



# Daffodil International University

Faculty of Science & Information Technology

Department of Computer Science and Engineering

Final Examination, Spring 2024

Course Code: CSE321, Course Title: Data Mining and Machine Learning

Level: 3 Term: 2 Batch: 60

Time: 02:00 Hrs

Marks: 40

## Answer ALL Questions

*[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]*

1.	a)	<p>Consider the following dataset of different businesses and their sales-related information.</p> <table border="1"> <thead> <tr> <th>Company</th><th>Clients</th><th>Rate of Return</th><th>Sales</th></tr> </thead> <tbody> <tr><td>A</td><td>150</td><td>15.4</td><td>50400200</td></tr> <tr><td>B</td><td>144</td><td>11.3</td><td>42100650</td></tr> <tr><td>C</td><td>120</td><td>9.9</td><td>39440420</td></tr> <tr><td>D</td><td>110</td><td>12.5</td><td>36500520</td></tr> <tr><td>E</td><td>100</td><td>9.7</td><td>40650005</td></tr> <tr><td>F</td><td>99</td><td>15.2</td><td>45665230</td></tr> <tr><td>G</td><td>56</td><td>9.2</td><td>25978080</td></tr> </tbody> </table> <p>Apply K-means clustering to the above dataset and show the first 4 iterations where K=3 considering the initial centroids as <math>C_1(150, 15.4, 50400200)</math>, <math>C_2(110, 12.5, 36500520)</math>, <math>C_3(56, 9.2, 25978080)</math> and the Euclidean distance as the measure of distance.</p>	Company	Clients	Rate of Return	Sales	A	150	15.4	50400200	B	144	11.3	42100650	C	120	9.9	39440420	D	110	12.5	36500520	E	100	9.7	40650005	F	99	15.2	45665230	G	56	9.2	25978080	[8]	CO3
Company	Clients	Rate of Return	Sales																																	
A	150	15.4	50400200																																	
B	144	11.3	42100650																																	
C	120	9.9	39440420																																	
D	110	12.5	36500520																																	
E	100	9.7	40650005																																	
F	99	15.2	45665230																																	
G	56	9.2	25978080																																	
b)																																				
2.	a)	Explain why the 'Apriori principle' is useful in association rule mining.	[2]	CO3																																
	b)	Consider the following market basket transactions.	[8]																																	
		<table border="1"> <thead> <tr> <th>Transactions ID</th><th>Items</th></tr> </thead> <tbody> <tr><td>T1</td><td>Coffee, Sugar, Cream</td></tr> <tr><td>T2</td><td>Tea, Biscuits, Honey</td></tr> <tr><td>T3</td><td>Coffee, Cream, Cake</td></tr> <tr><td>T4</td><td>Tea, Honey, Biscuits</td></tr> <tr><td>T5</td><td>Sugar, Coffee, Cake</td></tr> <tr><td>T6</td><td>Tea, Biscuits, Cream</td></tr> <tr><td>T7</td><td>Coffee, Sugar, Biscuits</td></tr> <tr><td>T8</td><td>Tea, Cream, Honey</td></tr> <tr><td>T9</td><td>Cake, Coffee, Cream</td></tr> <tr><td>T10</td><td>Biscuits, Tea, Sugar</td></tr> </tbody> </table> <p>Find all the frequent item sets using the Apriori algorithm and generate the significant decision rules considering the <u>minimum support threshold as 40%</u> and <u>minimum confidence threshold as 60%</u></p>	Transactions ID	Items	T1	Coffee, Sugar, Cream	T2	Tea, Biscuits, Honey	T3	Coffee, Cream, Cake	T4	Tea, Honey, Biscuits	T5	Sugar, Coffee, Cake	T6	Tea, Biscuits, Cream	T7	Coffee, Sugar, Biscuits	T8	Tea, Cream, Honey	T9	Cake, Coffee, Cream	T10	Biscuits, Tea, Sugar												
Transactions ID	Items																																			
T1	Coffee, Sugar, Cream																																			
T2	Tea, Biscuits, Honey																																			
T3	Coffee, Cream, Cake																																			
T4	Tea, Honey, Biscuits																																			
T5	Sugar, Coffee, Cake																																			
T6	Tea, Biscuits, Cream																																			
T7	Coffee, Sugar, Biscuits																																			
T8	Tea, Cream, Honey																																			
T9	Cake, Coffee, Cream																																			
T10	Biscuits, Tea, Sugar																																			

3.	a) Consider the following dataset of different types of foods.	[6]	CO2																																												
	<table border="1"> <thead> <tr> <th>Temperature</th> <th>Taste</th> <th>Size</th> <th>Appealing</th> </tr> </thead> <tbody> <tr><td>Hot</td><td>Salty</td><td>Small</td><td>No ..</td></tr> <tr><td>Cold</td><td>Sweet</td><td>Large</td><td>No ..</td></tr> <tr><td>Cold</td><td>Sweet</td><td>Large</td><td>No ..</td></tr> <tr><td>Cold</td><td>Sour</td><td>Small</td><td>Yes ..</td></tr> <tr><td>Hot</td><td>Sour</td><td>Small</td><td>Yes ..</td></tr> <tr><td>Hot</td><td>Salty</td><td>Large</td><td>No ..</td></tr> <tr><td>Hot</td><td>Sour</td><td>Large</td><td>Yes ..</td></tr> <tr><td>Cold</td><td>Sweet</td><td>Small</td><td>Yes ..</td></tr> <tr><td>Cold</td><td>Sweet</td><td>Small</td><td>Yes ..</td></tr> <tr><td>Hot</td><td>Salty</td><td>Large</td><td>No ..</td></tr> </tbody> </table>	Temperature	Taste	Size	Appealing	Hot	Salty	Small	No ..	Cold	Sweet	Large	No ..	Cold	Sweet	Large	No ..	Cold	Sour	Small	Yes ..	Hot	Sour	Small	Yes ..	Hot	Salty	Large	No ..	Hot	Sour	Large	Yes ..	Cold	Sweet	Small	Yes ..	Cold	Sweet	Small	Yes ..	Hot	Salty	Large	No ..		
Temperature	Taste	Size	Appealing																																												
Hot	Salty	Small	No ..																																												
Cold	Sweet	Large	No ..																																												
Cold	Sweet	Large	No ..																																												
Cold	Sour	Small	Yes ..																																												
Hot	Sour	Small	Yes ..																																												
Hot	Salty	Large	No ..																																												
Hot	Sour	Large	Yes ..																																												
Cold	Sweet	Small	Yes ..																																												
Cold	Sweet	Small	Yes ..																																												
Hot	Salty	Large	No ..																																												
	Demonstrate whether the following foods will be appealing or not using the Naïve Bayes classifier.																																														
	<table border="1"> <thead> <tr> <th>Temperature</th> <th>Taste</th> <th>Size</th> </tr> </thead> <tbody> <tr><td>Cold</td><td>Sweet</td><td>Large</td></tr> <tr><td>Hot</td><td>Salty</td><td>Small</td></tr> </tbody> </table>	Temperature	Taste	Size	Cold	Sweet	Large	Hot	Salty	Small																																					
Temperature	Taste	Size																																													
Cold	Sweet	Large																																													
Hot	Salty	Small																																													
b)	Why is the Naïve Bayes classifier called "Naïve"? How will you perform the classification task if that assumption is not true?	[2]																																													
c)	How can you solve the problem of one of the conditional probabilities becoming zero in the Naïve Bayes classifier?	[2]																																													
4.	a) Consider the following feed-forward neural network where all the hidden layers and the output layer have the sigmoid function as the activation function.	[8]	CO3																																												
	Find out the outputs $Y_1$ and $Y_2$ of the above feed-forward neural network if $X_1=1$ and $X_2=0$ are given to the network as inputs.																																														
b)	Differentiate between Supervised and Unsupervised learning	[2]																																													



# Daffodil International University

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Final Examination, Fall-2023

Course Code: CSE321, Course Title: Data Mining and Machine Learning

Level: 3 Term: 2 Batch: 58 and 59

Time: 2 Hours

Marks: 40

**Answer ALL Questions [Optional]**

*[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]*

1. Apply Algorithms for the following dataset is used to predict whether an animal is a pet or not and compare the performances.

CO3  
[15]

ID	Animals	Size of Animal	Body Color	Can we pet them
0	Dog	Medium	Black	Yes
1	Dog	Big	White	No
2	Rat	Small	White	Yes
3	Cow	Big	White	Yes
4	Cow	Small	Brown	No
5	Cow	Big	Black	Yes
6	Rat	Big	Brown	No
7	Dog	Small	Brown	Yes
8	Dog	Medium	Brown	Yes
9	Cow	Medium	White	No
10	Dog	Small	Black	Yes
11	Rat	Medium	Black	No
12	Rat	Small	Brown	No
13	Cow	Big	White	Yes

If we consider the following test data, test = (Cow, Medium, Black). What it will predict?

2. A market basket transactions data is shown in Table 1 where we have 10 transactions of different computer products and need to find out the frequent item sets from the following table. Then, find the confidence of the generated rules based on the support=60% and confidence=60%.

CO4  
[15]

P.T.O

Table 1. Market basket transactions

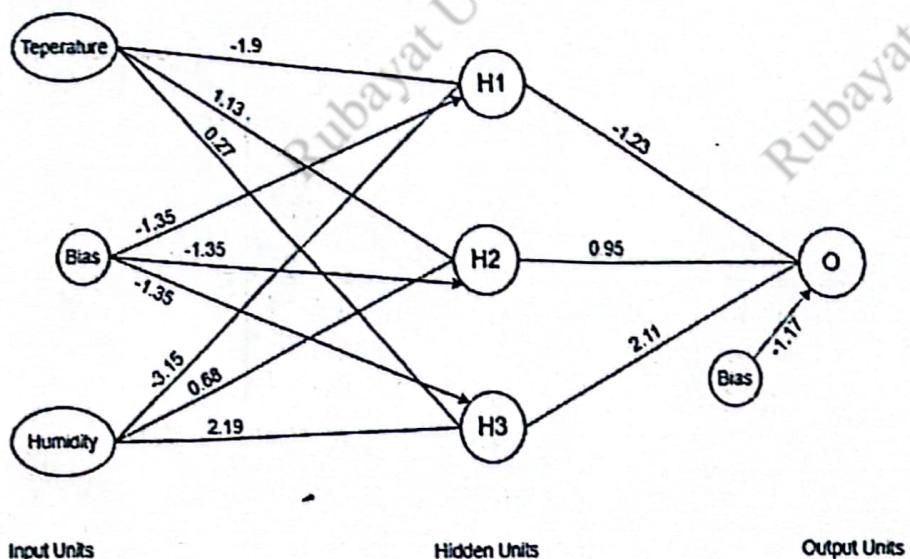
Transaction ID	Items Bought
1	{Laptop, Printer, Tablet, Headset}
2	{Printer, Monitor, Tablet}
3	{Laptop, Printer, Tablet, Headset}
4	{Laptop, Monitor, Tablet, Headset}
5	{Printer, Monitor, Tablet, Headset}
6	{Printer, Tablet, Headset}
7	{Monitor, Tablet}
8	{Laptop, Printer, Monitor}
9	{Laptop, Tablet, Headset}
10	{Printer, Tablet}

3. An artificial neural network has been trained on a binary classification dataset where the independent attributes are Temperature and Humidity, whose target variables are Sunny and Overcast, encoded as 0 and 1, respectively, before training the network. In this network, the hidden nodes are activated using Rectified Linear Units (ReLU), and the last node is activated using Sigmoid. The trained network is shown below. Classify the data shown in Table 1 using this network.

CO4  
[10]

Table 1: Test Data

Temperature	Humidity
1.05	1.8
2.3	1.5



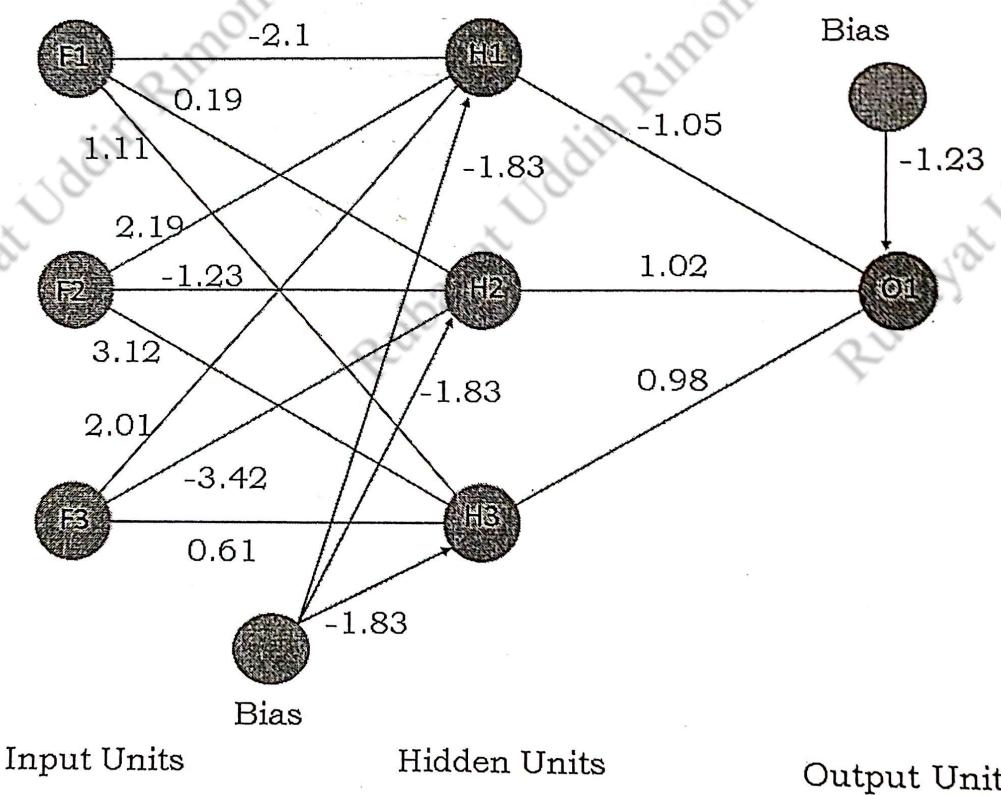


Daffodil International University  
Department of Computer Science and Engineering  
Faculty of Science and Information Technology

Final Examination Semester: Spring 2023 Section: All  
Course Code: CSE321 Course Title: Data Mining and Machine Learning  
Time— 1.5 hours Full Marks – 40

[N.B. Answer all questions. Marks of all questions have been indicated on the right side of the question paper.]

- An artificial neural network has been trained on a binary classification dataset, whose target variables are red and green, encoded as 0 and 1, respectively, prior to training the network. In this network, the hidden nodes are activated using Rectified Linear Units (ReLU), and the last node is activated using Sigmoid. The trained network is shown below. Classify the data shown in Table 1 using this network. CO4 10



F1	F2	F3
2.1	1.9	1.02
3.4	.04	1.2

Table 1: Test data

(2)

Below is a dataset with three features: Age, Annual Income, and Spending Score. Using a clustering approach, identify the number of clusters, boundary points, and outliers in the dataset.

CO2  
5+5+5

Age	Annual Income	Spending Score
19	15	39
21	15	81
23	16	77
31	17	40
21	600	1

(3)

Consider the transaction database shown in the following table.

CO3  
15

ID	Items
1	A, <del>B, C, D</del>
2	G, <del>B, C, D</del>
3	A, C, <del>D, G</del>
4	A, B, F, H
5	<del>B, D, N, E</del>

Let minimum support be 50% and minimum confidence be 65%. Predict all frequent item sets and the association rules.



Daffodil International University  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology

Final Examination, Fall 2022

Course Code: CSE321, Course Title: Data Mining and Machine Learning

Level: 3 Term: 2 Batch: 56

Time: 2:00 Hrs

Marks: 40

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1. Consider the following data for A and B:

$$A = \{184, 170, 165, 176, 181, 187\}$$

$$B = \{71, 55, 61, 67, 71, 77\}$$

You need to arrange those points into 2 groups. Initial values of  $m_1 = (184, 71)$  &  $m_2 = (170, 55)$ .

- a) Evaluate the K-MEANS clustering algorithm for the following data.

7

- b) Compare between Clustering and Classification.

3

CO3

2. Consider the following four sets of items (itemsets) bought together:

Transaction ID	Purchased Items
1	A, B, C
2	D, B, E
3	A, D, B, E
4	D, E

Now, answer the following questions by applying the association algorithm to these shopping transactions.

- a) Using the support threshold  $s = 0.5$ , estimate the frequent itemsets.

8

- b) Using the confidence threshold  $c = 0.95$ , estimate the association rules that are generated.

12

CO3

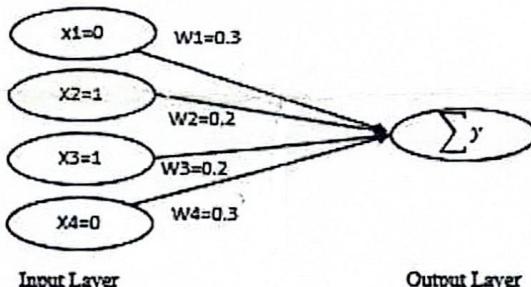
- c) Differentiate what initiatives the seller could take (based on the findings of the questions 1(a) & 1(b)) that would help to boost up his business.

4

- d) Consider the basic neural network you designed for an experiment.

4+2

CO4



Now, estimate the value of Y for this network and if you want to apply the network for research purpose then classify the activation function which you want to apply.



Daffodil International University  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Final Examination, Spring 2022

Course Code: CSE321 (Day), Course Title: Data Mining and Machine Learning  
Sections, Teachers: All

Time: 2:00 Hrs

Marks: 40

**Answer ALL Questions**

*[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]*

1. Consider the following data for persons **Height** and **Weight**:

$$\text{Height} = \{184, 170, 165, 176, 181, 187\}$$

$$\text{Weight} = \{71, 55, 61, 67, 71, 77\}$$

You need to arrange those points into 2 groups. Initial values of  $m_1 = (184, 71)$  &  $m_2 = (170, 55)$ .

a) Now, apply the K-MEAN clustering algorithm for the following data.

7

b) Compare between Clustering and Classification.

3

CO3

2. Consider the following five sets of items (itemsets) bought together from “**City Cafe**” restaurant:

TID	Item set
T1	Kabab, Luchi, Khichuri, Beef
T2	Kabab, Luchi, Chicken Bhuna
T3	Khichuri, Beef, Bhorta
T4	Khichuri, Kabab, Beef, Chichken Bhuna
T5	Tehari, Beef, Kabab

Now, answer the following questions by applying the Apriori algorithm to these shopping transactions.

a) Using the support threshold  $s = 0.45$ , find the frequent itemsets.

8

b) Using the confidence threshold  $c = 0.85$ , find the association rules that are generated.

12

c) Depending on the itemsets and questions of 2(a) and 2(b) findings how association mining for “**City Cafe**” restaurant that would help to enlarge restaurant owner business.

4

CO3

a) Suppose, you are working as a Research Scientist in a health-related Government Office. The government have a large number of COVID patient data. We know that COVID patients face various common symptoms. Recent research shows that these COVID symptoms depend on the COVID variant. Recently, the government approved a project on COVID patients. They want to make different groups of those COVID patients using the COVID variant. Now, If you worked as a Researcher on this project, explain the research method that you will follow for this project using your Data Mining and Machine Learning course knowledge.

6

CO4



Day

**Daffodil International University**

Department of Computer Science and Engineering

Faculty of Science &amp; Information Technology

Final Examination

Semester: Fall 2021

Course Code: CSE 321

Course Title: Data Mining and Machine Learning

Course Teacher: ALL

**Time: 2 hours****Full Marks: 40****Answer all of the following *three* questions. Figures in the right-hand margin indicate full marks.**

1. Consider the following four sets of items (itemsets) bought together:

Transaction ID	Purchased Items
1	A, B, C
2	D, B, E
3	A, D, B, E
4	D, E

Now, answer the following questions by applying the Apriori algorithm to these shopping transactions.

- a) Using the support threshold  $s = 0.5$ , find the frequent itemsets. 8
- b) Using the confidence threshold  $c = 0.95$ , find the association rules that are generated. 12
- c) Suggest what initiatives the seller could take (based on the findings of the questions 1(a) & 1(b)) that would help to boost up his business. 4

2. Consider the following proximity matrix: 3 + 3

	I1	I2	I3	I4	I5
I1	1.00	0.90	0.10	0.65	0.20
I2	0.90	1.00	0.70	0.60	0.50
I3	0.10	0.70	1.00	0.40	0.30
I4	0.65	0.60	0.40	1.00	0.80
I5	0.20	0.50	0.30	0.80	1.00

Now, apply the following agglomerative clustering algorithms to this matrix.

- i) MIN or single link  
ii) MAX or complete linkage

3. Write the answer to the following questions in *a single sentence*.
- a) How many possible candidate itemsets are there if  $d$  items are given? 1
  - b) Why is matching each transaction against every candidate computationally expensive in the brute-force approach? 1
  - c) Write a mathematical relation between  $k$  (from  $k$ -itemset) and  $w$  (maximum transaction width)? 1
  - d) Given a transaction  $t$  of  $n$  items, what are the possible subsets of size 3? 1
  - e) Write the anti-monotone property of support. 1
  - f) Write the main two objectives of cluster analysis. 1
  - g) Why cannot bisecting  $K$ -means clustering algorithm yield an empty cluster? 1
  - h) Write the equation of proximity of the agglomerative clustering algorithm (group average). 1
  - i) Write the names of the three types of points used in the DBSCAN algorithm. 1
  - j) Write the equation of density in the DBSCAN algorithm. 1



## Daffodil International University Department of Computer Science and Engineering

### Faculty of Science & Information Technology

Semester Final Examination, Summer-2021 @ DIU Blended Learning Center  
Course Code: CSE 321 (Day), Course Title: Data Mining and Machine Learning

Level: 3      Term: 2      Section: E

Instructor: DMR      Modality: Open Book Exam

Date: Sunday 29 August, 2021      Time: 01:30 pm - 05:00 pm

**Three and a half hours (3.5 hrs) to support online open/case study based assessment      Marks: 40**

**Directions:**

- Students need to go through the CASE STUDY shown in this exam paper.
- Analyze and answer specific section based on your own thinking and work.
- Do not share as this will be treated as plagiarism by Blended Learning Center.

Answer all of the following questions. Figures in the right-hand margin indicate full marks.

1.	Body weight, age and smoking have been proven to affect bone mineral density (BMD) directly or indirectly. It is expected that the data points representing high BMD are similar as like the data points of low BMD. Ten (10) such data points of high or low BMD are collected which have been provided below.  <table border="1"><thead><tr><th>Sl #</th><th>Age (Years)</th><th>Weight (Kg)</th><th>Smoking (0/1)</th></tr></thead><tbody><tr><td>1</td><td>15</td><td>32</td><td>0</td></tr><tr><td>2</td><td>10</td><td>23</td><td>0</td></tr><tr><td>3</td><td>41</td><td>44</td><td>0</td></tr><tr><td>4</td><td>49</td><td>77</td><td>0</td></tr><tr><td>5</td><td>57</td><td>72</td><td>1</td></tr><tr><td>6</td><td>25</td><td>43</td><td>1</td></tr><tr><td>7</td><td>37</td><td>71</td><td>1</td></tr><tr><td>8</td><td>61</td><td>67</td><td>0</td></tr><tr><td>9</td><td>33</td><td>50</td><td>0</td></tr><tr><td>10</td><td>23</td><td>70</td><td>0</td></tr></tbody></table> Now, answer the following questions based on the above data set:	Sl #	Age (Years)	Weight (Kg)	Smoking (0/1)	1	15	32	0	2	10	23	0	3	41	44	0	4	49	77	0	5	57	72	1	6	25	43	1	7	37	71	1	8	61	67	0	9	33	50	0	10	23	70	0	
Sl #	Age (Years)	Weight (Kg)	Smoking (0/1)																																											
1	15	32	0																																											
2	10	23	0																																											
3	41	44	0																																											
4	49	77	0																																											
5	57	72	1																																											
6	25	43	1																																											
7	37	71	1																																											
8	61	67	0																																											
9	33	50	0																																											
10	23	70	0																																											
a)	Apply preprocessing on the data set except outlier elimination.	5																																												
b)	Apply K-means clustering algorithm on the preprocessed data for $K = 2$ up to 3 iterations.	10																																												
c)	Apply K-means clustering algorithm with incremental approach for updating the centroids after each assignment on the preprocessed data for $K = 2$ up to 1 iteration.	5																																												

2.	Write the answer to the following questions in <b>a single sentence</b> .	
	<b>a)</b> What is the problem of using an even value of $k$ in the $k$ -NN classifier?	<b>1</b>
	<b>b)</b> What is the reason that has led the Bayesian Belief Network to emerge?	<b>1</b>
	<b>c)</b> What is the necessity of using scaling in $k$ -NN?	<b>1</b>
	<b>d)</b> Write a mathematical relation between Manhattan distance and Euclidean distance.	<b>1</b>
	<b>e)</b> Why is a dendrogram not applicable on $K$ -means clustering algorithm?	<b>1</b>
	<b>f)</b> What is the appropriacy of using minimum spanning tree (MST) other than all other types of trees to divisive hierarchical clustering?	<b>1</b>
	<b>g)</b> What are the observations, for which the size of proximity matrix can be reduced from $m^2$ to about $m^2/2$ ?	<b>1</b>
	<b>h)</b> Why is the matching each transaction against every candidate computationally expensive in brute-force approach?	<b>1</b>
	<b>i)</b> Write a mathematical relation between $k$ (from $k$ -itemset) and $w$ (maximum transaction width)?	<b>1</b>
	<b>j)</b> Given a transaction $t$ of $n$ items, what are the possible subsets of size 3?	<b>1</b>
	<b>k)</b> If number of items, $d = 3$ is given, calculate the total number of possible association rules in brute-force approach using two different ways.	<b>3</b>
3.	<b>a.</b> Describe the data set of the project that you have done in the corresponding lab course of this course “Data Mining and Machine Learning”.	<b>4</b>
	<b>b.</b> What is advantage of Naïve Bayes over decision tree?	<b>3</b>

----- Good Luck -----