R13

Code No: 126EW

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech III Year II Semester Examinations, December - 2017
DATA WAREHOUSING AND DATA MINING
(Information Technology)

Time: 3 hours **Max. Marks: 75**

Note: This question paper contains two parts A and B.
Part A is compulsory which carries 25 marks. Answer all questions in Part A. Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

PART - A

(25 Marks)

1.a)	Write brief notes on OLAP.	[2]
b)	State Data mining primitives.	[3]
c)	Define Euclidean and Jaccard similarity measure.	[2]
d)	Define KDD.	[3]
e)	Define the two step process of association rule mining.	[2]
f)	State data mining tasks.	[3]
g)	State the confidence and support measures with formulas.	[2]
h)	Explain pre-pruning and post -pruning decision tree.	[3]
i)	Mention types of clustering techniques.	[2]
j)	Define outlier detection. Explain database scan.	[3]

PART - B

(50 Marks)

2.a) What is a Data Warehouse? Explain three types of schemas that are used for modeling data- warehouse with examples.
b) Explain the Data cube computation techniques. [5+5]

OR

3.a) With a neat sketch explain the architecture of Online analytical Mining.
b) Discuss the OLAP operations used in multi dimensional data model. [5+5]

4.a) What is the need for Data preprocessing?
b) Discuss briefly various forms of Data -Preprocessing. [5+5]

OR

5.a) Explain data mining as a step in process in knowledge discovery.
b) Differentiate operational database systems and data warehousing. [5+5]

6.a) Explain a method that mines the complete set of frequent item sets without candidate generation for the table below. Explain the procedure in detail with minimum support = 3.

TID	Items
100	F, A, C, D, G, I, M, P
200	A, B, C, F, L, M, O
300	B, F, H, J, O, W
400	B, C, K, S, P
500	A, F, C, E, L, P, M, N

b) Mine the possible association rules with the frequent item -sets derived from above example. [5+5]

OR

7.a) Define Association Rule mining. Explain Apriori algorithm with suitable illustration.

b) Explain constraint based rule mining. [5+5]

8.a) What is Classification? With an example explain how Support Vector machines can be used for classification, evaluating the accuracy of a classifier or a predictor? Justify this statement with suitable illustrations.

b) Explain Decision tree induction algorithm. [5+5]

OR

9.a) Explain Naïve - bayes algorithm.

b) Discuss on classification by back propagation. [5+5]

10.a) Explain PAM algorithms.

b) Differentiate density based and model based clustering Techniques. [5+5]

OR

11.a) Explain the approaches in short Compare and contrast any two outlier analysis Approaches.

b) State key issues in hierarchical clustering. [5+5]

---ooOoo---