

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

Ph.D./M.S.(Engg.) by Research Degree Examination, Aug./Sep. 2023
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 the Venn diagram of data science. (07 Marks)
 b. Describe the reasons for Big Data and Data Science type. (05 Marks)
 c. “The Data Scientists play a major role in Academics and industry”. Justify. (08 Marks)

OR

- 2 a. Explain the concept of Datafication with example. (06 Marks)
 b. “There are varieties of types of data that a Data Scientist has to deal”, explain the statement with examples. (04 Marks)
 c. Explain the following concepts involved in building models :
 (i) Statistical modeling.
 (ii) Probability distributions
 (iii) Fitting and overfitting (10 Marks)

Module-2

- 3 a. With a diagram, describe the data science process. (08 Marks)
 b. Explain the K-means algorithm. List the issues associated with it. (07 Marks)
 c. Discuss the various distance metrics that can be used in KNN. (05 Marks)

OR

- 4 a. Explain how to build and fit a linear regression model along with its evaluation metrics. (10 Marks)
 b. Describe the Exploratory Data Analysis in Data Science. (05 Marks)
 c. Discuss the real direct real estate business and their data strategy. (05 Marks)

Module-3

- 5 a. What is Spam filtering? Explain why linear regression does not fit for filtering spam. (05 Marks)
 b. Explain Naïve Bayes with example. (07 Marks)
 c. List out the differences between Naïve Bayes and KNN algorithms. (08 Marks)

OR

- 6 a. Explain how data can be scrapped from web using API keys and extensions. (10 Marks)
 b. Discuss spam filter for individual words. Explain the same with an example. (10 Marks)

Module-4

- 7 a. What are filters and wrappers? Explain the ways of selecting algorithms and selection criteria for good subset of features. (10 Marks)
 b. Describe Decision tree, entropy and Random forest algorithms. (10 Marks)

OR

- 8 a. Explain singular value decomposition method to overcome dimensionality problem. (08 Marks)
 b. What is privacy? Explain its categories. (06 Marks)
 c. Describe the various problems associated with nearest neighbors. (06 Marks)

Module-5

- 9 a. What are social networks? Explain its characteristics. (10 Marks)
 b. Find the Laplacian matrix for the graph Fig. Q9.

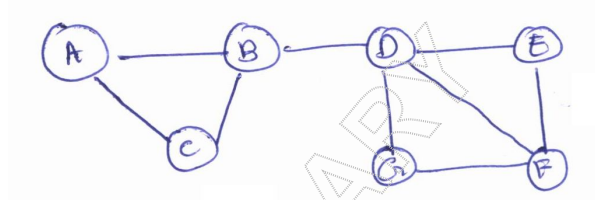


Fig. Q9

(10 Marks)

OR

- 10 a. Write Girvan-Newman Algorithm with explanation on data visualization. (10 Marks)
 b. “Next Generation data scientists should problem solvers” Justify. (10 Marks)
