Total No. of Questions – [06] Total No. of Printed Pages: 2

G.R. No.	

DECEMBER 2021 - ENDSEM EXAM

T. Y. B. TECH. (COMPUTER ENGINEERING) (SEMESTER - I)

COURSE NAME: DATA SCIENCE

COURSE CODE: CSUA31182

(PATTERN 2018)

Time: [1 Hr] [Max. Marks: 30]

Instructions to candidates:

1) **Answer Q.1 OR Q.2, Q.3 OR Q.4, Q.5 OR Q.6.**
2) **Figures to the right indicate full marks.**
3) **Use of scientific calculator is allowed.**
4) **Use suitable data where ever required.**

Q.1) a) Examine Logistic regression model. [4]

b) Decide the two regression equations of X on Y and Y on X from the data given below, taking deviations from a actual means of X and Y. Estimate the likely demand when the price is Rs.20. [6]

Price(Rs)	10	12	13	12	16	15
Amount demanded	40	38	43	45	37	43

OR

Q.2) a) Examine various types of linear regression. [4]

b) Examine the different assumptions of Linear Regression Model. [6]

Q.3) a) Examine Naïve Bayes theorem for classification. [4]

b) Inspect the significance of Information Gain. Explain the mathematical formulation associated with it. [6]

OR

Q.4) a) Analyze attribute selection measures used by the ID3 algorithm to construct a Decision Tree. [4]

b) Discuss the trade-offs between the different types of classification algorithms? How do you choose the best one? [6]

Q.5) a) Estimate the use of pivot tables in handling granularity in visual representation of data. [4]

b) Identify Tufte's design principles, explain with examples. [6]

OR

Q.6) a) Elaborate the effects of poor data cleaning and wrangling on quality of visual representation. [4]

b) Discuss the role of image processing techniques in medical visualization. Give one or more examples to illustrate your explanation. [6]