Total No. of Printed Pages:1

SUBJECT CODE NO:- P-8216

FACULTY OF ENGINEERING AND TECHNOLOGY

M.E.(Comp. Sci. & Engg.) Examination MAY/JUNE-2016

Data Mining & Big Data

(Revised)

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

"Please check whether you have got the right question paper."

N.B 1) Solve any two questions from each section.

2) Assume suitable data if necessary and state it clearly.

Section A

Q.1 a) Find the frequent item sets for following transactions using Apriori Algorithm. Minimum support count=2. 12

TID	List if 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 types of constraints? 08

Q.2 a) The daily expenditures on food (X1) and Clothing (X2) of five persons are shown below: given data is – (a,1,1), (b,1.5,1.5), (c,5,5), (d,3,4), (e ,4,4), (f,3,3.5) plot observations on a scatter diagram. How many clusters are formed visually? Calculate how the clusters will be formed using Single linkage method and complete linkage method. Use the Euclidean distance measure. Draw the dendogram in each case. 12

b) What are outliers? Describe any two methods of outlier detection. 08

Q.3 a) What is a closely related model of a Social Network? Give any two examples of Scale-free networks. 10

b) What is k-means algorithm? What are its limitations? 10

Section B

Q.4 a) What are the characteristics of Big Data? What are its challenges? 05

b) Distinguish clearly between- map reduce and Parallel DBMS technology. 05

c) Explain in brief the process of how the application of 'Spreadsheet' got enriched to 'dashboard'. 10

Q.5 a) Describe Apache Hadoop Architecture. How data access can be improve with HBase, Sqoop, and Flume. 10

b) How will you apply drive train approach for recommender system? Design and explain the steps. 10

Q.6 a) Explain the use of Big Data in Digital Publishing Domain. 10

b) Describe in brief: 10

i) The Dark side of Big data

ii) Use of Machine Learning in data products.