

9.5.19
B.E / B.Tech/ B.Arch (Full Time) END SEMESTER EXAMINATIONS, April / May 2019

INFORMATION TECHNOLOGY / CSE

CS8007 – DATA WAREHOUSING AND DATA MINING

(REGULATIONS 2012)

Time: 3 Hours

Answer ALL Questions

Max. Marks 100

PART-A (10 x 2 = 20 Marks)

1. Differentiate Data warehouse and Data mart
2. What is the role of metadata in DWH construction?
3. What are the differences between Business Intelligence and Business Analytics?
4. Differentiate OLAP and OLTP
5. What are the factors on which data mining system can be classified?
6. List and brief about the methods of attribute subset selection in dimensionality reduction
7. Name and brief about any two outlier detection methods
8. Differentiate Eager and Lazy learning methods
9. Comment on Web usage Mining
10. Define Outlier detection



Part – B (5 x 16 = 80 marks)

11. (i) Explain the data warehousing components and explain various steps involved in building a data warehouse (8)
- (ii) Explain the various schemas and OLAP operations for Multidimensional Data Model (8)
12. (a) (i) Explain 3-tier DWH architecture and ETL process in detail (10)
- (ii) Why are Business Intelligence applications so important? List down and brief about some applications of Business intelligence (6)

(OR)

12. (b) (i) Explain about the various types of OLAP servers (8)
(ii) Write short notes on distributed and virtual data warehouses (8)

13. (a) (i) Explain in detail about Data Integration and Data Transformation in data pre-processing technique (8)
(ii) Discuss the need for Data Mining & Explain about various data mining Functionalities (8)

(OR)

13. (b) A database has following transactions. Let $\text{min_sup}=60\%$ & $\text{min_conf}=70\%$ (16)

TID	List of item_Ids
T100	I1,I2,I5
T200	I2,I4
T300	I2,I3
T400	I1,I2,I4
T500	I1,I3
T600	I2,I3
T700	I1,I2,I3,I5
T800	I1,I2,I3



Find all frequent Itemsets using Apriori & FP growth and compare the efficiency of two Algorithms

14. (a) (i) The following table consists of training data from an employee database. The data have been generalized. Let status be the class label attribute. Construct Decision tree from the given data (8)
- (ii) Given a data tuple having the values "systems", "26....30" and "46-50K" for the attributes department, age and salary, respectively, what would a naïve Bayesian classification of the status for the tuple be? (8)

DEPT	STATUS	AGE	SALARY	COUNT
sales	Senior	31...35	46K...50K	30
sales	Junior	26...30	26K...30K	40
sales	Junior	31...35	31K...35K	40
systems	Junior	21...25	46K...50K	20
systems	Senior	31...35	66K...70K	5
systems	Junior	26...30	46K...50K	3
systems	Senior	41...45	66K...70K	3
marketing	Senior	36...40	46K...50K	10
marketing	Junior	31...35	41K...45K	4
secretary	Senior	46...50	36K...40K	4
secretary	Junior	26...30	26K...30K	6

(OR)

14. (b) Write short notes of the following:

(16)

- (i) k-means clustering
- (ii) Hierarchical based clustering
- (iii) Constraint based clustering
- (iv) Grid based clustering



15. (a) Write short notes on the following:

(8+8)

- (i) Mining Spatial Association and Co-location Patterns
- (ii) Multimedia Data mining

(OR)

15. (b) Explain the role of data mining applications in Health care, Agriculture and Telecommunication industry

(16)