

Exam.Code:1007
Sub. Code: 7717

1078
M.E. (Information Technology)
Third Semester
MEIT-315: Advanced Data Mining

Time allowed: 3 Hours

Max. Marks: 50

NOTE: Attempt five questions in all, including Question No. 1 which is compulsory and selecting two questions from each Unit.

x-x-x

I. Attempt the following:-

- a) What is entity identification problem?
- b) Name any two descriptive measures for data skewness.
- c) What is meant by spatio-temporal data mining?
- d) Differentiate between categorical, binary and continuous data.
- e) Define any two distance measures used in clustering.
- f) List any two features of parallel RDBMS.
- g) Distinguish between classification and prediction.
- h) State me main underlying assumption in using Naive Bayes method for classification.
- i) List the various BI infrastructure components.
- j) What are objective measures of pattern interestingness? (10x1)

UNIT – I

- II. Explain the various methods of data cleaning. Why is it needed? (10)
- III.
 - a) Discuss the multidimensional schema architecture of a data warehouse.
 - b) Distinguish between dependent data marts, enterprise warehouse and virtual warehouse. (5,5)
- IV. What is attribute relevance analysis? Explain the entire process including computation of relevant information theoretic measure. (10)

P.T.O.

(2)

UNIT - II

- V. Compare and contrast methods for mining association rules with and without candidate generation. (10)
- VI. Explain what are bagging and boosting classifiers. Describe any one such classifier in detail. (10)
- VII. Write short notes on:-
a) Embedded data mining
b) Applications of data mining (5,5)

x-x-x