Paper / Subject Code: 94728 / Information Technology : Data Science (R.2019)

DATA SCIENCE (Time: $2\frac{1}{2}$ hours)

[Total Marks: 60]

N. B.: (1) **All** questions are **compulsory**.
  (2) Make **suitable assumptions** wherever necessary and **state the assumptions** made.
  (3) Answers to the **same question** must be **written together**.
  (4) Numbers to the **right** indicate **marks**.
  (5) Draw **neat labeled diagrams** wherever **necessary**.
  (6) Use of **Non-programmable** calculator is **allowed**.

1. Attempt *any two* of the following: 12
   a. Explain any five data processing tools in data science technology.
   b. Discuss the Cross-Industry Standard Process for Data Mining (CRISP-DM).
   c. Explain the maintenance and processing utilities in the utility layer
   d. Describe the functional requirements in the business layer of the data science framework

2. Attempt *any two* of the following: 12
   a. Explain the five fundamental steps that form the core of the data science process.
   b. Write a short note on the six super steps for processing the data.
   c. Explain in brief the indicators used in the Audit sub layer.
   d. Explain the operational management layer.

3. Attempt *any two* of the following: 12
   a. Explain Assess super step. What are the different ways to handle errors in the Assess super step.
   b. State and explain the six data quality dimensions used in analysis of data
   c. What are the different ways of treating missing values in data using pandas package? Explain with example.
   d. Explain node, edge and directed acyclic graph.

4. Attempt *any two* of the following: 12
   a. Explain the Time-Person-Object-Location-Event (T-P-O-L-E) design principle.
   b. Explain hypothesis testing, t-test and chi-square test with respect to data science.
   c. Explain over fitting and under fitting. Discuss the common fitting issues.
   d. What is feature engineering? What are the common feature extraction techniques?

5. Attempt *any two* of the following: 12
   a. Explain univariate analysis, bivariate analysis and multivariate analysis.
   b. What is clustering? Explain the different clustering techniques
   c. What are random forests? Explain with examples.
   d. Explain the organize super step.

_____

17376   **Page 1 of 1**