

Code No: XXXXX

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD**B. Tech IV Year I Semester****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 are 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.

MODEL PAPER - 1**PART-A(Answer all the Questions)**

- 1.a) Write the differences between data warehousing and data mining.(3M)
 - b) Define multi dimensional data mining.(2M)
 - c) State the various views of data warehouse design?(3M)
 - d) Name the steps involved in data mining?(2M)
 - e) Name the pruning strategies in mining closed frequent item sets?(3M)
 - f) List the applications of pattern mining?.(2M)
 - g) Differentiate the supervised and unsupervised learning(2M)
 - h) Write short notes on the back propagation algorithm? (3M)
State the applications of clustering.(3M)
 - j) Explain briefly about the grid based method (2M)
2. Write the differences between operational databases and data warehousing? (10M)
OR
3. Explain in detail about the evolution of database technology.(10M)
4. Discuss briefly about multi dimensional data models?(10M)
OR
5. State and explain the methods used for efficient data cube computation(6M)
6. Discuss the FP-Growth algorithm with an example.(10M)
OR
7. Explain how to mine the multidimensional association rules from relational databases and data warehouses?
8. Discuss in detail about the decision tree induction algorithm.(10M)
OR
9. Write in detail about the k-nearest neighbor classifier and case-based reasoning?
10. Define and explain the two hierarchical clustering methods: BIRCH and CHAMELON.(10M)
OR
11. Explain about
 - a) Statistical-based outlier detection.(5M)
 - b) Distance-based outlier detection. (5M)

Code No: XXXXX

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD**B. Tech IV Year I Semester****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 are 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.

MODEL PAPER - 2**PART-A(Answer all the Questions)**

- 1.a) Write a short notes about the issues in data mining.(3M)
 - b) Define characterization and discrimination.(2M)
 - c) Explain in short about the virtual data warehouse.(3M)
 - d) Define concept hierarchy. Explain the types of concept hierarchies.(2M)
 - e) Define closed item set and maximal frequent item set.(3M)
 - f) What is an association? Write a short notes about association rule mining.(2M)
 - g) Define regression analysis..(2M)
 - h) Write a short notes about the attribute selection measures.(3M)
 - i)Write neatly about the data types used in cluster analysis.(3M)
 - j) Write about the density based clustering.(2M)

PART-B

2. Define data mining and explain in detail about the data warehouse architecture with a neat diagram.(10M)

OR

3. What are the primitives that specify the data mining task? Explain in detail about the data smoothing techniques.(10M)

4. Write neatly about different schemas used in multi dimensional data mining with an example for each.(10M)

OR

5. Define ROLAP, MOLAP, and HOLAP. Explain in detail about the efficient methods of data cube computation. (10M)

6. Write and explain the APRIORI algorithm with an example.

OR

7. Write a short notes about the interestingness measures. Discuss about constraint based association rule mining.

8. What measures are used to find best split in Decision Tree Induction algorithm? How Can we improve the scalability in Decision Tree Induction algorithm?(10M)

OR

9. Describe the working procedures of simple Bayesian classifier. Discuss the Back propagation algorithm.

10. Explain in detail about the categories of major clustering methods.(10M)

OR

11. What is an outlier? Explain about (10M)

- a) Distance-based outlier detection
- b) Statistical based outlier detection
- c) Density-based outlier detection.

Code No: XXXXX**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD****B. Tech IV Year I Semester****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.

MODEL PAPER - 3**PART-A(Answer all the Questions)**

- 1) a)What are the advantages of data warehouse?(3M)
- b)Define OLAP. (2M)
- c)List the major issues in data mining. (3M)
- d)List the reasons for using data mining? (2M)
- e)List the reasons for using data mining? (3M)
- f)Define FP-tree? (2M)
- g)What is Hunt's Algorithm?(2M)
- h)What is holdout technique? (3M)
- i)List the requirements of clustering.(3M)
- j) Write in brief about index based algorithms (2M)

PART-B

2.Explain the steps for designing and constructing data warehouse? (10M)

OR

3.What is data mining? List and describing the motivating challenges of data mining.(10M)

4.Discuss in brief about fact table.(10M)

OR

5.Explain in detail about transformation? (10M)

6. Explain in detail the construction of FP tree?(10M)

OR

7.Discuss in brief about,

- a)Maximum frequent item set.(5M)
- b)Closed frequent item set.(5M)

8 Write notes on evaluating the performance of a classifier?.(10M)

OR

9.How a Naive Bays classifier works? Explain with an example?(10M)

10.What is cluster analysis? Explain with suitable Example.(10M)

OR

11.What are different types of hierarchical methods? Explain?(10M)

Code No: XXXXX

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD**B. Tech IV Year I Semester****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 are 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.

MODEL PAPER - 4**PART-A(Answer all the Questions)**

- 1
 - a) Compare OLAP system versus statistical databases?(3M)
 - b) What are the differences between fact and dimension tables?(2M)
 - c) Write a brief note on data discretization? (3M)
 - d) Discuss briefly about similarities between data objects? (2M)
 - e) Describe Brute-forces method to generate candidate's item sets? (3M)
 - f) Discuss briefly the monotonicity property. (2M)
 - g) List the advantages of information gain?(2M)
 - h) List the weaknesses of k-means? (3M)
 - i) Write short note on density based outlier detection?(3M)
 - j) Write short notes on density based outlier detection.?(2M)

PART-B

2. Draw and explain the three tier data warehouse architecture? (10M)

OR

3. Illustrate and explain the OLAP architecture.(10M)

4. Explain different data pre- processing techniques.(10M)

OR

5. Explain with example the Jaccard coefficient? (10M)

6. With an example, explain the Fp-growth algorithm?.(10M)

OR

7. Explain in detail the candidate generation procedures (10M)

8. Discuss various types of classification techniques.(10M)

OR

9. Write an algorithm for decision tree induction? (10M)

10. What are the issue K-means?(10M)

OR

11. Explain briefly about statistical distribution based outlier detection. (10M)

Code No: XXXXX

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD

B. Tech IV Year I Semester

**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 are 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.

MODEL PAPER - 5

PART-A(Answer all the Questions)

1.
 - a) Discuss the characteristics of fact table(3M)
 - b)What are the tools used for designing data warehouse?(2M)
 - c)What is data cleaning? (3M)
 - d)Write the algorithm for discrete wavelet transform (DWT)? (2M)
 - e)Define support and confidence? (3M)
 - f)What are the drawbacks of FP growth algorithm ? (2M)
 - g)What is meant by classification? What are the applications of classification model? (2M)
 - h)List the advantage of Bayesian classification? (3M)
 - i)Make a comparison of complete and partial clusters?(3M)
 - j) Discuss the times and space complexity of K-means.?(2M)

PART-B

2.Explain in detail about ETL? (10M)

OR

3.Write a short note on OLAP cube.(10M)

4.Explain the process of knowledge discovery in database.(10M)

OR

5.What is data cleaning? What are the different techniques for handling missing values? (10M)

6. With an example, explain the frequent item set generation in the Apriori algorithm.(10M)

OR

7.Explain the partition algorithm with an example. (10M)

8. Explain briefly the test conditions for different types of attributes?(10M)

OR

9.What is the role of nearest neighbor classifier? Explain it briefly? (10M)

10. Write a short note on partitioning clustering?(10M)

OR

11.Explain agglomerative hierarchical clustering. (10M)

Code No: XXXXX

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD**B. Tech IV Year I Semester****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 are 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.

MODEL PAPER - 6**PART-A(Answer all the Questions)****Answer all of the following.**

1. a) Define KDD process. (2M)
b) What is Data Discretization? (3M)
c) Explain Minimum Support and Confidence Threshold (2M)
d) Write the difference between OLAP and OLAM. (3M)
e) What are Ensemble methods (2M)
f) Explain the OLAP operations with examples. (3M)
g) Write the short note on Frequent pattern sequences. (2M)
h) Explain Graph Mining. (3M)
i) Explain metadata repository. (2M)
j) What is Classifier? Write the Bayesian Classification technique. (3M)
- 2.Explain Data Warehouse Implementation steps. (10M)
OR
- 3.Explain Attribute Oriented Induction Technique.
- 4.Explain Data Mining Functionalities. (10M)
OR
- 5.Explain Data Mining Task primitives.
- 6.Explain Mining Frequent Patterns using APRIORI. (10M)
OR
- 7.Explain Mining Frequent Patterns using FP-Growth.
- 8.Write a prototype for data mining application for Insurance DWH. (10M)
OR
- 9.Write about mining the web link structures to identify authoritative web pages.
- 10.Explain about Spatial data cube construction and spatial OLAP. (10M)
OR
11. Explain the Class Composition Hierarchies.

Code No: R15A0526

MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY
 (Autonomous Institution – UGC, Govt. of India)
IV B. Tech I Semester Supplementary Examinations, May 2019
Data Warehousing and Data Mining
 (CSE)

Roll No										
----------------	--	--	--	--	--	--	--	--	--	--

Time: 3 hours**Max. Marks: 75****Note:** This question paper contains two parts A and B

Part A is compulsory which carries 25 marks and Answer all questions.

Part B Consists of 5 SECTIONS (One SECTION for each UNIT). Answer FIVE Questions, Choosing ONE Question from each SECTION and each Question carries 10 marks.

PART-A (25 Marks)

- | | | | |
|-----|---|---|------|
| 1). | a | What is Fact Constellation? | [2M] |
| | b | Discuss about the Fact Table | [3M] |
| | c | List out the Major issues in Data Mining | [2M] |
| | d | What is the Need for Preprocessing the Data | [3M] |
| | e | Define The Partition Algorithms | [2M] |
| | f | Discuss about the FP-Growth Algorithms | [3M] |
| | g | What is K- Nearest neighbor classification-Algorithm? | [2M] |
| | h | Explain about the Accuracy and Error measures | [3M] |
| | i | Give the different types of data in cluster analysis. | [2M] |
| | j | Discuss about the Grid-Based Methods | [3M] |

PART-B (50 MARKS)**SECTION-I**

- | | | | |
|---|----|--|-------|
| 2 | a) | Differentiate between Fully Addictive and Semi-Addictive (5 M) | [10M] |
| | b) | Explain about the Non Addictive Measures and Fact-Less-Facts (5 M) | |

OR

- | | | | |
|---|----|---|-------|
| 3 | a) | Discuss about the Dimension Table Characteristics (5 M) | [10M] |
| | b) | Explain about the OLAP Cube and OLAP Operations (5 M) | |

SECTION-II

4 Write Short notes on Data Cleaning, Data Integration & Transformation, [10M]

OR

5 a) Explain about the Data Reduction (5 M) [10M]

b) Discuss about the Discretization and Concept Hierarchy Generation (5 M)

SECTION-III

6 Explain about the APRIORI Principle of Frequent Item Set [10M]

OR

7 Differentiate between Maximal Frequent Item Set and Closed Frequent Item Set [10M]

SECTION-IV

8 Explain about the Naive Baye's classifier with an example. [10M]

OR

9 Describe the Ensemble Methods [10M]

SECTION-V

10 Discuss about the Model based Clustering Methods. [10M]

OR

11 Explain about the Outlier Analysis [10M]
