

Reg. No. :

--	--	--	--	--	--	--	--	--	--

**Question Paper Code: 31889**

B.E. / B.Tech. DEGREE EXAMINATION, MAY 2017

Elective

Information Technology

01UIT909 - DATA ANALYTICS

(Regulation 2013)

Duration: Three hours

Maximum: 100 Marks

Answer ALL Questions

PART A - (10 x 2 = 20 Marks)

1. State the significance of resampling.
2. State the challenges of conventional systems in handling big data.
3. Differentiate between examples and instances from rule induction.
4. What is Language bias?
5. State the definition of data stream.
6. State the different filtering schemes used in streams.
7. What is the purpose of Apriori Algorithm?
8. Compare the Agglomerative and Divisive Hierarchical Clustering.
9. State the general form of Map and Reduce function in Hadoop Map Reduce.
10. What is the use of Hive Metastore?

PART - B (5 x 16 = 80 Marks)

11. (a) Analyze sampling distributions and explain types sampling. (16)

Or

(b) Discuss about statistical inference in detail. (16)

12. (a) What are the various inference in Bayesian network? Explain in detail. (16)

Or

(b) (i) Write a short notes on fuzzy decision trees. (8)

(ii) Explain the fuzzy decision tree algorithm. (8)

13. (a) Sketch the stream data model diagram and explain in detail. (16)

Or

(b) Describe in detail about estimating moments. (16)

14. (a) (i) Write a short note on hierarchical methods in the clustering. (6)

(ii) List the hierarchical methods in the clustering techniques and explain its type. (10)

Or

(b) (i) Discuss about Clustering High-Dimensional Data. (8)

(ii) Distinguish between CLIQUE and PROCLUS method. (8)

15. (a) Illustrate the basics of Hadoop Distributed File System with its architecture. (16)

Or

(b) Evaluate the techniques of visual data analysis. (16)