

Code No: **R41053****R10****Set No. 1****IV B.Tech I Semester Supplementary Examinations, February/March - 2018****DATA WARE HOUSING AND DATA MINING****(Common to Computer Science and Engineering and Information Technology)****Time: 3 hours****Max. Marks: 75**

Answer any FIVE Questions
All Questions carry equal marks

- 1 a) What meant by data quality? Mention different techniques to measures data quality [8]
b) Why preprocessing is required? explain [7]
- 2 a) Differentiate between Jaccard coefficient and extended jaccard coefficient [8]
b) How to measure dissimilarities between data objects. [7]
- 3 a) Explain different data cube operations [8]
b) What is Data warehousing Modeling? Why it is necessary. [7]
- 4 a) What are the different measures for selecting the best split [8]
b) Explain about overfitting and underfitting [7]
- 5 a) What is meant by model representation? Explain about model building [8]
b) Write the characteristics of Bayesian Belief Networks [7]
- 6 a) What is bayes theorem? Explain classification with help of bayes theorem. [8]
b) Explain the characteristics of Association analysis [7]
- 7 a) What does clustering mean? Explain different types of clustering [8]
b) Explain strength and weakness of K-means [7]
- 8 Write short notes on
a) Center based approach
b) Rule generation
c) Optimization problem [15]