

Exam.Code:0918
Sub. Code: 33442

2056
B.E. (Computer Science and Engineering)
Sixth Semester
CS-601: Data Mining and Machine Learning

Time allowed: 3 Hours

Max. Marks: 50

NOTE: Attempt five questions in all, including Question No. 1 which is compulsory and selecting two questions from each Unit.

x-x-x

1. Attempt the following:-

- a) Differentiate between data mining and machine learning.
- b) Define ordinal and nominal attributes with examples.
- c) What is interquartile range (IQR)?
- d) Define support and confidence in association rule mining.
- e) What is overfitting?
- f) Define bias-variance tradeoff.
- g) What is cross-validation?
- h) What is kernel trick in SVM?
- i) Define silhouette coefficient.
- j) What is boosting?

(10x1)

UNIT - I

2. A retail supermarket chain collects transaction data (items purchased together, timestamps, customer IDs).
 - a) Identify suitable data mining tasks for this dataset.
 - b) Design a data preprocessing pipeline.(5,5)
3.
 - a) Explain measures of central tendency and dispersion. Compare their suitability for skewed datasets.
 - b) Given a dataset, how will you identify whether normalization or standardization is more appropriate and justify.(5,5)
4.
 - a) Explain Apriori algorithm with candidate generation and pruning steps.
 - b) Discuss limitations of Apriori for large datasets.(5,5)

P.T.O.

(2)

UNIT - II

5. A bank wants to predict whether a customer will default on a loan.
 - a) Compare Decision Tree, Naïve Bayes, and k-NN for this problem.
 - b) Suggest how class imbalance can be handled.

(7,3)

6.
 - a) Explain k-means clustering and compare with hierarchical clustering.
 - b) Suggest how to determine optimal number of clusters.

(7,3)

7. A fraud detection system needs high accuracy and adaptability.
 - a) Explain how ensemble methods (Bagging, Boosting, AdaBoost) improve performance.
 - b) Suggest use of transfer learning in fraud detection.

(5,5)

x-x-x