

Kadi Sarva Vishwavidyalaya

M.E Semester-I Examination

Subject: Major Elective-I Data Mining and Business Intelligence

Date: 22/01/2013

Time: 10:00 am to 01:00 pm

Maximum Marks: 70

Instructions:

1. Answer each Section in Separate Answer sheet.
2. Use of Scientific Calculator is permitted.
3. Make suitable assumptions wherever necessary.

Section-I

- Q.1 (A) Define the term "Data Mining". Write down short note on KDD process. [05]
(B) Clearly state difference between Data Warehouse and Operational Database system. [05]
(C) Explain the "Star" and "Snowflake" Schemas of Data Warehouse. [05]

OR

- (C) Define the following terms: [05]
1) Spatial Database
2) Time Series Database
3) Heterogeneous Database
4) Legacy Database
5) Multimedia Database

- Q.2 (A) In real world data, tuples with missing values for some attributes are a Common occurrence. Describe various methods for handling this problem. [05]
(B) University course database for KSV University contains the following attributes: the name, address, status (e.g., undergraduate or graduate), and major of each student, and their cumulative grade point average (GPA). [05]
1) Propose a concept hierarchy for the attribute status, major, GPA and address.

OR

- Q.2 (A) Briefly explain issues to consider during Data Integration. [05]
(B) List and describe the primitives for specifying data mining task. [05]

- Q.3 (A) Explain process of designing and developing the Business Intelligence applications. [05]
(B) Explain Factors and Obstacles Driving BI. [05]

OR

- Q.3 (A) What is BI? Explain BI as a Process, Product, Solution and Tool for an Organization. [05]
(B) Importance of Business Intelligence Applications [05]

Section-II

- Q.4 (A) Write steps of the k-means clustering algorithm with its limitation [05]
(B) Why naïve Bayesian classification is called "naïve"? Briefly outline the major ideas of naïve Bayesian classification. [05]
(C) List the various clustering methods. Explain any two in details. [05]

OR

- (C) Briefly outline the major steps of decision tree classification [05]
Q.5 (A) Explain the steps of the "Apriori Algorithm" for mining frequent itemsets. [05]
(B) Compare the merits and demerits of eager classification (e.g., decision tree, Bayesian, Neural Networks) Versus lazy classification (e.g., k-nearest neighbor, case-based reasoning) [05]

[P.T.O]

OR

Q.5 (A) List approaches to mining multilevel association rules. Explain any two approaches in detail. [05]

(B) Briefly explain various methods for the generation of concept hierarchies for categorical data. [05]

Q.6 (A) Explain Types of BI Applications. [05]

(B) Explain process of Deploying and Supporting the DW/BI System. [05]

OR

Q.6 (A) What is Dashboard? Explain structured approach of designing Dashboard. [05]

(B) Explain ETL (Extract, Transport and Load) Process. [05]

OR

- 1) Spatial Database
- 2) Time Series Database
- 3) Heterogeneous Database
- 4) Legacy Database
- 5) Multischema Database

OR

- (A) In real world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem.
- (B) University course database for KSV University contains the following attributes: the name, address, status (e.g., undergraduate or graduate), and major of each student and their cumulative grade point average (GPA).
- (C) Propose a concept hierarchy for the attribute status, major, GPA and address.

OR

- (A) Briefly explain issues to consider during Data Integration.
- (B) List and describe the primitives for specifying data mining task.

OR

- (A) Explain process of designing and developing the Business Intelligence applications.
- (B) Explain Factors and Obstacles Driving BI.

Section-II

OR

- (A) Write steps of the k-means clustering algorithm with its limitation.
 - (B) Why naive Bayesian classification is called "naive"? Briefly outline the major ideas of naive Bayesian classification.
 - (C) List the various clustering methods. Explain any two in details.
- OR**
- (A) Explain the steps of the "Apriori Algorithm" for mining frequent itemsets.
 - (B) Compare the merits and demerits of eager classification (e.g., decision tree, Bayesian, Neural Networks) Versus lazy classification (e.g., k-neighbor, case-based reasoning).