Unit – I

1. a) What is Data Mining? Discuss goals of Data Mining.
2. Describe the database system evolutionary path.

   b) Explain any two Data Mining Techniques.

2. a) Detail on the architecture of data mining system with a suitable diagram.

   b) What are the difference between DBMS and Machine learning

3. a) Define each of the following data mining functionalities:

   i) Data Cleaning ii) Data Discretization

   ii) With the schematic diagram, describe the architecture of a data mining system.

4. Discuss the issues in data mining in detail.
5. Discuss in detail about the steps in knowledge discovery in data base.
6. Explain different techniques in data mining.

Unit – II

1. With a neat diagram explain the components of data warehouse.
2. Discuss the various OLAP operations in the Multidimensional Data Model.
3. Describe the different methods for data cleaning
4. Write short notes on: OLAP and Statistics
5. Explain Discretization and Concept Hierarchy Generation for Numeric Data
6. What are the various forms of visualizing the discovered patterns in data mining?
7. How does a data warehouse handle multi dimensional data? Explain the data structures and schema that support multi dimensional data with suitable illustration.

UNIT III

1. With examples, discuss the different attribute types and the transformations that attribute levels
2. Explain similarity and dissimilarity measures between simple attributes based on different types of attributes.
3. Explain FP tree algorithm with an example.
4. Explain Analytical Characterization?
5. Methods of Attribute Relevance Analysis?

6 (b) (i) Write an algorithm for FP-Tree construction and explain how frequent itemsets are generated from FP-Tree. (8)

(ii) Consider the following database has twelve transactions, let min_sup = 25% and find all frequent item set using the above algorithm. (8)

TID	List of items _ IDs
T1	11, 12, 15
T2	12, 14, 16
T3	12, 13
T4	11, 12, 14
T5	11, 13
T6	12, 13
T7	11, 13
T8	11, 12, 13, 15
T9	11, 12, 13
T10	11, 12, 14, 16
T11	15, 16
T12	13, 14, 15

5 a. What is a FP – growth algorithm? For the following transaction data set, explain and construct a FP – tree. If the ordering scheme is reversed, give the resulting FP – tree.

TID	Items
1	{a, b}
2	{b, c, d}
3	{a, c, d, e}
4	{a, d, e}
5	{a, b, c}
6	{a, b, c, d}
7	{a}
8	{a, b, c}
9	{a, b, d}
10	{b, c, e}

8. Define classification. Explain general approach or solving a classification problem
9. Explain how the decision tree works, with an example.
10. Explain K-nearest neighbor classification algorithm.
11. What are Baysian classifiers? Explain Baye's theorem for classification.
12. Explain outlier evaluation criteria for classification method.
13. Explain the classification method for estimating predictive accuracy.
14. Differentiate classification and predication

1. Classify the tuple
   X= (age=youth, income=medium, student=yes, credits-rating=fair)
   Using naïve Bayesian classification on the following training data set

RID	Age	Income	Student	Credit-rating	Class: buys-computer
1	Youth	High	No	Fair	No
2	Middle-aged	High	No	Fair	Yes
3	Senior	Medium	No	Fair	Yes
4	Senior	Low	Yes	Excellent	Yes
5	Youth	High	No	Excellent	No
6	Youth	Medium	No	Fair	No
7	Middle-aged	Medium	No	Excellent	Yes
8	Senior	Medium	Yes	Fair	Yes
9	Youth	High	No	Fair	No
10	Middle-aged	Medium	No	Excellent	Yes

12. (a) Write an algorithm for constructing a decision tree. Construct a decision tree for the following data set using information gain. Predict the class label for a data point with values < Female, 2, standard, high>. (16)

Gender	Car ownership	Travel cost	Income level	Transport mode
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car
Female	1	Cheap	Medium	Train
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train

UNIT V

1. Explain different types data in cluster analysis.
2. Write K-means and K-medoids algorithms and explain with examples.
3. Discuss the following clustering algorithm using examples :
   (i) K-means.
   (ii) K-medoid.
4. Explain agglomerative hierarchical clustering.