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**	**23CSBTB33**
Assignment Number: 01			

Q. No.	Question	Expected Time to complete
1	**Chill Scoops – Ice Cream Outlet**	

Context:

Chill Scoops monitors how hot days (temperature in °C) affect daily ice cream sales.

Temperature (°C) (x)	Ice Creams Sold (y)
30	300
32	330
29	290
31	310
33	350

Objective:

Analyze the effect of temperature on daily ice cream sales for chill Scoops by performing Linear Regression and interpreting SHAP values.

Requirements:

1. **Perform Linear Regression Analysis**
   - Use the given dataset where:
     - **Independent Variable (x):** Temperature in °C
     - **Dependent Variable (y):** Ice Creams Sold
2. **Calculate the Baseline Value**
   - Compute the **mean of all ice cream sales (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 temperature.
4. **Compute Final Prediction**
   - Use the **linear regression model** to calculate predicted ice cream sales for each temperature value.
   - Confirm that:

   Final Prediction=Baseline+SHAP Value
5. **Interpret the Results**
   - Explain how the temperature influenced each predicted sales value.
   - Compare the predicted value to the actual value for each row.
   - Identify **underprediction** or **overprediction**, 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	**LearnNow – Online Course Platform – Enrollment Prediction using Multiple Linear Regression and SHAP Analysis**	

Objective:

Study how the number of emails sent and social media posts influence course enrollments by applying Multiple Linear Regression and interpreting the results using SHAP value analysis.

Given Dataset:

Emails Sent (x_1)	Social Posts (x_2)	Enrollments (y)
100	5	300
150	10	500
120	8	400
80	4	250
60	2	180

Tasks:

1. **Perform Multiple Linear Regression Analysis**
   - Use Emails Sent and Social Media Posts as independent variables
   - Use Enrollments as the dependent variable
2. **Calculate the Baseline Value**
   - Compute the mean of all enrollment values
3. **Calculate SHAP Values**
   - Compute SHAP value
   - Distribute SHAP contributions between Emails and Social Posts based on model coefficients
4. **Compute Final Prediction for Each Record**

   Verify: Prediction = Baseline + SHAP (Emails) + SHAP (Social Posts)
5. **Interpret the Results**
   - For each row, explain how Emails and Social Media Posts contributed to the prediction
   - Compare predicted vs actual enrollments
   - State if the model overpredicted or underpredicted and suggest 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(Feature$_1$) + SHAP(Feature$_2$) + ... + 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(Feature$_1$) + SHAP(Feature$_2$) + ... + 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.