1T01876 - T.E. Computer Science and Engineering (Artificial Intelligence and Machine Learning) (Choice Based) (R-19-20 'C' Scheme)SEMESTER - VI / 37471 - Data Analytics and Visualization
QP CODE: 10039481 DATE: 11/12/2023

(3 Hours) **(Total Marks: 80)**

N.B.: 1. Question No. 1 is compulsory.
2. Answer any three out of the remaining questions.
3. Assume suitable data if necessary.
4. Figures to the right indicate full marks.

Q1. Attempt the following (any 4): (20)

a. Why is data analytics lifecycle essential?

b. The regression lines of a sample are $x + 6y = 6$ and $3x + 2y = 10$.
Find (i) sample means $\bar{x}$ and $\bar{y}$.
(ii) coefficient of correlation between x and y

c. Differentiate between linear regression and logistic regression.

d. What is Pandas? State and explain key features of Pandas.

e. Explain term frequency (TF), document frequency (DF), and inverse document frequency (IDF).

Q2. Attempt the following:

a. Explain the data analytics lifecycle. (10)

b. Find two lines of regression from the following data: (10)

Age of husband (x)	25	22	28	26	35	20	22	40	20	18
Age of wife (y)	18	15	20	17	22	14	16	21	15	14

Estimate (i) the age of husband when the age of wife is 19 and (ii) the age of wife when the age of the husband is 30.

Attempt the following:

a. Explain Box-Jenkins intervention analysis. (10)

b. What is text mining? Enlist and explain the seven practice areas of text analytics. (10)

Q4. Attempt the following:

a. Explain different types of data visualizations in R programming language. (10)

b. Fit a regression equation to estimate β_0, β_1, and β_2 to the following data of a transport company on the weights of 6 shipments, the distances they were moved and the damage of the goods that was incurred. (10)

Weight X_1 (1000 kg)	4.0	3.0	1.6	1.2	3.4	4.8
Distance X_2 (100 km)	1.5	2.2	1.0	2.0	0.8	1.6
Damage Y (Rs.)	160	112	69	90	123	186

Estimate the damage when a shipment of 3700 kg is moved to a distance of 260 km.

Q5. Attempt the following:

a. From the following results, obtain two regression equations and estimate the yield when the rainfall is 29 cm and the rainfall when the yield is 600 kg. (10)

	Yield in Kg.	Rainfall in cm
Mean	508.4	26.7
SD	36.8	4.6
Coefficient of Correlation	0.52	

b. What is stepwise regression? State and explain different types of stepwise regression. (10)

Q6. Write short notes on (any 2): (20)

a. Time series analysis

b. Exploratory data analysis

c. Regression plot

d. Generalized linear model (GLM)

****