

USN

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

201CS007

Ph.D./M.S. (Engg.) by Research Degree Examination, December 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 the big data and data science hype. How to get past this hype? (10 Marks)
- b. What is datafication? Why it is required in the current scenario? (05 Marks)
- c. Draw the Venn diagram of data science and discuss in brief. (05 Marks)

OR

- 2 a. Explain the following with examples
i) Statistical inference (10 Marks)
ii) Populations and samples (05 Marks)
- b. What is statistical modeling? How do you build a model? (05 Marks)
- c. Explain probability distributions with examples. (05 Marks)

Module-2

- 3 a. Discuss the data science process with a neat sketch. (10 Marks)
- b. Explain the basics and philosophy of Exploratory Data Analysis (EDA). (05 Marks)
- c. Write a short note on Real Direct Online real estate firm. (05 Marks)

OR

- 4 a. Discuss the different similarity or distance metrics. Explain the steps involved in K-Nearest Neighbor algorithm with an example. (10 Marks)
- b. Explain the linear regression with an example in brief. (05 Marks)
- c. Write a short note on K-Means algorithm. (05 Marks)

Module-3

- 5 a. Why linear regression and K-NN algorithms are poor choice for filtering spam? Explain with examples. (08 Marks)
- b. Explain the process for hand-written digit recognition using K-NN.. (07 Marks)
- c. Discuss the spam in brief with its features. (05 Marks)

OR

- 6 a. Explain the spam filter for individual words and the spam filter that combines words using Naïve Bayes algorithm. (10 Marks)
- b. Explain the Bayes law with an example. (05 Marks)
- c. Write a note on APIS and other tools for scrapping the web. (05 Marks)

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and/or equations written eg. 42+8 = 50, will be treated as malpractice.

Module-4

- 7 a. Explain the wrappers in detail with respect to features selection. (08 Marks)
b. Explain the decision tree with an example. (07 Marks)
c. Explain the feature selection with an example of user retention. (05 Marks)

OR

- 8 a. Explain the problems with nearest neighbors. (08 Marks)
b. Discuss the singular value decomposition along with properties. (07 Marks)
c. Explain the real-world recommendation engine using Bipartite graph. (05 Marks)

Module-5

- 9 a. Discuss the social network as graphs with an example. Also explain the different types of social networks. (10 Marks)
b. Explain the partitioning of graphs in brief with an example. (05 Marks)
c. Write a short note on Clustering of graphs based on social networks. (05 Marks)

OR

- 10 a. Explain neighborhood properties of graph in brief. (10 Marks)
b. Explain the basics of data visualization in brief. (05 Marks)
c. Discuss in brief about next generation data scientists. (05 Marks)

* * * * *