Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–VII(NEW) EXAMINATION – SUMMER 2019

Subject Code:2170715 Date:18/05/2019

Subject Name:Data Mining and Business Intelligence

Time:02:30 PM TO 05:00 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 cluster analysis and outlier analysis with example. **03**

(b) A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data – Justify. **04**

(c) Consider following database of ten transactions. Let min_sup = 30% and min_confidence = 60%.

TID	items bought
T1	pen, pencil
T2	book, eraser, pencil
T3	book, chalk, eraser, pen
T4	chalk, eraser, pen
T5	book, pen, pencil
T6	book, eraser, pen, pencil
T7	ink, pen
T8	book, pen, pencil
T9	eraser, pen, pencil
T10	book, chalk, pencil

A) Find all frequent itemsets using Apriori algorithm. **05**

B) Generate strong association rules. **02**

Q.2 **(a)** Discuss following terms. **03**

1) Supervised learning 2) Correlation analysis 3) Tree pruning

(b) What is noise? Explain binning methods for data smoothing. **04**

(c) Discuss data warehouse architecture in detail. **07**

OR

(c) Write and discuss the algorithm which is used to generate frequent itemsets using an iterative level-wise approach based on candidate generation. **07**

Q.3 **(a)** Which are the two measures of rule interestingness? Explain with example. **03**

(b) Discuss Hash-based technique to improve efficiency of Apriori algorithm. **04**

(c) Explain various data normalization techniques. **07**

OR

Q.3 **(a)** Discuss Big Data. **03**

(b) Discuss possible ways for integration of a Data Mining system with a Database or DataWarehouse system. **04**

(c) Enlist data reduction strategies and explain any two. **07**

Q.4 **(a)** Discuss various layers of multilayer feed-forward neural network with diagram. **03**

(b) What is apex cuboid? Discuss drill down and roll up operation with diagram. **04**

(c) Using Naive Bayesian classification method, predict class label of X = (age = youth, income = medium, student = yes, credit_rating = fair) using following training dataset. **07**

age	income	Student	credit_rating	Class: buys_computer
youth	high	no	Fair	no
youth	high	no	excellent	no
middle_aged	high	no	fair	yes
senior	medium	no	fair	yes
senior	low	yes	fair	yes
senior	low	yes	excellent	no
middle_aged	low	yes	excellent	Yes
youth	medium	no	fair	no
youth	low	yes	fair	yes
senior	medium	yes	fair	yes
youth	medium	yes	excellent	yes
middle_aged	medium	no	excellent	yes
middle_aged	high	yes	fair	yes
senior	medium	no	excellent	no

OR

Q.4 **(a)** Explain various conflict resolution strategies in rule based classification. **03**

(b) What is classification? Explain classification as a two step process with diagram. **04**

(c) Discuss fraud detection and click-stream analysis using data mining. **07**

Q.5 **(a)** Compare data mart and data warehouse. **03**

(b) Discuss star schema and fact constellation schema with diagram. **04**

(c) What do you mean by learning-by-observation? Explain k-Means clustering algorithm in detail. **07**

OR

Q.5 **(a)** Discuss following terms. **03**

1) DataNode 2) NameNode 3) Text mining

(b) Discuss attribute subset selection. **04**

(c) Compare OLAP and OLTP in detail. **07**

****