Information Technology : Paper I - Big Data Analytics (R2020)

(2½ Hours)   **[Total Marks: 75]**

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** calculators is **allowed**.

Q1. Attempt the three of the following.   **15 Marks**

A. Explain the definition of Big Data with a suitable diagram.

B. Write a short note on Current Analytical Architecture with a suitable diagram.

C. Explain Data Science and its expertise with a suitable diagram.

D. Explain the ACID property in RDBMS.

E. Explain the challenges of Big Data.

Q2. Attempt three of the following   **15 marks**

A. Explain Apriori Algorithm.

B. Explain N-Fold Cross-Validation.

C. Write a note on Reasons to choose and cautions.

D. Explain validation and testing.

E. Write a note on the Applications of Association Rules.

Q3. Attempt three of the following.   **15 marks**

A. Write a short note on Decision Trees.

B. Write a short note on NAIVE BAYES.

C. Write a short note on Text Analysis Steps.

D. Explain Time Series Analysis.

E. How to categorize documents by topics? Explain in detail.

Q4. Attempt three of the following. 15 marks

    A. Explain HADOOP as Big Data Operating System.

    B. Write a short note on Hadoop Distributed File System.

    C. Explain Resilient Distributed Datasets.

    D. Explain Big Data Pipeline with suitable diagram.

    E. Explain the concepts of Advanced Map Reduce with suitable diagram (any two).

Q5. Attempt three of the following. 15 marks

    A. Explain Indexing and Inverted index.

    B. Write a short note on Data Ingestion.

    C. Explain Design Patterns.

    D. Explain the concept of filtering.

    E. Explain DataFrames with suitable diagram.

_____