C **H1064** **Pages: 2**

Reg No.:_______________ Name:__________________________

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MAY 2019

Course Code: CS466

Course Name: DATA SCIENCE

Max. Marks: 100 Duration: 3 Hours

PART A

Answer all questions, each carries 4 marks. Marks

1 Categorise the different roles associated with a data analysis project. (4)

2 A retail store is having a database stored as spreadsheet documents and text files. Design suitable procedure for accessing the files for data analysis. (4)

3 List some similarity measures used for clustering. (4)

4 Create an array with 4 rows and 5 columns and with elements from 1 to 20. Also print the array (use R) (4)

5 Why box plot is important? Explain how to create a box plot in Python (4)

6 Illustrate add_subplot(2 2 1) in Python (4)

7 What are the advantages of Hadoop? (4)

8 Which are the nodes in HDFS, and what do they contain/maintain? (4)

9 What is the purpose of knitr? (4)

10 How to create a matrix plot in R? (4)

PART B

Answer any two full questions, each carries 9 marks.

11 a) Illustrate with an example different stages of data science project. (9)

12 a) List various real life problems that can be mapped to machine learning techniques. Deduce suitable models in solving them. (9)

13 a) Write a note on logistic regression. (3)

b) Illustrate with a data analysis example, the use of linear regression methods in solving the problem. (6)

PART C

Answer any two full questions, each carries 9 marks.

14 a) Explain data frames in R. Illustrate attach(), detach() and search() functions in R (6)

b) Write the function in R to build a linear model with an example (3)

15 a) Which are the probabilistic distribution functions available in R? Explain any 4 (4)

functions.

b) Discuss statistical models in R. Write two examples. (5)

16 a) Discuss and Illustrate user-based collaborative filtering in Python based on Euclidean distance score. (9)

PART D

Answer any two full questions, each carries 12 marks.

17 a) With a neat diagram, explain MapReduce architecture. (4)

b) Give an overview of the execution of MapReduce program with a neat diagram (8)

18 a) How to cope with node failures in Hadoop MapReduce? (6)

b) What is the difference between mfrow=c(3,2) and mfcol=c(3,2).Explain its operation with a figure. (6)

19 a) What should be the contents of an effective presentation? (12)

****