

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	23CSBTB32														
Assignment Number: 01																	
Q. No.	Question				Expected Time to complete												
1	Learn Now – Online Course Platform:																
Objective:																	
Analyze the effect of email marketing on course enrollments for LearnNow by performing Linear Regression and interpreting SHAP values.																	
Requirements:																	
1. Perform Linear Regression Analysis																	
○ Use the given dataset where:																	
▪ Independent Variable (x): Number of Emails Sent																	
▪ Dependent Variable (y): Enrollments																	
<table><tr><th>Number of Emails Sent (x)</th><th>Enrollments (y)</th></tr><tr><td>1</td><td>80</td></tr><tr><td>3</td><td>120</td></tr><tr><td>2</td><td>95</td></tr><tr><td>1</td><td>85</td></tr><tr><td>3</td><td>130</td></tr></table>						Number of Emails Sent (x)	Enrollments (y)	1	80	3	120	2	95	1	85	3	130
Number of Emails Sent (x)	Enrollments (y)																
1	80																
3	120																
2	95																
1	85																
3	130																
2. Calculate the Baseline Value																	
• Compute the mean of all enrollments (y values).																	
3. Calculate SHAP Values																	
• For each record, calculate the difference between the predicted value and the baseline.																	
• This difference is the SHAP value, attributed to the number of emails sent.																	

4. Compute Final Prediction

- Use the linear regression model to calculate predicted enrollments for each x.
- Confirm that: Final Prediction=Baseline+SHAP Value

5. Interpret the Results

- Explain how the number of emails influenced each predicted enrollment.
- Compare the predicted value to the actual value for each row.
- Identify under prediction or over prediction and provide reasoning.

Deliverables:

A notebook or document containing:

- 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	ShopEase – E-commerce Revenue Prediction using Multiple Linear Regression and SHAP Analysis	

Objective:

Analyze how ad spend and discount percentage influence daily revenue using Multiple Linear Regression and interpret the results with SHAP value analysis.

Given Dataset:

Ad Spend (x ₁)	Discount (%) (x ₂)	Revenue (y)
200	10	1500
300	15	2000
250	5	1700
150	10	1400
100	0	1000

Tasks:

1. Perform Multiple Linear Regression Analysis

- Use Ad Spend and Discount (%) as independent variables

<ul style="list-style-type: none"> • Use Revenue as the dependent variable <ol style="list-style-type: none"> Calculate the Baseline Value <ul style="list-style-type: none"> • Compute the mean of all revenue values Calculate SHAP Values <ul style="list-style-type: none"> • Compute SHAP value • Distribute SHAP contribution between Ad Spend and Discount based on model coefficients Compute Final Prediction for Each Record <ul style="list-style-type: none"> • Verify prediction = Baseline + SHAP (Ad Spend) + SHAP (Discount) Interpret the Results <ul style="list-style-type: none"> • For each row, explain how Ad Spend and Discount % influenced the predicted revenue • Compare predicted vs actual revenue • Comment on over/under prediction 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:
Prediction = Baseline + SHAP(Feature₁) + SHAP(Feature₂) + ... + SHAP(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:
Predicted Score = Baseline + SHAP(Feature₁) + SHAP(Feature₂) + ... + SHAP(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.