Code: 15A05602 **R15**

B.Tech III Year I Semester (R15) Supplementary Examinations October 2020

DATA WAREHOUSING & MINING

(Information Technology)

Time: 3 hours Max. Marks: 70

PART – A

(Compulsory Question)

*****

1 Answer the following: (10 X 02 = 20 Marks)

(a) Define data mining.
(b) What is algebraic measure?
(c) What is the need for separate data warehouse?
(d) What are the components of data warehouse architecture?
(e) What is lift?
(f) What are strong association rules?
(g) What are lazy learners?
(h) What is the best split point?
(i) Write the types of data in clustering?
(j) What is stream data?

PART – B

(Answer all five units, 5 X 10 = 50 Marks)

UNIT – I

2 (a) Discuss integration of data mining with database or data warehouse system.
(b) Explain the primitives of data mining task.

OR

3 (a) Write about major issues in data mining system.
(b) Explain classification of data mining system.

UNIT – II

4 (a) Explain in detail about snowflake schema.
(b) Explain about Fact Constellation schema.

OR

5 (a) Explain the 3 tier architecture of data warehouse system.
(b) What is meta data? Explain different types of OLAP servers.

UNIT – III

6 (a) Explain frequent itemset using candidate generation.
(b) Explain the technique for improving the efficiency of Apriori algorithm.

OR

7 (a) Explain about tree pruning algorithm.
(b) Explain about support vector machine.

UNIT – IV

8 (a) Explain the requirements of clustering.
(b) Explain about binary variables in clustering.

OR

9 (a) Explain about hierarchical clustering.
(b) Explain about DBSCAN algorithm.

UNIT – V

10 (a) Discuss about mining sequence patterns in biological data.
(b) Explain periodicity analysis for time related sequence data.

OR

11 (a) Explain in detail about text mining.
(b) Explain about social network analysis.

*****

Code: 15A05602 **R15**

B.Tech III Year I Semester (R15) Supplementary Examinations June/July 2019

DATA WAREHOUSING & MINING

(Information Technology)

Time: 3 hours Max. Marks: 70

PART – A

(Compulsory Question)

*****

1 Answer the following: (10 X 02 = 20 Marks)

(a) What is the need for pre-processing the data?
(b) What are the types of concept hierarchies?
(c) List out the functions of OLAP servers in the data warehouse architecture.
(d) Define Slice and Dice operation.
(e) What is over fitting and what can you do to prevent it?
(f) Define the concept of classification.
(g) List basic requirements of cluster analysis.
(h) What is outlier?
(i) What are the challenges in mining the WWW?
(j) Define time series analysis.

PART – B

(Answer all five units, 5 X 10 = 50 Marks)

UNIT – I

2 What is data cleaning? Describe in detail the different methods for data cleaning.

OR

3 How data mining system can be classified? Describe the different data mining functionalities with example.

UNIT – II

4 With relevant examples discuss multidimensional online analytical processing and multi relational online analytical processing.

OR

5 Illustrate three-tier data warehouse architecture and the different schemas for multidimensional database.

UNIT – III

6 A database has five transactions. Let minimum support = 60% and minimum confidence = 80%.

Transaction	Items
T10	M, O, N, K, E, Y
T20	D, O, N, K, E, Y
T30	M, A, K, E
T40	M, U, C, K, Y
T50	C, O, O, K, I, E

Find all frequent item sets using Apriori and FP-growth, respectively.

OR

7 Explain in detail classification method using decision tree induction with an example.

Contd. in page 2

UNIT – IV

8 Design statistical approaches in outlier detection with neat design and examples.

OR

9 Explain the following clustering methods in detail:

(a) BIRCH.

(b) CURE.

UNIT – V

10 What is spatial database? Explain the methods of mining spatial databases.

OR

11 Discuss in detail about mining time series data and mining data streams.

*****

Code: 15A05602 **R15**

B.Tech III Year II Semester (R15) Regular & Supplementary Examinations May/June 2019

DATA WAREHOUSING & MINING

(Computer Science & Engineering)

Time: 3 hours Max. Marks: 70

PART – A

(Compulsory Question)

*****

1 Answer the following: (10 X 02 = 20 Marks)

(a) Define data characterization.
(b) What information can be uncovered by mining text data?
(c) Why do we need to have a separate data warehouse?
(d) What is curse of dimensionality?
(e) What is frequent itemset mining?
(f) Define prediction.
(g) What are binary variables?
(h) How is dissimilarity computed between objects described by ordinal variables?
(i) What is time series data?
(j) Define the terms minimum support and minimum confidence.

PART – B

(Answer all five units, 5 X 10 = 50 Marks)

UNIT – I

2 What are the types of data on which mining can be performed? Explain with examples.

OR

3 What is the need for data cleaning? Explain the steps involved in the process of data cleaning.

UNIT – II

4 Explain about multidimensional data model with an example indicating the operations performed.

OR

5 Write the basic attribute oriented induction algorithm.

UNIT – III

6 Describe FP growth algorithm with an example.

OR

7 Explain Naïve Bayes classification with an example.

UNIT – IV

8 Explain the clustering process of categorical attributes using ROCK algorithm.

OR

9 How are clusters identified using DBSCAN algorithm?

UNIT – V

10 How are frequent patterns mined in data streams?

OR

11 Write the Apriori-based approach for frequent substructure mining.

*****

Code: 15A05602 **R15**

B.Tech III Year II Semester (R15) Regular & Supplementary Examinations October/November 2020

DATA WAREHOUSING & MINING

(Computer Science and Engineering)

Time: 3 hours Max. Marks: 70

PART – A

(Compulsory Question)

*****

1 Answer the following: (10 X 02 = 20 Marks)

(a) What is a data mart?
(b) What is the need for preprocessing?
(c) What is a fact table?
(d) Mention four OLAP operations.
(e) What is maximal frequent item set?
(f) What is classification?
(g) What is an outlier?
(h) List the demerits of k-means algorithm.
(i) Give applications for time series mining.
(j) Why sequence patterns are important in biological data mining?

PART – B

(Answer all five units, 5 X 10 = 50 Marks)

UNIT – I

2 (a) Explain the major issues in data mining.
(b) Explain the data mining system architecture with the help of a diagram.

OR

3 (a) Discuss the data preprocessing techniques giving examples for each.
(b) Give various functionalities of data mining systems.

UNIT – II

4 (a) What are the various components of data warehouse? Explain their functionality.
(b) Differentiate between the features of data warehouse and normal databases.

OR

5 (a) Differentiate between OLAP and OLTP operations.
(b) Explain attribute oriented induction with example.

UNIT – III

6 (a) Explain the Apriori algorithm for mining frequent item sets.
(b) Discuss various kinds of association rules.

OR

7 (a) Explain back propagation algorithm for classification.
(b) Explain decision tree classification with the help of a neat diagram.

UNIT – IV

8 (a) Write short notes on grid based clustering methods.
(b) Explain any hierarchical clustering method with a diagram.

OR

9 (a) Give the merits and demerits of k-means algorithm with a neat diagram.
(b) Categorize various clustering methods mentioning a technique for each category.

UNIT – V

10 (a) Explain sequence mining in detail.
(b) Explain the importance of web log mining. Give examples.

OR

11 (a) Explain text mining and mention its applications.
(b) Explain social network analysis and mention its significance.

*****