

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY

MCA - SEMESTER-IV • EXAMINATION – SUMMER • 2015

Subject Code: 640005

Date: 20-05-2015

Subject Name: Data Warehousing and Data Mining (DWD)

Time: 10:30 am - 01:00 pm

Total Marks: 70

Instructions:

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

- Q.1 (a)** Explain the following terms: **07**
1. Attribute Oriented Induction.
 2. Bootstrap.
 3. Name some conventional visualization techniques.
 4. Clustering.
 5. Decision tree.
 6. Association Rule.
 7. Snowflake Schema.
- (b)** State whether the following statements are True or False. Justify your answer. **07**
1. The confusion matrix is a useful tool for analyzing how well a classifier can recognize tuples of different classes.
 2. A time-series database stores sequences of ordered events, with or without a concrete notion of time.
 3. The 0-D Cuboid, which holds the highest level of summarization stated as apex cuboid.
 4. Drill-Down operation performs aggregation on a data cube.
 5. Cross-validation and bootstrap are two techniques that increase the accuracy of decision tree.
 6. Poisson Regression is the extension of straight-line regression that can involve more than one predictor variable.
 7. Classification and numeric prediction are the two major types of prediction problems.
- Q.2 (a)** In the process of data cleaning, how can we fill up the missing values? Write down its methods. **07**
- (b)** Discuss the following as attribute selection measure with example **07**
1. Information gain
 2. Gain ratio
- OR**
- (b)** Describe the differences between the following approaches for the integration of a data mining system with a database or data warehouse system: no coupling, loose coupling, semi-tight coupling, and tight coupling. State which approach you think is the most popular, and why. **07**
- Q.3 (a)** Suppose the data for analysis is- **07**
- 13, 15, 16, 16, 19, 20, 20, 21, 25, 30, 33, 111
1. Compute mean, median, midrange, five number summary & inter-quartile range.
 2. What is mode & modality of data?
 3. Show a box-plot of data.
- (b)** Describe Decision Tree Induction algorithm. You can describe it with the help of an example. **07**

OR

- Q.3 (a)** Explain different OLAP servers. **07**

- (b) Explain the three-tier architecture of a data warehouse. 07
- Q.4** (a) What is Apriori property? How the Apriori property is used in finding frequent itemset. 07
- (b) Explain several OLAP operations in multidimensional data model with example. 07
- OR**
- Q.4** (a) Explain Indexing techniques for OLAP data. 07
- (b) Describe Data Reduction. List and explain its several strategies. 07
- Q.5** (a) What is Bayes's theorem? Explain the working of Naïve Bayesian Classifier. 07
- (b) 1. Describe the following applications of Data Warehousing & Data Mining for the Finance Sector: 04
- (i) Loan payment prediction and Customer credit policy analysis.
- (ii) Detection of money laundering and other financial crimes.
2. What do you mean by cross validation? 03
- OR**
- Q.5** (a) Discuss two Ensemble methods for increasing the accuracy of classifier. 07
- (b) Discuss the application of Data Mining in Retail Industry. 07
