SUBJECT CODE NO:- E-8060
FACULTY OF ENGINEERING AND TECHNOLOGY
M.E. (Comp.Sci. & Engg.) Examination Nov/Dec 2017
Data Mining & Big Data
(Revised)

[Time: Three Hours]   **[Max.Marks:80]**

Please check whether you have got the right question paper.
N.B   i) Solve any two questions from each section.
                 ii) Figures right indicates full marks.
                 iii) Assume Suitable data if necessary

Section A

Q.1   a) Find frequent item sets for following transactions using Apriori Algorithm.   10
            Minimum support count = 2

            TID   List of item
             1   11,12,15
             2   12,13
             3   12,14
             4   11,12,13
             5   11,14
             6   12,14
             7   11,14
             8   11,12,14,15
             9   11,12,14

        b) What is constraint based association mining? What are the different type of constraints?   10

Q.2   a) Cluster following points in three clusters. Take initially A1, B1, C1 as a center points. Use K-   10
            means algorithms to show only the three cluster centers after the final round of execution.
            A1(2,10) A2(2,5) A3(8,4) B1(5,8) B2 (7,5) B3 (6,4) C1(1,2) C2 (4,9) (Use Euclidean
            distance).

        b) With an example explain how hierarchical clustering works   10
            1) Single linkage
            2) Complete Linkage

Q.3   a) What is the concept of page rank by which the page popularity is captured on web? Explain   10
            with the page rank algorithm.

        b) What is social network analysis (SNA)? How graph techniques are used for SNA?   10

2017

14414640D234007D1BC18F2CD343AD03

Section B

Q.4 a) How to improve Data Access in Hadoop using HBase, Sqoop, and Flume?   10

     b) A cloud uses 2000 nodes for data processing and has a processing capacity of 50GB per hour. 10
         Considering the charges as 0.4USD per node per hour calculate the total cost and time
         required for processing one zettabyte of data on this cloud.

Q.5 a) What is object base data products ? How will you apply drive train approach for marketing   10
         purpose?

     b) Explain "What it takes to build great machine learning products"?   10

Q.6 a) Explain how 'Data Personalization' can be used for 'discrimination'. Justify with example.   10

     b) Explain in brief how the application of 'spreadsheet' got enriched to 'dashboard'. What is the 10
         role of privacy in design?

2017