

(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

- Enumerate and explain the challenges with bi data.
- Explain Classification of Analytics with respect to Second School of thought.
- What are the requirements of technologies to meet challenges of big data? Explain the Responsibilities of Data Scientists.
- What are the main phases of the Data Analytics? Explain using suitable diagram.

2. Attempt any two of the following:

12

- What is K-means clustering? Describe the steps to find k clusters using k-means algorithm.
- What is the role of support in Apriori algorithm? Explain how the Apriori property works with a neat diagram
- What is Linear regression? Explain in detail.
- Explain any three applications of logistic regression model applied to situations in Government and the Private sector.

3. Attempt any two of the following:

12

- Explain Autoregressive Integrated Moving Average Model.
- Explain a probabilistic classification method based on Naive Bayes' theorem.
- What is the critical problem in using Term frequency? How can it be fixed?
- What are the components of time series? Explain Box Jenkins Methodology.

4. Attempt any two of the following:

12

- How to refactor the data science pipeline into an iterative model? Explain all its phases with a neat diagram.
- How Map Reduce can be implemented on a cluster? Explain with neat diagram.
- Write a short note on job chaining with a neat diagram.
- What is Apache spark? Explain its different components.

5. Attempt any two of the following:

12

- Explain with example Projection in context of Pig
- Write the entire procedure with appropriate commands for importing data from MySQL to Hive.
- Explain the concept of structured queries with HIVE.
- Write a short note on DATA INGESTION.