

Register No:

Name:

SAINTGITS COLLEGE OF ENGINEERING (AUTONOMOUS)

(AFFILIATED TO APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY, THIRUVANANTHAPURAM)

SIXTH SEMESTER B.TECH DEGREE EXAMINATION(R,S), MAY 2024**Computer Science and Engineering****(2020 SCHEME)****Course Code : 20CST322****Course Name : Data Analytics****Max. Marks : 100****Duration:3 Hours**

Scientific calculator and statistical table is allowed in the examination hall.

PART A*(Answer all questions. Each question carries 3 marks)*

1. A study is conducted to investigate the association between gender and voting preference in a local election. A sample of 300 voters is surveyed, and their gender (male/female) and voting preference (Candidate A/Candidate B) are recorded. The data is summarized in the following contingency table:

	Candidate A	Candidate B
Male	70	30
Female	60	140

Calculate the observed frequencies and expected frequencies for each cell in the contingency table.

2. Explain the different measures of Central Tendency.
3. How outlier detection and treatment can be done in analytics?
4. What are the different stages of Life cycle of Data Analytics?
5. What is partitional algorithm?
6. Explain Classification by Backpropagation.
7. What is credit risk modelling?
8. Define unstructured datastructures.
9. Which R command(s) would you use to remove null values from a dataset?
10. Explain Data analytics tools with an example.

PART B*(Answer one full question from each module, each question carries 14 marks)***MODULE I**

11. Explain Point Estimation and Interval Estimation. 14

OR

12. Discuss measures of central tendency (mean and median) and measures of dispersion (range and variance) in the context of ordinal and continuous variables. 14

MODULE II

13. Discuss the key considerations and criteria for selecting an appropriate analytical model for a given problem. 14

OR

14. With neat sketches explain Analytics Process Model. 14

MODULE III

15. Explain Naive Bayes classifier. Using relevant examples, explain different classification techniques. 14

OR

16. Explain K-means algorithm considering the challenges in hierarchial algorithm. 14

MODULE IV

17. Describe the architecture of HDFS and its features. 14
- OR**
18. Explain the Hadoop Ecosystem. What are the functions of HDFS and Map Reduce? 14
- MODULE V**
19. Explain about the attributes and Datatypes of R programming with examples. 14
- OR**
20. Explain in detail the descriptive analytics. 14
