

Government College of Engineering Aurangabad
(An Autonomous Institute of Government of Maharashtra)

BE (IT) FT Examination
End Semester Examination November 2016

IT442: DATA MINING

Time: Three Hours

11 NOV 2016

Max. Marks: 60

“Verify the Course Code & check whether you have got the correct question paper”

N.B:-

- 1. All questions are compulsory*
- 2. Figures to the right indicate full marks*
- 3. Assume suitable data if necessary & state it clearly*
- 4. Use of non-programmable calculator is allowed*

-
- Q.1 Attempt any Two** (12)
- i) List and explain steps in the process of Knowledge discovery
 - ii) Discuss three major issues in Data Mining
 - iii) Briefly describe any one of the following advanced database system and discuss an application of the same : Spatial Databases or Sequence Databases
- Q.2 Attempt any Two** (12)
- i) Why data reduction is necessary? Explain discretization and concept hierarchy with respect to data reduction
 - ii) Suppose that the data for analysis includes the attribute age. Age values are 13,15,16,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35,36,40,45,46,52,70 Use min max[range 0.0-0.1], z score normalization (std. dev =12.94 years) to transform value 35 for age
 - iii) Write and explain Apriori Algorithm with suitable example.
- Q.3 Attempt any Two** (12)
- i) Illustrate Decision tree classification with suitable example
 - ii) What is clustering? List and explain any three requirements of clustering.
 - iii) Write a Note on: Hierarchical clustering
- Q.4 Attempt any Two** (12)
- i) Draw and explain MapReduce data flow with multiple reduce task
 - ii) How does the Namenode choose datanodes to store replicas on?
 - iii) Explain ,how client read data from HDFS
- Q.5 Write a note on any two** (12)
- i) Limitations of HAR files
 - ii) File permission in HDFS
 - iii) Characteristics of HDFS