

# CBCS SCHEME



USN

--	--	--	--	--	--	--	--	--	--

18AI644

## Sixth Semester B.E. Degree Examination, June/July 2024 Foundations for Data Science

Time: 3 hrs.

Max. Marks: 100

*Note: Answer any FIVE full questions, choosing ONE full question from each module.*

### Module-1

- 1 a. Explain the process of data analysis for prediction of customer churn. (08 Marks)
- b. List the popular data-mixing tasks and give a brief description of any four tasks. (12 Marks)

OR

- 2 a. Illustrate the salient differences in evolution from Big Data 1.0 to Big Data 2.0. (08 Marks)
- b. Explain the data mining process through the CRISP model. (12 Marks)

### Module-2

- 3 a. Explain the concept of supervised segmentation, with suitable example. (08 Marks)
- b. Sketch the solution for addressing churn problem with tree induction. (12 Marks)

OR

- 4 a. Demonstrate the concept of supervised segmentation with Tree-structure model. (12 Marks)
- b. Explain the concept of Induction and prediction, with suitable example each. (08 Marks)

### Module-3

- 5 a. Explain the concept of Linear Discriminant functions taking example of IRIS Dataset. (10 Marks)
- b. Briefly explain the concept of Support Vector Machines. (10 Marks)

OR

- 6 a. Describe "Loss function" with respect to data science. Explain following Loss function :
  - (i) Hinge loss.
  - (ii) Zero-one loss.
  - (iii) Squared loss. (08 Marks)
- b. Explain the concept of overfitting. Illustrate the techniques of avoiding overfitting. (08 Marks)
- c. Explain the avoidance of overfitting with Tree Induction. (04 Marks)

### Module-4

- 7 a. Explain the concept of Nearest Neighbours algorithm for predictive algorithm with an example. (08 Marks)
- b. List the issues associated with Nearest – Neighbour method. Briefly detail salient features of each of issues. (12 Marks)

OR

- 8 a. Briefly explain the concept of hierarchical clustering, with diagrammatic illustration. (10 Marks)
- b. Briefly explain concept of clustering around centroids, with an example. (10 Marks)

**Module-5**

- 9 a. Explain the need for confusion matrix, with its salient features. (06 Marks)  
b. Illustrate steps involved in simple bag-of-words using term frequency. (10 Marks)  
c. Describe briefly problem with unbalanced classes. (04 Marks)

OR

- 10 a. Explain the concept of association with respect to data-science task. Describe the following association surprise measuring techniques:  
(i) Lift (10 Marks)  
(ii) Leverage  
b. Briefly explain the concept of Inverse Document Frequency (IDF) for text classification, with steps. (10 Marks)

\*\*\*\*\*