Q.P. Code : 595001

(3 Hours) [Total Marks : 80

N.B. : (1) **Q1** is **compulsory**.
    (2) Attempt any **three** from the remaining.
    (3) Assume suitable data.

1. (a) Define "Data Mining". Enumerate five example applications that can benefit by using Data Mining.   5

   (b) Clearly explain the data preprocessing phase for data mining.   5

   (c) Describe one hierarchical clustering algorithm using an example dendrogram.   5

   (d) Explain the concept of a decision support system with the help of an example application.   5

2. (a) Partition the given data into 4 bins using Equi-depth binning method and perform smoothing according to the following methods.   10
Smoothing by bin mean
Smoothing by bin median
Smoothing by bin boundaries

Data: 11,13,13,15,15,16,19,20,20,20,21,21,22,23,24,30,40,45,45, 45, 71, 72, 73, 75

   (b) For the same set of data points in question 2. (a)   10
     (a) Find Mean, Median and Mode.
     (b) Show a boxplot of the data. Clearly indicating the five- number summary.

[TURN OVER

3. (a) The table below shows a sample dataset of whether a customer   10
responds to a survey or not. **"Outcome"** is the class label.
Construct a Naive Bayes' Classifier for the dataset. For a new example
(Rural, semidetached, low, No), what will be the predicted class label?

District	House Type	Income	Previous Customer	Outcome
Suburban	Detached	High	No	Nothing
Suburban	Detached	High	Yes	Nothing
Rural	Detached	High	No	Responded
Urban	Semi-detached	High	No	Responded
Urban	Semi-detached	Low	No	Responded
Urban	Semi-detached	Low	Yes	Nothing
Rural	Semi-detached	Low	Yes	Responded
Suburban	Terrace	High	No	Nothing
Suburban	Semi-detached	Low	No	Responded
Urban	Terrace	Low	No	Responded
Suburban	Terrace	Low	Yes	Responded
Rural	Terrace	High	Yes	Responded
Rural	Detached	Low	No	Responded
Urban	Terrace	High	Yes	Nothing

(b) Briefly explain Regression based Classifiers   10

4. (a) Use the Apriori algorithm to identify the frequent item-sets in the following database. Then extract the strong association rules from these sets.   **10**

Min. Support = 30% Min. Confidence = 75%

TID	Items
01	A, B, D, E, F
02	B, C, E
04	A, B, D, E
04	A, B, C, E,
05	A, B, C, D, E, F
06	B, C, D
07	A, B, D, E

(b) Explain multidimensional and multi level Association rules with examples.   **10**

5. (a) What is clustering? Explain k-means clustering algorithm. Suppose the date for clustering is {2,4,10,12,3,20,11,25} Consider k = 2, cluster the given data using K-means algorithm   **10**

(b) What is an outlier? Describe methods that can be used for outlier analysis.   **10**

6. (a) Consider the following case study: A telecom company wants to analyze and improve its performance by introducing a series of innovative mobile payment plans. For this case study design a BI system, clearly explaining all steps from data collection to decision making.   **10**

(b) Clearly explain the working of the DBSCAN algorithm using appropriate diagrams.   **10**

(3 Hours)   [Total Marks :80

N.B. : (1) Question No. 1 is compulsory.
       (2) Attempt any **three** questions from Q.2 to Q.6.

1. (a) Explain CMM   5
    (b) Give difference between Waterfall and Prototype Model   5
    (c) List Software Engineering Practice core Principles.   5
    (d) Explain following design concepts: Abstraction, Modularity   5

2. (a) What is agility? Explain XP   10
    (b) What is Design? Explain Design Principles   10

3. (a) Draw control flow graph and find cyclomatic complexity for the following PDL   10
         if(c1or c2 and c3) s1;
         else s2;
         while(c4) s3; s4;
         do s5; while(c5);
         s6

    (b) Explain testing strategies   10

4. (a) How important is requirement analysis. Elaborate on different requirement   10
         engineering tasks.

    (b) For the given Order processing system scenario draw DFD level 0, 1and 2. The   10
         customer can place order, cancel order, do modification in the placed order
         before it is delivered. The order is delivered to the customer address by the
         courier company. Customer can make payment for the placed order using credit
         card or net banking. If customer is not satisfied by the product he can return
         the product within 15 days from the delivery date. The refund is deposited back
         to the customer account within 10 days from the date of product return.

5. (a) Explain Refinement, Refactoring and design classes.   10
    (b) Explain different architectural styles.   10

6. (a) Explain Coupling and Cohesion. What is preferred in the component? Why?   5
    (b) Identify risk of completing graduation with good marks but without knowledge   5
    (c) Explain software quality attributes.   5
    (d) Explain change management process.   5

———————

(3 Hours) [Total Marks : 80

N.B. : (1) Question No. **1** is **compulsory**.
(2) Attempt any **THREE** Questions out of remaining **FIVE** questions.

1. (a) Give two techniques to establish a covert channel. 5
   (b) Compare and contrast discretionary access control and mandatory access control. 5
   (c) Define with examples i) SQL injection ii) Cross-site scripting. 5
   (d) What are the different phases of a virus? Explain. 5

2. (a) What are the different kinds of malware? How do they propagate? 10
   (b) Explain RSA algorithm for public key encryption. Given modulus N = 143 and public key =7, find the values of p, q, phi (n), and private key d. Can we choose value of e=5? Justify. 10

3. (a) What is a firewall? Explain different types of firewalls and specify at which layer of the Internet stack do they operate? 10
   (b) What is a denial of service attack? Discuss different ways in which an attacker can mount a DOS attack. 10

4. (a) Distinguish between the ESP and AH protocol in IPSec. Show the working of each in transport and tunnel mode. 10
   (b) What is an IDS? How does it differ from a honeypot? Discuss the different types of IDS. 10

5. (a) Explain the process of generation and verification of digital certificate. 10
   (b) With respect to SSL protocol explain the following:- 10
       (i) Generation of master Key
       (ii) Authentication of server to client.

6. Write short notes on (**any four**) : 20
   (a) Windows Security
   (b) Federated Identity Management
   (c) Software Reverse Engineering
   (d) Knapsack cryptosystem
   (e) Non-malicious programming errors

Q.P. Code : 581702

(3 Hours) [Total Marks : 80

N.B. : (1) Question No.1 is Compulsory.
  (2) Attempt any 3 questions out of the rest
  (3) Figure to the right indicate full marks.
  (4) All question carry equal marks.

1. a) What are advantages and disadvantage od Distributed DBMS 5
   b) What are the features ofDDBMS? 5
   c) Explain the basic Timestamp Ordering Algorithm. 5
   d) What are the objectives of Distributed Query Processing? 5

2 a) What is horizontal and vertical fragmentation? What are the types of 10
     horizontal fragmentation. Perform horizontal fragmentation for student
     relation as given below.
     Also give the correctness criteria for it.
     Student (Studentrollno., StudentName, CourseNarne, CourseFees, year)
  b) What are the various kinds of transparencies in distributed database design? 10
     Explain each with the help of an example

3 a) What are the various concurrency control techniques? Compare Lock based 10
     Concurrency Control strategies in detail.
  b) Compare Distributed Deadlock prevention to Distributed Deadlock 10
     Avoidance
     Explain one scheme of Distributed deadlock Detection and Recovery.

4 a) A banking database should contain the customer's information along with 10
     the types of accounts customer is maintaining. Customer information is
     its full profile information along with his current address, PAN ID, adhar
     Card no. included and account information should include type of account
     (Saving, fixed, demat, recuring, current), date and time of access and the
     transactions details.
     i)   Write the DTD rules for the above XML documents.
     ii)   Create an XML schema for the above XML document.
  b) What are homogenous and heterogeneous database. Give the architecture 10
     of heterogeneous databases along with some query processing issues.

TURN OVER

10

5 a) What problems can occur in a distributed system due to the failure of link? What are the ways by which recovery can take place?

10

b) Explain the phases of query processing in distributed database.

20

6 Answer any two:
   a) Bond Energy Algorithm
   b) Design issues of Distributed Database
   c) 3PC
   d) Transaction management model for distributed System.