

Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Eighth Semester B.Tech. Degree Examinations, September 2020

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) Describe any two parallel processing technologies that improve analytic scalability. (8)
- b) What is regression analysis? Give an example for a linear regression model and explain why they are widely used in statistical analysis? (7)
- 2 a) What is the significance of re-sampling techniques? Illustrate any two resampling techniques. (8)
- b) What is a perceptron? Give the procedure for back propagation learning algorithm. (7)
- 3 a) Explain the frequentist approach in statistical inference. (7)
- b) Why is dimensionality reduction important and explain the multivariate technique used in dimensionality reduction? (8)

PART B*Answer any two full questions, each carries 15 marks.*

- 4 a) Briefly describe and give examples of each of the following approaches to clustering: (i) partitioning methods, (ii) hierarchical methods. (8)
- b) What is big data? What are the challenges of big data? (7)
- 5 a) Give the Apriori property? How is the Apriori property used for finding frequent itemsets? (7)
- b) Discuss big data in terms of four dimensions volume, velocity, variety and veracity (8)
- 6 a) What is sequential pattern mining? Give one application where sequential pattern mining is used. (7)
- b) Explain any two organizational applications of big data analytics. (8)

PART C

Answer any two full questions, each carries 20 marks.

- 7 a) Explain the following data types using R command and give examples (6)
- i. Vector
 - ii. Data frames
 - iii. Contingency tables
- b) What is p-value? What is the significance of p-value in statistical tests? (4)
- c) Explain the working of recommender system with an example. (10)
- 8 a) What is dirty data? How is dirty data detected in the data exploration phase? (7)
- b) What is ANOVA? (3)
- c) How is fraud detection done in big data analytics? (10)
- 9 a) What is exploratory data analysis? Illustrate any three visualization techniques using R functions for finding the relationship among multiple variables. (10)
- b) How is MapReduce implemented in Apache Hadoop? (10)
