

**Time: 2 Hrs**

**[Total Marks: 60]**

**N.B**

- 1) All questions are compulsory
- 2) Figures to right indicate marks
- 3) Illustration, depth answers and diagrams will be appreciated.
- 4) Mixing of sub-questions are not allowed.
- 5) Each question carries 5 Marks.

**Q 1. Attempt any two of the following** **[12]**

- a) Differentiate between structured and unstructured data.
- b) What is Re-sampling? Explain the two commonly used re-sampling methods.
- c) List and explain the components of time series data.
- d) Draw and explain the architecture of fuzzy logic based software systems.

**Q 2. Attempt any two of the following** **[12]**

- a) Explain the term Distributed File System.
- b) What is Map-Reduce? Explain operations perform by mapper and reducer class.
- c) Explain Matrix-Vector Multiplication by MapReduce with suitable example.
- d) Explain Union, Intersection and Difference concept by MapReduce.

**Q 3. Attempt any two of the following** **[12]**

- a) Explain applications of near-neighbour search.
- b) Explain the term: Minhashing
- c) Write down the definition of a Distance Measure. Explain Euclidean Distances.
- d) Explain the three conditions that we need for a family of functions.

**Q 4. Attempt any two of the following** **[12]**

- a) Explain the Data-Stream-Management System with diagram.
- b) Explain the sources of stream data.
- c) How to perform Analysis of Bloom Filtering?
- d) What is Stream Queries? Explain two ways that queries get asked about streams.

**Q 5. Attempt any two of the following** **[12]**

- a) Explain the Hill climbing algorithm
- b) Explain the steps of execution for word count algorithm with an example.
- c) What is Apache Hadoop? What are the characteristics of a Distributed File System?
- d) Compute the Jaccard bag similarity of each pair of the following three bags: {1, 1, 1, 2}, {1, 1, 2, 2, 3}, and {1, 2, 3, 4}.

\*\*\*\*\*