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**	**23CSBTB45**
Assignment Number: 01			

Q. No.	Question	Expected Time to complete
1	**Snap Shoots – Photography Studio**	

Context:

Snap Shoots tracks the impact of Facebook ads (in ₹000s) on the number of wedding bookings.

Facebook Ads (x in ₹k)	Bookings (y)
1	8
2	15
3	20
1	10
2	17

Objective:

Analyze how Facebook ad spending influences wedding bookings using Linear Regression and SHAP value interpretation.

Requirements:

1. **Perform Linear Regression Analysis**
   - Use the dataset where:
     • Independent Variable (x): Facebook Ads (₹k)
     • Dependent Variable (y): Bookings
2. **Calculate the Baseline Value**
   - Compute the mean of all y-values (bookings).
3. **Calculate SHAP Values**
   - For each entry, calculate:
     SHAP Value = Prediction - Baseline
   - Attribute this value to the number of Facebook ads.
4. **Compute Final Prediction**
   - Use the regression equation to predict bookings for each Facebook ad value.
   - Verify that:
     Final Prediction = Baseline + SHAP Value
5. **Interpret the Results**
   - Explain how each input (Facebook Ads) contributed to the final prediction.
   - Compare predicted vs actual values.
   - Indicate whether the model overpredicted or underpredicted, with possible reasoning.

Deliverables:

• Regression equation (slope and intercept)

- Baseline booking value
- Table with Facebook ads, predicted bookings, SHAP values, and actual bookings
- Interpretation for each case
- Overall summary covering model accuracy, feature influence, and SHAP-based insights

Q. No.	Question	Expected Time to complete
2	**HomeGenie – Service Requests using Multiple Linear Regression and SHAP Analysis**	

Objective:

Analyze the influence of advertisement clicks and response time on the number of service bookings using Multiple Linear Regression and SHAP value interpretation.

Given Dataset:

Ad Clicks (x_1)	Response Time (min) (x_2)	Bookings (y)
150	10	300
100	15	250
120	8	280
90	12	240
80	20	200

Tasks:

1. **Perform Multiple Linear Regression Analysis**
   - Independent variables: Ad Clicks and Response Time
   - Dependent variable: Bookings
2. **Compute the Baseline Value**
   - Use the average of the bookings
3. **Calculate SHAP Values**
   - Calculate SHAP Value
   - Break down SHAP contributions by each feature
4. **Generate Model Predictions**
   - Apply the regression equation
   - Confirm decomposition: Prediction = Baseline + SHAP(Ad Clicks) + SHAP(Response Time)
5. **Interpret SHAP Contributions for Each Entry**
   - For every data point, explain the effect of each feature
   - Comment on whether the model overpredicts or underpredicts, and why, using SHAP values

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.

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.

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.