

14/05/2025 TE CSE-AIML SEM-VI C SCHEME DAAV QP CODE: 10081183

Time: 3 hours

Max. Marks: 80

N.B. (1) Question one is Compulsory.**(2) Attempt any 3 questions out of the remaining.****(3) Assume suitable data if required.****Q1. Attempt the following (any 4):****[20]**

- What are the application and use cases for text mining?
- What is Seaborn Library? State and explain key features of seaborn.
- Explain TFIDF with an example.
- Explain Logistic Response Function.
- List and explain the various key roles for a successful analytics.

Q2. Attempt the following**[20]**

- Calculate the linear regression using least square method for the given dataset.

Independent Variable (X)	Dependent Variable (Y)
1	2
2	4
3	5
4	4
5	5

- List and explain the different type of data visualization used in R.

Q3. Attempt the following:**[20]**

- Explain the data analytics life cycle.
- Explain AR and MA model in detail.

Q4. Answer the following**[20]**

- 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.

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)	160	112	69	90	123	186

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

- Describe in detail about building and evaluating an ARIMA Model.

Q5. Attempt the following

[20]

- a) List and explain the steps in text analysis.
- b) Describe in detail about removing the dirty data using R.

Q.6 Write short notes on:

[20]

- a) Box Plot
 - b) Data types in R
 - c) Box-Jenkins Methodology
 - d) Fitted value and residuals in Linear Regression.
-