

Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MAY 2019

Course Code: IT404

Course Name: Data Analytics

Max. Marks: 100

Duration: 3 Hours

PART A

Answer any two full questions, each carries 15 marks.

Marks

- 1 a) Compare and contrast analysis and reporting in data analytics with suitable example. (7)
- b) Consider the population of size $N = 6$, displayed in the table given below. 'X' is the population variable. A random sample of size $n = 2$ is drawn with replacement from this population. Show the distributions of the sample and the sampling distribution of the sample statistics \bar{x} and S^2 . (8)

unit	1	2	3	4	5	6
X	1	1	2	2	2	3

- 2 a) What is a neural network? How can it be used in analytics? (8)
- b) Explain decision tree with an example (7)
- 3 a) What is the need for sampling? Explain about different sampling methods. (8)
- b) Explain principal component analysis (7)

PART B

Answer any two full questions, each carries 15 marks.

- 4 a) A database has five transactions. Let $min_sup = 60\%$ and $min_conf = 80\%$. (10)

TID	Items_bought
T100	{M,O,N,K,E,Y}
T200	{D,O,N,K,E,Y}
T300	{M,A,K,E}
T400	{M,U,C,K,Y}
T500	{C,O,O,K,I,E}

- (i) Find all frequent itemsets using Apriori algorithm.
- (ii) List all of the *strong* association rules (with support s and confidence c)
- b) List the advantages and disadvantages of K Means clustering. (5)
- 5 a) What are the various stages in big data analytics life cycle? Illustrate with a figure, explaining each of them. (9)
- b) What are the characteristics of Big data? (6)
- 6 a) Define sequential rule mining. Give example (5)
- b) Discuss the trends in big data generation and acquisition (10)

PART C

Answer any two full questions, each carries 20 marks.

- 7 a) List six R functions which are used in descriptive statistics. (12)
b) What is the significance of scatter plot matrix? (8)
- 8 a) Illustrate and explain the concept of Map Reduce framework (10)
b) With an example, explain the term *social media analytics*. (10)
- 9 a) Write R function to check whether the given number is prime or not? (10)
b) What is HDFS? How does it handle Big Data? (10)
