

Reg. No.

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

B.E. / B.TECH. DEGREE EXAMINATIONS, DEC 2019

Seventh Semester

CS16004 – DATA ANALYTICS*(Computer Science and Engineering)***(Regulation 2016)****Time: Three Hours****Maximum : 100 Marks**

Answer ALL questions

PART A - (10 X 2 = 20 Marks)

	CO	RBT
1. Distinguish analysis versus reporting.	1	AN
2. Why does one choose analytical system over conventional system?	1	E
3. What is meant by a support-vector machine?	2	R
4. What is main idea of decision trees in fuzzy logic?	2	R
5. Define decay window.	3	R
6. Compute the surprise number (second moment) for the stream 3, 1, 4, 1, 3, 4, 2, 1, 2. What is the third moment of this stream?	3	AP
7. Compare and contrast the Multistage and Multi-Hash algorithm.	4	U
8. What is meant by curse of dimensionality?	4	U
9. Examine the features of Hive.	5	AN
10. Illustrate the advantages of MapR.	5	AP

PART B - (5 X16 = 80 Marks)

11. (a) (i) Discuss web data in detail. How web data is collected, analysed and used for understanding user behaviours? **(8)** 1 U
- (ii) Discuss in detail the types of Analytic sandboxes with neat diagrams. **(8)** 1 U

(OR)

- (b) Describe in detail the Statistical inference techniques for data analysis. **(16)** 1 U

12. (a) Investigate how Bayesian modeling is used in analyzing data. (16) 2 AN
- (OR)**
- (b) Examine how Rule learning is regarded as a search problem. (16) 2 AN
13. (a) (i) Describe the Stream data model architecture with a neat diagram. (8) 3 U
- (ii) Examine the Flajolet-Martin Algorithm for counting distinct elements in a stream. Apply FM algorithm on the stream 1,3,2,1,2,3,4,3,1,2,3,1. Use $h(x)=(6x+1) \bmod 5$ as the hash function and length of the binary string as 5. (8) 3 AN
- (OR)**
- (b) (i) How Alon-Matias-Szegedy Algorithm is used for estimating moments? Explain with an example. (8) 3 AN
- (ii) Assuming a real time stock market situation, bring out the various ideas used in prediction analysis. (8) 3 C
14. (a) (i) Explain Apriori algorithm and with an example for discovering frequent itemsets. (8) 4 U
- (ii) Explain SON and Toivonen's algorithms in detail. Explain how they can be used to identify frequent itemsets in limited number of passes. (8) 4 U
- (OR)**
- (b) (i) Summarize the hierarchical clustering in Euclidean and non-Euclidean Spaces with its efficiency. (8) 4 U
- (ii) Examine how the data is processed in BFR Algorithm. (8) 4 AN
15. (a) What is Hadoop? Explain HDFS in detail. (16) 5 U
- (OR)**
- (b) (i) Write short notes on NoSQL Databases and its types. (8) 5 R
- (ii) Explain the Visual data analysis techniques in detail. (8) 5 U