Atal Bihari Vajpayee Indian Institute of Information Technology & Management, Gwalior

IT405: Data Mining

Major Examination (Session 2024–25)

Maximum Time: 3 Hours   Max Marks: 70

Note: Attempt all questions. Answer concisely but clearly. Numerical steps must be shown.

1. (a) Define data mining and explain its role in knowledge discovery in databases (KDD). (6 Marks)
   (b) Differentiate between descriptive and predictive data mining tasks with examples. (4 Marks)

2. (a) Describe the Apriori algorithm with an example. Why is candidate pruning important? (7 Marks)
   (b) For the following transaction database: T1: A,B,C, T2: A,C, T3: A,B, T4: B,C, T5: A,B,C Find all frequent itemsets with minimum support = 40%. Derive at least two strong rules with min confidence = 60%. (7 Marks)

3. (a) Explain ID3 decision tree algorithm with steps. (6 Marks)
   (b) A dataset has 14 examples, 9 positive and 5 negative. Calculate the entropy of the dataset. (4 Marks)

4. **Clustering (12 Marks)**: (a) Compare partitioning methods (k-means) and hierarchical methods with pros/cons. (5 Marks)
   (b) Apply one iteration of k-means to 2D data points: (1,1),(2,2),(8,8),(9,9) with initial centroids (1,1) and (9,9). Show calculations. (7 Marks)

5. (a) Define overfitting in classification. How can cross-validation help prevent it? (5 Marks)
   (b) Compute Precision, Recall, F1-score and Accuracy given: TP=50, FP=10, FN=15, TN=25. (5 Marks)

6. Short notes (any two, 6 Marks each): (a) Ensemble learning: Bagging vs Boosting.
   (b) PCA for dimensionality reduction.
   (c) Data preprocessing: normalization and discretization.   (12 Marks)

7. **Case Study (18 Marks)**: An e-commerce company wants to use data mining to:
   - Recommend products to customers based on purchase history. - Detect fraudulent transactions. - Segment customers for targeted marketing.

   For each task: (i) Identify the suitable data mining technique. (ii) Outline preprocessing and features required. (iii) Suggest evaluation metrics. (iv) Mention one risk (e.g., bias, privacy) and mitigation.