



Reg. No. :

--	--	--	--	--	--	--	--	--	--	--	--

Question Paper Code : X 60353

B.E./B.Tech. DEGREE EXAMINATIONS, NOVEMBER/DECEMBER 2020
Seventh Semester
Computer Science and Engineering
CS 2032/CS 701/10144 CSE 32 – DATA WAREHOUSING AND DATA MINING
(Common to Sixth Semester Information Technology)
(Regulations 2008/2010)
(Common to PTCS 2032 – Data Warehousing and Data Mining for
B.E. (Part-Time) Sixth Semester – CSE – Regulations 2009)

Time : Three Hours

Maximum : 100 Marks

Answer ALL questions

PART – A

(10×2=20 Marks)

1. What is a data mart ?
2. State why one of the biggest challenges when designing a data warehouse is the data placement and distribution strategy.
3. Define how the complex aggregation at multiple granularities is achieved using multi-feature cubes ?
4. What is time series analysis ?
5. Differentiate between data characterization and discrimination.
6. Give the need for data pre-processing.
7. List the two interesting measures of an association rule.
8. What is decision tree induction ?
9. Let $x_1 = (1, 2)$ and $x_2 = (3, 5)$ represent two points. Calculate the Manhattan distance between the two points.
10. How outliers may be detected by clustering ?



11. a) Explain the three tier architecture of a data warehouse with diagrammatic illustration. **(16)**
- (OR)
- b) Explain star schema and snowflake schema with example and diagrammatic illustration. **(16)**
12. a) i) Diagrammatically illustrate and discuss the architecture of MOLAP and ROLAP. **(12)**
- ii) Compare MOLAP and ROLAP. **(4)**
- (OR)
- b) List and discuss the features of Cognos Impromptu. **(16)**
13. a) i) Describe the various descriptive statistical measures for data mining. **(8)**
- ii) What are the major issues in data mining ? Explain. **(8)**
- (OR)
- b) i) What is attribute-oriented induction ? Describe how this is implemented. **(8)**
- ii) Discuss the various issues that have to be addressed during data integration. **(8)**
14. a) Discuss about mining association rules using the apriori algorithms. **(16)**
- (OR)
- b) Explain as to how neural networks are used for classification of data. **(16)**
15. a) Consider five points $\{X_1, X_2, X_3, X_4, X_5\}$ with the following coordinates as a two dimensional sample for clustering :
- $X_1 = (0, 2.25)$; $X_2 = (0, 0.25)$; $X_3 = (1.25, 0)$; $X_4 = (4.5, 0)$; $X_5 = (4.5, 2.5)$.
- Illustrate the K-means partitioning algorithm (clustering algorithm) using the above data set. **(16)**
- (OR)
- b) Explain with an example density-based local outlier detection. **(16)**
-