

**APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY**  
**SECOND SEMESTER M.TECH DEGREE EXAMINATION, MAY 2016**

**Computer Science and Engineering**  
**(Computer science and Systems Engineering)**  
**04 CS 6434—Big Data Management and Analytics**

Max. Marks: 60

Duration: 3 Hours

**PART A**

*Answer All Questions*

*Each question carries 3 marks*

1. Write any 3 sources of big data.
2. Write the features of Hadoop Distributed File system.
3. Write about any two failures possible in a hadoop cluster.
4. What are the two styles for distributing data?
5. Write notes on 2 basic types of machine learning algorithms.
6. Write a note on the features of R.
7. What is clustering? Why it is known as unsupervised machine learning?
8. Write an example for the implementation of any machine learning technique in R.

**PART B**

*Each question carries 6 marks*

9. Explain the difference between predictive and descriptive analytics.  
OR
10. What is big data? How it is generated? Write about any two tools to process big data.
11. Describe the working of Hadoop Distributed File system.  
OR
12. Explain map reduce programming model with an example.
13. Explain different job scheduling techniques in hadoop.  
OR
14. With a neat diagram describe the working of hadoop map reduce cluster.
15. Explain graph databases with an example.  
OR
16. Explain aggregate data model with an example.
17. Briefly explain the following classification techniques.  
(i) Support Vector Machine (ii) Naïve Bayes  
OR
18. Explain any unsupervised machine learning technique.
19. With an example explain the implementation of any machine learning technique in python.  
OR
20. Explain the uses of machine learning