ST. JOSEPH'S COLLEGE (AUTONOMOUS), BANGALORE-27
M.SC. BIG DATA ANALYTICS - II SEMESTER
SEMESTER EXAMINATION: APRIL 2020
BDA 2418: Data Mining

Time: 2 ½ hours   **Max marks:70**

This Question Paper Contains Two Printed Pages and One Part

Answer any seven questions

1. Define the Data Mining and illustrate the steps involved in the process of knowledge discovery from data.
2. Illustrate the working of a decision tree in a credit risk system.
3. Explain briefly the divisive hierarchical clustering algorithm with an example.
4. Explain the working apriori algorithm by solving the numerical on Apriori algorithm.

TID	List of Items
T1	I1, I2, I5, I6
T2	I2, I4, I5
T3	I2, I3, I5
T4	I1, I2, I4,I5
T5	I1, I3, I4
T6	I2, I3, I5
T7	I1, I3
T8	I1, I2 ,I3, I5
T9	I1, I2, I3

5. Define the following:
   1. Various types of attributes in data mining
   2. Proximity measures for binary data
   3. Cosine similarity
   4. Jaccards similarity
6. Explain the working of the Naïve Bayes classifier as a sentiment analyzer.
7. Write a note on data reduction, data cleaning, data integration and data transformation.
8. Explain the working of the multi-layered feed-forward neural network with the various activation functions.
9. Fit the K-means clustering algorithm for the given numerical, where k =2.

a	b	C
5.1	3.5	1.4
4.9	3	1.4
7	3.2	4.7
6.4	3.2	4.5
6.3	3.3	6
5.8	2.7	5.1
4.9	2.4	3.3