7th Semester DATA MINING SESSIONAL EXAM -1, 2021

Time : 1 hour 15 Minutes (9:00 -10:15 A.M., Uploading 10:15 to 10:30 a.m. Max. Marks: 15

Answer all questions

Instructions: Use your brain to design the datasets in the best possible manner so as to complete in the required time. Credit will be given based on your original design and results.

Upload the answer as a single file before 10:30 a.m. In case of any problem you may send the answer sheets on my email. Don't send by WhatsApp.

Q.1. Consider the market basket transactions shown in Table.

TID	Items
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, Beer, Diapers, Bread, Butter}
10	{Beer, Cookies}

(a) What is the maximum number of association rules that can be extracted from this data (including rules that have zero support)?

(b) What is the maximum size of frequent itemsets that can be extracted (assuming minsup > 0)?

(c) Write an expression for the maximum number of size-3 itemsets that can be derived from this data set.

(d) Find all frequent itemsets using Minsup=2 using any algorithm.

Q.2. Consider a sequence dataset containing at least 4 sequences which are made using a combination of 6 different characters. The sequences should be of reasonable length. Use any Sequence Mining algorithm to extract frequent sequences having a minimum support =2. **The dataset should be your own and not chosen from any example given anywhere in the books.**

Q.3. A credit card company receives thousands of applications for issue of new cards. The application contains several attributes like Name, Age and many more. The problem is to categorize the applications into two classes **"Good Credit"** or **"Bad credit"** based on past experience of the company. Think of a suitable, manageable and sufficient dataset for the above problem and using a decision tree classifier find the root of the tree.