



Sixth Semester B.E./B.Tech. Degree Examination, Dec.2024/Jan.2025
Data Science and Visualization

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 the Venn diagram of data science. (08 Marks)
- b. Explain the concept of datafication with an example. (04 Marks)
- c. Explain the following concepts with examples:
 - i) Statistical Interference
 - ii) Population
 - iii) Samples
 - iv) Type of data. (08 Marks)

OR

- 2 a. Explain the role of a data scientist in the context of big data. How does the skill set of a data scientist different from traditional data analysts? (08 Marks)
- b. Explain the process of fitting a model to data. How do you evaluate the goodness of fit and what metrics are commonly used in evaluation? (08 Marks)
- c. List out the common probability distributions with a brief explanation and examples with respect to usage in big-data. (04 Marks)

Module-2

- 3 a. Describe the data science process with neat diagram. (06 Marks)
- b. Explain the K-means algorithm. List the issues associated with it. (06 Marks)
- c. Explain the basic tools of exploratory data analysis with proper example. (08 Marks)

OR

- 4 a. Explain the concept of model evaluation in machine learning. What metrics are commonly used to evaluate the performance of linear regression, K-NN and K-means model? (10 Marks)
- b. Discuss the importance of visualizing data in exploratory data analysis. What are the common tools and techniques used in EDA? (10 Marks)

Module-3

- 5 a. Define feature generation. Explain in detail how information can be categorized during feature generation. (08 Marks)
- b. Explain and construct decision tree with a suitable example. (06 Marks)
- c. What is dimensionality problem? Explain real world recommendation engine with neat diagram. (06 Marks)

OR

- 6 a. Explain singular value decomposition with a suitable illustration. (06 Marks)
 b. Explain the three primary methods used while building the regression model. (06 Marks)
 c. Write a short note on:
 i) Under fitting
 ii) Filters
 iii) Over fitting
 iv) Wrappers. (08 Marks)

Module-4

- 7 a. What is data wrangling with a neat diagram, explain the steps involved in data wrangling process. (08 Marks)
 b. Describe data visualization. Explain why data visualization is important. (06 Marks)
 c. Explain line chart, bar chart and radar chart with suitable examples. (06 Marks)

OR

- 8 a. Write a note on following:
 i) Scatter plot
 ii) Correlogram
 iii) Bubble plot
 iv) Heat map (08 Marks)
 b. Explain the stacked bar chart with an example. Explain the design practices to be followed while plotting stacked bar. (06 Marks)
 c. Explain the different map plots with a suitable example. (06 Marks)

Module-5

- 9 a. With a neat diagram, explain the components involved in anatomy of a matplotlib figure. (08 Marks)
 b. With proper illustration, explain how the bar chart and pie chart can be implemented using matplotlib. (06 Marks)
 c. Write a short notes on different customization options available while drawing any plot. (06 Marks)

OR

- 10 a. With a proper assumptions, explain how to use a scatter plot to visualize correlation between various animals. (06 Marks)
 b. Explain stacked area charts with a proper example plot the area chart using matplotlib. (06 Marks)
 c. Explain the basic operations that can be performed on image using matplotlib. (08 Marks)

* * * * *