



Daffodil International University
Department of Computer Science and Engineering

Faculty of Science & Information Technology

Final Examination, Fall 2020 @ DIU Blended Learning Center

Course Code: CSE450 (Day), Course Title: Data Mining

Level: 4 Term: 3 Section: PC-A

Instructor: SMAH Modality: Open Book Exam

Date: Thursday 17 December, 2020 Time: 02:00pm-06:00pm

Four hours (4:00) 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.**

Case-1 of 2:

Shwapno is one of the largest retail chains in Bangladesh. Recently, due to the pandemic caused by COVID-19, their sales around the country has been decreased. The management of the chain has decided to apply some data analytics for increasing the sales. Shwapno typically preserves all its sales data in their database. Now they are looking for some expert in the data analytic field and they contacted you in this regard. On a pilot basis, they gave you the following transaction data. Now your task is to analyze this data based on the following questions.

Transaction ID	Items Bought
1	{Milk, Beer, Diapers}
2	{Bread, Butter, Milk}
3	{Milk, Diapers, Cookies}
4	{Bread, Butter, Cookies}
5	{Beer, Cookies, Diapers}
6	{Milk, Diapers, Bread, Butter}
7	{Bread, Butter, Diapers}
8	{Beer, Diapers}
9	{Milk, Diapers, Bread, Butter}
10	{Beer, Milk}

- Q1. For minimal support of 0.6, find all the frequent item sets (8)
- Q2. Considering the above minimal support and minimal confidence of 0.7, find association rules of the form $item1 \rightarrow \{item2, item3\}$ (7)
- Q3. Assuming minimal support $> zero$, in k -itemset, what will be the maximum value of k ? (5)
- Q4. For the above dataset, how many pair of items a & b can be found such that $\{a\} \rightarrow \{b\}$ and $\{b\} \rightarrow \{a\}$ have the same confidence. (5)

Case-2 of 2:

Customer analysis is an important part for the growth of a business. It helps to design new campaign for existing customers and attract new ones. It further helps to understand customer needs and improve services accordingly. Typically, demographic data is used to understand customers and their characteristics.

Table. 1: Inputs for the demographic data

Feature	Demographic	# Values	Type
1	Sex	2	Categorical
2	Marital status	5	Categorical
3	Age	7	Ordinal
4	Education	6	Ordinal
5	Occupation	9	Categorical
6	Income	9	Ordinal

Q5. Table. 1 describes six demographic data, their possible values and corresponding types. Transform this table into a real dataset with at least possible number of values, for example, for the feature “Sex”, since its #Values is 2, provide at least two data (one for Male and one for Female) into the dataset. Similarly, there will be 5 data for the feature 2 and so on. Finally, draw the whole dataset with all the features. **(10)**

Q6. Propose a suitable data mining technique for analyzing the above data. Briefly describe how the technique can help to better understand the customers and how it can help to grow a business. **(5)**