USN | | | | | | | | | | |

18AI644

Sixth Semester B.E. Degree Examination, June/July 2023
Foundation 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. Discuss how data science capability in a strategic asset. **(08 Marks)**
   b. Explain with example, the fundamental concepts of data science that study the extraction of useful knowledge from data analytics. **(12 Marks)**

OR

2 Explain CRISP data mining process in detail with required diagram. **(20 Marks)**

Module-2

3 a. Explain about supervised segmentation and the process of selecting informative attributes. **(14 Marks)**
   b. Discuss on data querying methods for queries posed by an analyst. **(06 Marks)**

OR

4 a. Explain supervised segmentation with Tree structured models with an example of a simple classification tree. **(10 Marks)**
   b. How to represent trees as a set of rules? Explain with an example. **(05 Marks)**
   c. Discuss how to address the churn problem with tree Induction. **(05 Marks)**

Module-3

5 a. Explain Linear Discriminant functions. **(10 Marks)**
   b. How to optimize an objective function? **(05 Marks)**
   c. Discuss on how Linear Discriminant functions are used for scoring and ranking instances. **(05 Marks)**

OR

6 a. Discuss about Logistic Regression. **(10 Marks)**
   b. Why overfitting often causes models to become worse? Explain in detail. **(10 Marks)**

Module-4

7 a. How predictive modeling performed using nearest neighbors concept? Explain in detail with required example. **(10 Marks)**
   b. Briefly discuss about the formulas used for calculating the prediction of an instance from a set of the instances neighbors. **(10 Marks)**

OR

8 a. Explain Hierarchial clustering in detail with an example of Whiskey analysis. Support your explanation with a dendogram. **(10 Marks)**
   b. Explain the technique of clustering around centroids with the help of K-means algorithm.
                                                                                        (10 Marks)

Module-5

9 a. What is good model? Explain how to evaluate classifiers? **(10 Marks)**
   b. Discuss about confusion matrix with an example. **(05 Marks)**
   c. Why is text important and why is text difficult? Explain. **(05 Marks)**

OR

10 a. What is TFIDF? Explain how IDF is related to Entropy? **(10 Marks)**
    b. What do you mean by Co-occurrences and association? Explain with example. **(10 Marks)**

* * * * *