

R18

Code No: 157HM

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

B. Tech IV Year I Semester Examinations, December-2023/January-2024

PREDICTIVE ANALYTICS

(Computer Science and Engineering (Data Science))

Time: 3 Hours

Max. Marks: 75

Note: i) Question paper consists of Part A, Part B.

ii) Part A is compulsory, which carries 25 marks. In Part A, Answer all questions.

iii) In Part B, Answer any one question from each unit. Each question carries 10 marks and may have a, b as sub questions.

PART – A

(25 Marks)

- a) Give examples for quantitative variables. [2]
- b) Illustrate the use of F-statistic measure. [3]
- c) Define training error. [2]
- d) Give Akaike information criterion. [3]
- e) What is meant by exponential loss? [2]
- f) What is the need of classification tree? [3]
- g) What is cross-entropy? [2]
- h) What is the effect of large k-value in kNN? [3]
- i) Define support of association rule. [2]
- j) List the steps in k-means clustering algorithm. [3]

PART – B

(50 Marks)

2. Discuss the drawbacks of linear regression and explain multiple regression with an example. [10]

OR

3. Make a comparison of ridge regression and logistic regression. [10]
4. Describe the bias-variance decomposition in detail. [10]

OR

5. How to choose k-value for k-fold validation? Explain your answer with suitable examples. [10]

6. Explain local scoring algorithm for the additive logistic regression model. [10]

OR

7. Illustrate gradient boosting on a multi classification problem with necessary data. [10]

8. Summarize the important issues in training neural network. [10]

OR

9. How does SVM handle inseparable data? Explain with examples. [10]

10. What is the goal of clustering? What are the applications of cluster analysis? [10]

OR

11. "Random forest is a substantial modification of bagging that builds a large collection of de-correlated trees." Substantiate this statement with appropriate example. [10]

---ooOoo---