

25/5  
AN

Reg. No. :

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

**Question Paper Code : 72045**

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

Seventh Semester

Information Technology

IT 6006 — DATA ANALYTICS

(Common to Computer Science and Engineering)

(Regulations 2013)

Time : Three hours

Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. Write down the characteristics of big data.
2. Outline the need for resampling.
3. List the benefits of regression modeling.
4. What is competitive learning?
5. Outline the need for sampling data in a stream.
6. What is sentiment analysis?
7. What is market basket analysis?
8. Define clustering.
9. What is NoSQL database?
10. Outline the need for a distributed file system.

PART B — (5 × 16 = 80 marks)

11. (a) What is data analytics? Appraise the features of data analytic tools. (16)

Or

- (b) Appraise with any two examples the role of statistics in data analytics. (16)



12. (a) Elaborate the role of Bayesian networks in data analysis with an example. (16)

Or

- (b) (i) Appraise the need for principal component analysis (PCA) with an example. (12)
- (ii) What is a fuzzy decision tree? Explain with an example. (4)
13. (a) (i) With a neat sketch, explain the architecture of a data-stream-management system. (8)
- (ii) Outline the algorithm used for counting distinct elements in a data stream. (8)

Or

- (b) Explain with an example a real time analytics platform (RTAP). (16)
14. (a) Explain the Apriori algorithm for mining frequent item sets with an example. (16)

Or

- (b) Explain the k-means clustering algorithm with an example. (16)
15. (a) (i) Describe the mapreduce frame work. (8)
- (ii) Outline the features of Hadoop distributed file system. (8)

Or

- (b) Discuss any two visual data analysis techniques with an example. (16)