Printed Pages: 4 NCS-066

(Following Paper ID and Roll No. to be filled in your Answer Books)	
Paper ID : 110666	Roll No.

B. TECH.

Theory Examination (Semester-VI) 2015-16

DATA WAREHOUSING & DATA MINING

Time : 3 Hours *Max. Marks : 100*

Note: Attempt questions from all Sections as per directions.

Section-A

1. **Attempt all parts of this section. Answer in brief.** [2×10=20]

   (a) Write some of the facts of the association rule mining.

   (b) Briefly explain the concept of Frequent Item sets and Closed Item sets.

   (c) Briefly explain important approaches to build the data warehouse.

   (d) Define KDD. Identify the phases in KDD process.

   (e) Why data warehouse is maintained separately from database?

quescol.com
https://quescol.com | Free PYQs, Solutions, Syllabus, Notes etc

(f) Describe the market basket analysis.

(g) Write short notes on Linear Regression.

(h) Sketch a neat diagram of architecture of a typical Data Mining System.

(i) Compare Roll up, Drill down, Slice and Dice operations.

(j) Name and describe the main features of Genetic Algorithm.

Section-B

2. Attempt any five questions from this section. (110×5=50)

(a) Draw the 3-tier data warehouse architecture. Explain ETL process.

(b) Explain the various types of OLAP servers. What are the steps for efficient processing of OLAP queries?

(c) Write the algorithm of decision tree induction. What are the methods that can be used for selecting the splitting criteria?

(d) Draw a box-and-whisker plot for the following data set :

126, 132, 138, 140, 141, 141, 142, 143, 144, 144, 144, 145, 146, 147, 148, 148, 149, 149, 150, 150, 150, 154, 155, 158, 158.

Also find the outliers.

https://quescol.com | Free PYQs, Solutions, Syllabus, Notes etc

(e) What is Hierarchical method for clustering? Explain BIRCH method.

(f) Give the difference between the star and fact constellation multidimensional data model.

(g) Describe the types of data that often occur in cluster analysis and briefly explain ' how to pre process that data for clustering:

(h) Explain how query performance can be improved by cascading the operations.

Section-C

Attempt any two questions from this section. (15×2=301)

3. Classify the tuple X= {Color = 'RED', Type = 'SUV' Origin = 'DOMESTIC'} using Naive Bayesian classification. Training data is given in the following table where class label is {STOLEN}.

Color	Type	Origin	Stolen?
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Yellow	Sports	Domestic	No
Yellow	Sports	Imported	Yes

https://quescol.com | Free PYQs, Solutions, Syllabus, Notes etc

Yellow	SUV	Imported	No
Yellow	SUV	Imported	Yes
Yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes

4. (i) Describe the difference between the following approaches for the integration of data mining system with database or data warehouse systems: no coupling, loose coupling and semi tight coupling.

(ii) Define and describe the basic similarities and differences among ROLAP, MOEAP and HOLAP.

5. Explain Chi-square test method. Show using chi-square test that gender and preferred reading are independent or not from given table. (Given are the observed counts).

	Male	**Female**	**Total**
Fiction	250	200	450
Non-Fiction	50	1000	1050
Total	300	1200	1500

https://quescol.com | Free PYQs, Solutions, Syllabus, Notes etc