



विश्वजीवनामृतं ज्ञानम्

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

## IT405: Data Mining

Minor Examination (Session 2024-25)

Maximum Time: 1.5 Hours

Max Marks: 25

---

*Note: Attempt all questions. Write all steps for numerical parts.*

1. **MCQs (1 mark each):** (a) Which of the following is a density-based clustering algorithm? (i) K-means (ii) DBSCAN (iii) Decision Tree (iv) Apriori  
(b) The curse of dimensionality refers to: (i) Increased sparsity of data in high dimensions (ii) Reduced noise in data (iii) Automatic feature selection (iv) None (2 Marks)
2. **True/False with justification (2 marks each):** (a) Confidence of an association rule is always less than or equal to its support.  
(b) Clustering is a supervised learning task. (4 Marks)
3. Distinguish between **classification** and **clustering** with suitable examples. (3 Marks)
4. **Numerical (k-means):** Cluster the following 1D data {1, 2, 3, 10, 12} into 2 clusters using k-means with initial centroids 2 and 10. Perform one full iteration (assignment + recompute centroids). (6 Marks)
5. Short answer (any two): (a) What is feature selection and why is it needed in data mining? (b) Explain the concept of lift in association rules. (c) Give two differences between supervised and unsupervised learning. (6 Marks)
6. Write short notes (any one): (a) Outlier detection in data mining. (b) Dimensionality reduction using PCA. (4 Marks)