JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech IV Year I Semester Examinations, November/December – 2017
DATA WAREHOUSING AND DATA MINING
(Computer Science and Engineering)
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

1.a) Define data warehouse. [2]
b) List the Data warehouse Characteristics. [3]
c) How can you go about filling in the missing values for this attribute? [2]
d) Why is the word data mining a misnomer? [3]
e) Give a note on Closed Frequent Item Set. [2]
f) Write the FP-graph algorithm. [3]
g) How prediction is different from classification? [2]
h) What is rule classification? [3]
i) Give a note on k means algorithm. [2]
j) List the Key Issues in Hierarchical Clustering. [3]

PART – B (50 Marks

2.a) Make a comparisons between the MOLAP and HOLAP

. b) Discuss the star and snowflake schema in detail with suitable example. [5+5]

OR

3.a) Write the difference between designing a data warehouse and an OLAP cube.
b) Give a brief note on ROLAP. [5+5]

4. Explain concept hierarchy generation for the nominal data. [10]

OR

5.a) Describe the Feature Subset Selection. b) Illustrate the Data Transformation by Normalization. [5+5]

6. Make a comparison of Apriori and ECLAT algorithms for frequent item set mining in transactional databases.
Apply these algorithms to the following data:

TID	LIST OF ITEMS
1	Bread, Milk, Sugar, TeaPowder, Cheese, Tomato
2	Onion, Tomato, Chillies, Sugar, Milk
3	Milk, Cake, Biscuits, Cheese, Onion
4	Chillies, Potato, Milk, Cake, Sugar, Bread
5	Bread, Jam, Mik, Butter, Chilles
6	Butter, Cheese, Paneer, Curd, Milk, Biscuits
7	Onion, Paneer, Chilies, Garlic, Milk
8	Bread, Jam, Cake, Biscuits, Tomato [10]

OR

7. Briefly explain the Partition Algorithms. [10]

8. Discuss K- Nearest neighbor classification-Algorithm and Characteristics. [10]

OR

9.How does the Naïve Bayesian classification works? Explain in detail. [10]

10.a) Give a brief note on PAM Algorithm.
b) What is the drawback of k-means algorithm? How can we modify the algorithm to diminish that problem? [5+5]

OR

11.What are the different clustering methods? Explain in detail. [10]

Code No: 117CD

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

B. Tech IV Year I Semester Examinations, April/May - 2018

DATA WAREHOUSING AND DATA MINING

(Computer Science and Engineering)

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

1.a) List out the operations of OLAP. [2]
b) What is fact table? Write its uses. [3]
c) Define discretization. [2]
d) What is predictive mining? Explain it briefly [3]
e) Write the purpose of Apriori algorithm. [2]
f) Define support and confidence measure. [3]
g) What is boosting? [2]
h) Define decision tree. [3]
i) Write the strengths of hierarchical clustering. [2]
j) Compare agglomerative and divisive methods. [3]

PART-B

(50 Marks)

2. a) With a neat sketch, Explain three tier architecture of data ware housing.
   b) Explain various data warehouse models. [5+5]

OR

3. Write a note on
   a) Relational OLAP
   b) Multi dimensional OLAP. [5+5]

4.a) Discuss in detail about the steps of knowledge discovery?
b) Write a note on subset selection in attributes for data reduction

OR

5.a) Explain various data mining tasks.
b) Discuss briefly about data cleaning techniques. [5+5]

6.a) Write FP- growth algorithm.
b) Explain how association rules are generated from frequent item sets. [5+5]

OR

7.a) Explain the procedure to mining closed frequent data item sets.

b) Explain, how can you improve the performance of Apriori algorithm.

8.a) What is Bayesian belief network? Explain in detail.
b) Write a note attribute selection measures. [5+5]

OR

9.a) Write k-nearest neighbor classification algorithm and its characteristics.
b) Write decision tree induction algorithm. [5+5]

10.a) What is outlier detection? Explain distance based outlier detection.
b) Write partitioning around mediods algorithm. [5+5]

OR

11.a) Write K-means clustering algorithm.
b) Write the key issue in hierarchical clustering algorithm

15. Advanced topics beyond the syllabus

- Active learning
- Multi-label learning
- Graph mining
- Link prediction
- Data mining in bioinformatics
- Social media analytics
- Privacy-aware data mining