CBCS SCHEME

USN |__|__|__|__|__|__|__|__|__|__|

First Semester M.Tech. Degree Examination, Feb./Mar. 2022
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. What is Data Science? Explain in detail about Big data and Data Science hype. Why hype?
(10 Marks)

b. What is Datafication? Discuss in detail about the current landscape of perspectives and skill sets.
(10 Marks)

OR

2   a. What is Statistical Thinking? Discuss in detail statistical inferences and also explain about population and samples.
(10 Marks)

b. What is Statistical Modelling? Explain in detail about probability distribution and fitting a model. Write about over fitting.
(10 Marks)

Module-2

3   a. What is Exploratory Data Analysis? Write about the Philosophy of EDA.   **(10 Marks)**

b. What is Data Science Process? Write about a data scientists role in this process.   **(10 Marks)**

OR

4   a. Write about Real Direct Case Study.   **(10 Marks)**

b. What are the three basic Machine Learning Algorithms? Explain the following in brief Linear Regression, K – nearest neighbor and K - means.
(10 Marks)

Module-3

5   a. Illustrate in detail why Linear regression and KNN are poor choices for filtering Spam.
(10 Marks)

b. Compare Naïve Bayes to KNN. Explain in detail about scraping the web through APIS and other tools.
(10 Marks)

OR

6   a. What is Naïve Bayes? Explain about Bayes Law. Also write about Spam filter for individual word module.
(10 Marks)

b. Write briefly about Data Wrangling and Filtering Spam. Also explain about Spam filter that combines words : Naïve Bayes.
(10 Marks)

Module-4

7   a. Write about the dimensionality problem. Describe Singular Value Decomposition (SVD) and the important properties of SVD.
(10 Marks)

b. Demonstrate Principal Component Analysis (PCA). Write a sample code to build your own recommendation system.
(10 Marks)

OR

8 a. Illustrate about feature extraction and feature selection. Write about Random forests.

(10 Marks)

   b. Describe about User retention. Write about decision trees. Also explain about brainstroming and role of domain expertise in feature generation. (10 Marks)

Module-5

9 a. What is Data Visualisation? Explain its basic principles and write about the tools used.

(10 Marks)

   b. Write about Ethical issues faced in Data Science. Also explain graphs in Social networking.

(10 Marks)

OR

10 a. Discuss about Privacy, Security ethics in Data Science. Also write about Next – Gen Data Scientists. (10 Marks)

   b. Write about partitioning in Graphs , Neighborhood properties in Graphs and Clustering of Graphs. (10 Marks)

* * * * *