DATA AND TEXT MINING

DOCUMENTATION RESEARCH AND TRAINING CENTRE, ISI-BC

Final EXAMINATION (2026)

Total Marks: 100 **Time: 180 mins.**

ANSWER Any FIVE

Important Information: PROVIDE SUFFICIENT AND RELEVANT EQUATIONS, DESCRIPTIONS, AND ILLUSTRATIONS TO JUSTIFY YOUR ANSWER.

1. Illustrate the working principle of the minimum distance and K-NN classification models. Compare their advantages and disadvantages. [20]
2. Discuss the measures of Location, Spread, Shape, and Dependence with their applications. [20]
3. If X is a data set with 9 samples and 2 features each. Find the Euclidean and Mahalanobis distances between 2nd and 5th samples of XX. Explain the reasons for the different distance values and their merits and demerits. [20]

```
XX =[4  5
     3  2
     1  6
     4  0
     5  5
     6  7
     7  7
     8  9
     9  7];
```

4. [10+10]
   - Provide a complete interpretation of the Box and Whisker Plot Analysis of a data set. How it behave for symmetric and skewed data?
   - Describe the significance and operational steps of semisupervised and Reinforcement learning. How is it different from supervised and unsupervised learning?
5. State and describe the working steps of a linear classifier, starting from the given data set to the performance analysis of the model. [20]
6. What is the distance metric in the decision-making process of data mining? List out the common properties of a distance METRIC. Discuss the various distance metrics, starting from Minkowski distance to Mahalanobis distance. [20]

==========END of the Question Paper==========