JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATION- 2018

B.Tech. (IT), 6th Semester

COURSE CODE: 10B22CI622

COURSE NAME: Data Mining

COURSE CREDITS: 4

MAX. MARKS: 25

MAX. TIME: 1.5 HR

Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means.

1. Association rule mining often generated a large number of rules. Discuss effective methods that can be used to reduce the number of rules generated while still preserving most of the interesting rules. **[2 Marks]**

2. How does the tree pruning work? Briefly explain the two common approaches to tree pruning. **[3 Marks]**

3. Consider the Transactional Data for an all electronics branch as shown below:

TID	List of Items_IDs
T100	I_1, I_2, I_5
T200	I_2, I_4
T300	I_2, I_3
T400	I_1, I_2, I_4
T500	I_1, I_2, I_3, I_5
T600	I_1, I_2, I_3

Mine the frequent itemsets using vertical data format and consider the minimum support as two for this problem. **[4 Marks]**

4. Consider the class labeled tuples from the AllElectronics customer database as follows;

RID	Age	Income	Student	Credit-Rating	Class: buys_computer
1	Youth	High	No	fair	No
2	Youth	High	No	excellent	No
3	Middle_Aged	High	No	fair	Yes
4	Senior	Medium	No	fair	Yes
5	Senior	Low	Yes	fair	Yes
6	Senior	Low	Yes	excellent	No
7	Middle_Aged	Low	Yes	excellent	Yes
8	Youth	Medium	No	fair	No
9	Youth	Low	Yes	fair	Yes
10	Senior	Medium	Yes	fair	Yes
11	Youth	Medium	Yes	excellent	Yes
12	Middle_Aged	Medium	No	excellent	Yes
13	Middle_Aged	High	Yes	fair	Yes
14	Senior	Medium	No	excellent	No

Apply the information gain method and identify the first splitting attribute for the same.

[8 Marks]

4. Consider a database has five transactions. Let *min sup* = 60% and *min conf* = 80%.

TID	Items_bought
T100	{ M,O,O, N,K,E,Y}
T200	{D,O,N,K,E,Y}
T300	{M,A,K,E,E}
T400	{M,U,C,K,Y,Y}
T500	{C,O,O,K,I,E,E}

Find all frequent itemsets using Apriori and FP-growth, respectively. Compare the efficiency of the two mining processes.

[8 Marks]