विश्वजीवनामृतं ज्ञानम्

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

IT405: Data Mining

Minor Examination (Session 2023–24)

Maximum Time: 1.5 Hours **Max Marks: 25**

Note: Attempt all questions. Show calculations for numerical parts.

1. **Multiple Choice (1 mark each)**: (a) Which algorithm is primarily used for classification by recursive partitioning?
   (i) K-means (ii) Apriori (iii) Decision tree (C4.5) (iv) PCA
   (b) Lift in association rule mining measures:
   (i) Support of rule (ii) Strength relative to random occurrence (iii) Confidence only (iv) None
   (2 Marks)

2. **True/False (with one-line justification) — 2 marks each**: (a) In k-means clustering the number of clusters k must be known beforehand.
   (b) PCA is a supervised dimensionality reduction technique. (4 Marks)

3. Explain the difference between **support** and **confidence** in association rule mining. Provide a short example. (4 Marks)

4. **Numerical (k-means — 6 Marks)**: Given 1D data points: {2, 4, 5, 10, 12, 14} and initial centroids at 4 and 12, perform one iteration of k-means clustering (assign points and compute new centroids). Show steps and results. (6 Marks)

5. Briefly explain: (a) Overfitting in classification and two methods to avoid it. (b) ROC curve and AUC interpretation. (4 Marks)

6. Short notes (any one — 5 Marks): (a) Density-based clustering (DBSCAN) — idea and parameters.
   (b) Handling missing data — two common strategies and when to use them. (5 Marks)