataka</td><td>Cancer</td></tr><tr><td>5-25</td><td>Karnataka</td><td>Pneumonia</td></tr><tr><td>26-35</td><td>Tamil Nadu</td><td>Osteoporosis</td></tr><tr><td>5-25</td><td>Karnataka</td><td>Sinusitis</td></tr><tr><td>36-50</td><td>Kerala</td><td>Viral Infection</td></tr><tr><td>26-35</td><td>Tamil Nadu</td><td>Bronchitis</td></tr><tr><td>26-35</td><td>Tamil Nadu</td><td>Flu</td></tr><tr><td>36-50</td><td>Kerala</td><td>Gastritis</td></tr></table>	Age	Location	Disease	5-25	Karnataka	Cancer	5-25	Karnataka	Pneumonia	26-35	Tamil Nadu	Osteoporosis	5-25	Karnataka	Sinusitis	36-50	Kerala	Viral Infection	26-35	Tamil Nadu	Bronchitis	26-35	Tamil Nadu	Flu	36-50	Kerala	Gastritis																															
Age	Location	Disease																																																									
5-25	Karnataka	Cancer																																																									
5-25	Karnataka	Pneumonia																																																									
26-35	Tamil Nadu	Osteoporosis																																																									
5-25	Karnataka	Sinusitis																																																									
36-50	Kerala	Viral Infection																																																									
26-35	Tamil Nadu	Bronchitis																																																									
26-35	Tamil Nadu	Flu																																																									
36-50	Kerala	Gastritis																																																									
2.	Address the possible attacks on the data after applying the below mentioned method using relevant examples. a. k-anonymous [2 Mark] b. l-diverse [4 Mark] c. Perturbation of time series data with white noise [2 Mark] d. Perturbation of time series data with correlated noise. [2 Mark] Solution: The attacks given below has to be discussed with examples in detail. a. K-anonymous – Homogeneity Attacks b. L-diverse - skewness attack, similarity attack c. Perturbation of time series data with white noise – Filtering Attacks d. Perturbation of time series data with correlated noise – Regression attacks	10	CO3	BL4																																																							
3.	a. Consider any data set and convert it into the following: i. k-anonymous data set. [2.5 Mark] ii. l-diverse data set. [2.5 Mark] b. Consider the below mentioned 2-anonymous system and calculate classification metric. [5 Mark] <table><tr><td>Race</td><td>Date of birth</td><td>Gender</td><td>Zip</td><td>Problem</td></tr><tr><td>black</td><td>1965</td><td>male</td><td>2141</td><td>Asthma</td></tr><tr><td>black</td><td>1965</td><td>male</td><td>2141</td><td>Pulmonary vascular</td></tr><tr><td>black</td><td>1965</td><td>female</td><td>2138</td><td>Lung cancer</td></tr><tr><td>black</td><td>1965</td><td>female</td><td>2138</td><td>Mouth cancer</td></tr><tr><td>black</td><td>1965</td><td>female</td><td>2138</td><td>Lung cancer</td></tr><tr><td>white</td><td>1964</td><td>male</td><td>2139</td><td>Lung cancer</td></tr><tr><td>white</td><td>1964</td><td>male</td><td>2139</td><td>Coronary artery</td></tr><tr><td>white</td><td>1964</td><td>male</td><td>2139</td><td>Lung cancer</td></tr><tr><td>white</td><td>1967</td><td>male</td><td>2138</td><td>Pulmonary vascular</td></tr><tr><td>white</td><td>1967</td><td>male</td><td>2138</td><td>Asthma</td></tr></table> Solution:	Race	Date of birth	Gender	Zip	Problem	black	1965	male	2141	Asthma	black	1965	male	2141	Pulmonary vascular	black	1965	female	2138	Lung cancer	black	1965	female	2138	Mouth cancer	black	1965	female	2138	Lung cancer	white	1964	male	2139	Lung cancer	white	1964	male	2139	Coronary artery	white	1964	male	2139	Lung cancer	white	1967	male	2138	Pulmonary vascular	white	1967	male	2138	Asthma	10	CO2	BL3
Race	Date of birth	Gender	Zip	Problem																																																							
black	1965	male	2141	Asthma																																																							
black	1965	male	2141	Pulmonary vascular																																																							
black	1965	female	2138	Lung cancer																																																							
black	1965	female	2138	Mouth cancer																																																							
black	1965	female	2138	Lung cancer																																																							
white	1964	male	2139	Lung cancer																																																							
white	1964	male	2139	Coronary artery																																																							
white	1964	male	2139	Lung cancer																																																							
white	1967	male	2138	Pulmonary vascular																																																							
white	1967	male	2138	Asthma																																																							

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	13053	28	Russian	Heart Disease
2	13068	29	American	Heart Disease
3	13068	21	Japanese	Viral Infection
4	13053	23	American	Viral Infection
5	14853	50	Indian	Cancer
6	14853	55	Russian	Heart Disease
7	14850	47	American	Viral Infection
8	14850	49	American	Viral Infection
9	13053	31	American	Cancer
10	13053	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	35	American	Cancer

Figure 1. Inpatient Microdata

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	130***	< 30	*	Heart Disease
2	130***	< 30	*	Heart Disease
3	130***	< 30	*	Viral Infection
4	130***	< 30	*	Viral Infection
5	1485*	≥ 40	*	Cancer
6	1485*	≥ 40	*	Heart Disease
7	1485*	≥ 40	*	Viral Infection
8	1485*	≥ 40	*	Viral Infection
9	130***	3*	*	Cancer
10	130***	3*	*	Cancer
11	130***	3*	*	Cancer
12	130***	3*	*	Cancer

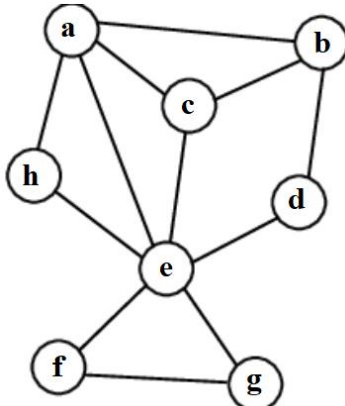
Figure 2. 4-anonymous Inpatient Microdata

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	1305*	≤ 40	*	Heart Disease
4	1305*	≤ 40	*	Viral Infection
9	1305*	≤ 40	*	Cancer
10	1305*	≤ 40	*	Cancer
5	1485*	> 40	*	Cancer
6	1485*	> 40	*	Heart Disease
7	1485*	> 40	*	Viral Infection
8	1485*	> 40	*	Viral Infection
2	1306*	≤ 40	*	Heart Disease
3	1306*	≤ 40	*	Viral Infection
11	1306*	≤ 40	*	Cancer
12	1306*	≤ 40	*	Cancer

Figure 3. 3-Diverse Inpatient Microdata

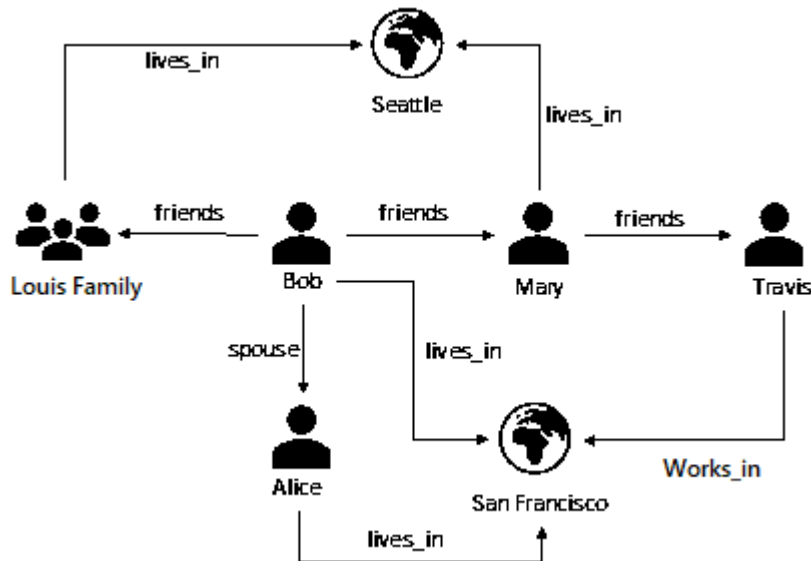
b.

Race	Date of birth	Gender	Zip	Problem	Penalty
black	1965	male	2141	Asthma	1
black	1965	male	2141	Pulmonary vascular	1
black	1965	female	2138	Lung cancer	0
black	1965	female	2138	Mouth cancer	1
black	1965	female	2138	Lung cancer	0
white	1964	male	2139	Lung cancer	0
white	1964	male	2139	Coronary artery	1
white	1964	male	2139	Lung cancer	0

	<table><tr><td>white</td><td>1967</td><td>male</td><td>2138</td><td>Pulmonary vascular</td><td>1</td></tr><tr><td>white</td><td>1967</td><td>male</td><td>2138</td><td>Asthma</td><td>1</td></tr></table> $CM = \frac{\sum_{All Rows} \text{penalty}(\text{row } r)}{N}$ Classification Metric = 6/10 = 0.6	white	1967	male	2138	Pulmonary vascular	1	white	1967	male	2138	Asthma	1									
white	1967	male	2138	Pulmonary vascular	1																	
white	1967	male	2138	Asthma	1																	
4.	Consider the below mentioned graph.  Calculate: a. Degree of centrality of each node. [2 Mark] b. Betweenness of node f. [4 Mark] c. Betweenness of node e. [4 Mark] Solution: a. Centrality <table><tr><td>Node</td><td>a</td><td>b</td><td>c</td><td>d</td><td>e</td><td>f</td><td>g</td><td>h</td></tr><tr><td>Centrality</td><td>4</td><td>3</td><td>3</td><td>2</td><td>6</td><td>2</td><td>2</td><td>2</td></tr></table> b. Betweenness of node f = 0 (Steps are expected) c. Betweenness of node e = 10 (Steps are expected)	Node	a	b	c	d	e	f	g	h	Centrality	4	3	3	2	6	2	2	2	10	CO2	BL3
Node	a	b	c	d	e	f	g	h														
Centrality	4	3	3	2	6	2	2	2														

5.

a. Consider the below graph and apply any two link protection methods. Exhibit the output of each method and shortly describe the procedure followed. [6 Mark]



b. Identify the type data of the below mentioned dataset and mention challenges in that type of data. [4 Mark]

Person_ID	Location	Salary	age
1	Vellore	20000	26
1	Chennai	25000	27
2	Chennai	28000	30
2	Chennai	29000	31
3	Ambur	20000	22
3	Vellore	25000	23

Solution:

- a. Link Protection techniques are naïve anonymization, random perturbation and clustering.

The graph has to be converted based on any of the two methods. Description and Output for each method is expected.

- b. The given dataset was longitudinal data.

Challenges of longitudinal data are

- Identity disclosure—prevent record linkage.
- Attribute disclosure—prevent sensitive data linkage.
- Correlation in the cluster.
- Pattern in the clustered data set.
- Take into account that the records in the clustered data set cannot be treated independent of each other as in the case of multidimensional data (relational data).
- Unknown background knowledge of the adversary.

10

CO2

BL6