

Reg. No. :

Question Paper Code : 10239

M.E./M.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2023.

First Semester

Big Data Analytics

BD 4151 — FOUNDATIONS OF DATA SCIENCE

(Common to: M.E. Computer Science and Engineering/M.E. Software Engineering)

(Regulations 2021)

Time : Three hours

Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. What are the requirements of Training, Testing and Validation?
2. Illustrate Bias-Variance trade-off.
3. Distinguish between Linear and Logistic regression
4. What are R-factors?
5. Distinguish between Poisson and Normal distribution.
6. Differentiate between data frames and lists.
7. What is the difference between HDFC block and InputSplit?
8. What are the main components of Map Reduce job?
9. What are the different graphical parameters in R?
10. What are Matrix plots?

PART B — (5 × 13 = 65 marks)

11. (a) (i) Explain the standard errors of regression coefficients. (7)
(ii) Explain betweenness and centrality in detail. (6)

Or

- (b) (i) Explain various stages in data science process. (7)
(ii) Explain overfitting and underfitting in detail with an example. (6)

12. (a) (i) Describe the statement "correlation is not causation" with an example in detail. (6)
(ii) Detail the any unsupervised method with an example. (7)

Or

- (b) (i) Explain bottom-up hierarchical clustering approach with an example in detail. (7)
(ii) Discuss the nearest neighbor model in detail. (7)

13. (a) (i) Describe dispersion and variance and write R code to compute the variance. (6)
(ii) Write a function to create a matrix given its shape and a function for generating its elements. Then use the function to generate to 5 x 5 identity matrix. (7)

Or

- (b) (i) Explain standard deviation and interquartile range and write R code to compute standard deviation and interquartile range. (6)
(ii) Describe how data can be manipulated by considering any example of your own. (7)

116

14. (a) (i) How to set mappers and reducers for Hadoop jobs? Illustrate with an example. (6)

(ii) Explain JobConf in MapReduce. (7)

Or

(b) (i) Explain Shuffling and Sorting in MapReduce. (7)

(ii) Detail the main components of MapReduce job. (6)

15. (a) (i) How will you display multi variate data? Explain with an example. (6)

(ii) What is the requirement of Graphical Analysis? Explain. (7)

Or

(b) What are the various types of graphs drawn using R? Explain. (13)

binils.com

PART C (1 × 15 = 15 marks)

16. (a) (i) Illustrate minibatch and stochastic gradient descent in detail. (7)

(ii) Write a R program to extract the date and time of a GitHub repository created and also the language of the last five repositories. (8)

Or

(b) (i) Write a R program to apply the logistic regression model and check for goodness of the fit. (7)

(ii) Describe theoretically the Naïve Bayes theorem to model a sophisticated spam filter. (8)