

Total No. of Printed Pages:2

**SUBJECT CODE NO:- H-1660**  
**FACULTY OF SCIENCE AND TECHNOLOGY**  
**M.E. (Comp. Sci. & Engg.)**  
**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. Figures right indicates full marks
3. Assume suitable data if necessary

**Section A**

- Q.1 a) What are outliers? Describe any two methods of outlier detection 10  
 b) What is a closely related model of a social Network? Give any two examples of scale- free networks 10
- Q.2 a) What is constraint based association mining? What are the types of constraints? 10  
 b) What is correlation analysis? The following contingency table summarizes transaction data. Find the correlation between A and I using lift and cosine measures. 10

	A	B
I	400	350
$\bar{I}$	300	50

- Q.3 a) What is spatial mining? Describe any one application which is based on it 10  
 b) 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 diagram in each case. 10

**Section B**

- Q.4 a) What are the characteristics of Big Data? What are its challenges? 10  
 b) Distinguish clearly between- Mapreduce and Parallel DBMS technology. 10
- Q.5 a) Explain in brief the process of how the application of ‘Spreadsheet’ got enriched to ‘dashboard’. 10  
 b) Describe Apache Hadoop Architecture. How data access can be improved with HBase, Sqoop, and Flume. 10

- Q.6 a) Describe Apache Hadoop Architecture. What is the use Oozie, Zookeeper and Mahout tools? 10
- b) How will you apply drivetrain approach for recommender system? Design and explain the steps. 10