

SCHOOL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE			DEPARTMENT OF COMPUTER SCIENCE ENGINEERING														
Program Name: B. Tech		Assignment Type: Lab		Academic Year: 2025-26													
Course Coordinator Name		Dr.Vairachilai Shenbagavel															
Instructor(s) Name		Srinivas Komakula															
Course Code	23CA201SE402	Course Title	Explainable AI (P)														
Year/Sem	III/V	Regulation	R24														
Date and Day of Assignment	28-07-2025	Time(s)	09:00AM -05:00PM														
Duration	2 Hours	Applicable to Batches	23CSBTB42														
Assignment Number: 01																	
Q. No.	Question				Expected Time to complete												
1	Green Cycle – E-Bike Company																
<div>Context:</div> <div>Green Cycle tracks how many social media reels affect demo test ride bookings.</div> <table><tr><th>Instagram Reels (x)</th><th>Test Rides (y)</th></tr><tr><td>1</td><td>20</td></tr><tr><td>2</td><td>35</td></tr><tr><td>3</td><td>48</td></tr><tr><td>1</td><td>25</td></tr><tr><td>2</td><td>38</td></tr></table>						Instagram Reels (x)	Test Rides (y)	1	20	2	35	3	48	1	25	2	38
Instagram Reels (x)	Test Rides (y)																
1	20																
2	35																
3	48																
1	25																
2	38																
<div>Objective:</div> <div>Analyze the effect of Instagram reel postings on the number of test ride bookings for Green Cycle by performing Linear Regression and interpreting SHAP values.</div> <div>Requirements:</div> <div><div>1. Perform Linear Regression Analysis</div><div><div>Use the given dataset where:</div><div><div>• Independent Variable (x): Instagram Reels</div><div>• Dependent Variable (y): Test Rides</div></div></div><div>2. Calculate the Baseline Value</div><div><div>Compute the mean of all test ride values (y values).</div></div><div>3. Calculate SHAP Values</div><div><div>For each record, calculate the difference between the predicted value and the baseline.</div><div>This difference is the SHAP value, attributed to the number of Instagram reels.</div></div><div>4. Compute Final Prediction</div><div><div>Use the linear regression model to calculate predicted test rides for each reel count.</div><div>Confirm that:</div><div>Final Prediction = Baseline + SHAP Value</div></div><div>5. Interpret the Results</div><div><div>Explain how the number of Instagram reels influenced each predicted test ride count.</div><div>Compare the predicted value to the actual value for each row.</div><div>Identify under prediction or over prediction, and provide reasoning.</div></div></div>																	
<div>Deliverables:</div> <div>A notebook or document containing:</div>																	

- Linear regression implementation with coefficients
- Baseline (mean of y)
- Table of SHAP values and predictions
- Explanation of how each input influenced the prediction
- Comparison of predicted vs actual values, with over/under prediction notes
- Summary analysis covering:
 - Accuracy of the model
 - Trend analysis
 - SHAP interpretation insights

Q. No.	Question	Expected Time to complete
2	GoFit – Gym Membership Signups using Multiple Linear Regression and SHAP Analysis	

Objective:

Analyze how trainer availability and advertising spend influence new gym membership signups using Multiple Linear Regression and interpret results using SHAP values.

Given Dataset:

Trainers (x_1)	Ad Spend (k₹) (x_2)	Signups (y)
5	50	80
4	40	70
6	60	90
3	30	60
2	20	50

Tasks:

- Perform Multiple Linear Regression Analysis**
 - Use Trainers and Ad Spend as independent variables
 - Use Signups as the dependent variable
- Calculate the Baseline Value**
 - Compute the mean of all signup values
- Calculate SHAP Values**
 - Calculate SHAP Value
 - Distribute SHAP contributions between Trainers and Ad Spend based on model coefficients
- Compute Final Prediction for Each Record**
 - Use the regression equation
 - Verify: Prediction = Baseline + SHAP (Trainers) + SHAP (Ad Spend)
- Interpret the Results**
 - For each record, explain how trainers and ad spend affected the prediction
 - Compare predicted vs actual signup values
 - Indicate overprediction or underprediction and possible reasons

Q. No.	Question	Expected Time to complete
3	Regression with Diabetes Dataset	

Objective:

Understand how patient features influence disease progression using Multiple Linear Regression and SHAP value analysis.

Tasks

1. Perform Multiple Linear Regression Analysis

- Use all available features from the Diabetes dataset as independent variables.
- Fit a Multiple Linear Regression model to predict disease progression.

2. Calculate the Baseline Value

- Compute the **mean** of the target variable (disease progression scores) from the training data.
- This will serve as the **baseline prediction**.

3. Calculate SHAP Values

- Apply SHAP to compute **feature contributions** to each prediction.
- Use model coefficients to proportionally attribute the difference from the baseline to each feature.

4. Compute Final Prediction for Each Record

- For every test record, verify that:

$$\text{Prediction} = \text{Baseline} + \text{SHAP}(\text{Feature}_1) + \text{SHAP}(\text{Feature}_2) + \dots + \text{SHAP}(\text{Feature}_n)$$

5. Interpret the Results

- For each patient record:
 - Explain how each feature contributed to the predicted disease progression.
 - Compare the **predicted value** vs the **actual observed value**.
 - Comment on whether the model **overpredicted or underpredicted** and **why**, based on SHAP values.

Q. No.	Question	Expected Time to complete
4	Regression with Student Performance Dataset	

Objective:

Investigate how student background and behavior influence final exam scores using Multiple Linear Regression and SHAP value analysis.

Tasks

1. Perform Multiple Linear Regression Analysis

- Use all relevant student attributes (e.g., study time, parental education, absences, etc.) as independent variables.
- Fit a regression model to predict the **final exam score**.

2. Calculate the Baseline Value

- Compute the **mean of the final exam scores** from the training set.
- This serves as the **baseline prediction** (expected value).

3. Calculate SHAP Values

- Use SHAP to compute the contribution of each student attribute to the final exam score prediction.
- Distribute the prediction deviation from the baseline among the features.

4. Compute Final Prediction for Each Record

- For each student record, confirm:

$$\text{Predicted Score} = \text{Baseline} + \text{SHAP}(\text{Feature}_1) + \text{SHAP}(\text{Feature}_2) + \dots + \text{SHAP}(\text{Feature}_n)$$

5. Interpret the Results

- For every prediction:
 - Explain how different features (e.g., study time, failures, health) impacted the exam score.
 - Compare predicted score to actual score.
 - Comment on overprediction or underprediction and possible reasons behind it.