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 Batch**	**23CSBTB31**
Assignment Number: 01			

Q. No.	Question	Expected Time to complete
1	**FitFuel – Linear Regression with SHAP Analysis**	

Objective:

Analyze the impact of influencer marketing on product sales for **FitFuel**, a health drink brand, by performing **Linear Regression and SHAP value interpretation**.

Requirements:

1. **Perform Linear Regression Analysis**
   - Use the given dataset with Number of Influencer Posts (x) as the independent variable and Bottles Sold (y) as the dependent variable.

Number of Influencer Posts (x)	Bottles Sold (y)
1	120
2	150
3	180
1	130
2	160

2. **Calculate the Baseline Value**

   Compute the **mean** of all y values (i.e., average number of bottles sold).

3. **Calculate SHAP Values**
   - Identify the **difference between model prediction and baseline** to compute SHAP values for each observation.
   - Attribute the contribution of the independent variable (x) to the output.

4. **Compute Final Prediction for each record**

   Use the linear regression equation to calculate predicted y.

   Verify that: Final Prediction=Baseline+SHAP Value

5. **Interpret the Results**

   - Explain how each input (number of influencer posts) influenced the final prediction.
   - Compare predicted vs actual value for each row.
   - Identify if the model **underpredicted or overpredicted** and explain why.

Deliverables:

A notebook or document with:

- Linear regression implementation and coefficients
- Baseline calculation
- Table of SHAP values and final predictions
- Interpretation of each result (in simple terms)
- A summary analysis of model behavior (SHAP interpretation)

Q. No.	Question	Expected Time to complete
2	**Study Boost – Multiple Linear Regression with SHAP Analysis**	

Objective:

Analyze how IQ and Study Hours impact student test scores using Multiple Linear Regression and interpret the results using SHAP value analysis.

Requirements:

1. Perform Multiple Linear Regression Analysis

- Use the given dataset with:
  - **IQ (x_1)** and **Study Hours (x_2)** as independent variables
  - **Test Score (y)** as the dependent variable

IQ (x_1)	Study Hours (x_2)	Test Score (y)
110	40	100
120	30	90
100	20	80
90	0	70
80	10	60

2. Calculate the Baseline Value

- Compute the mean of all **y values** (i.e., average test score).

3. Calculate SHAP Values

- Compute SHAP values for each record:

4. Compute Final Prediction for Each Record

Final Prediction = Baseline + SHAP value (sum of SHAP values for IQ and StudyHours)

5. Interpret the Results

For each row:

- **Explain how IQ and Study Hours contributed to the final prediction**

- **Compare predicted vs actual value**
- **Indicate if the model overpredicted or underpredicted and suggest why**

Deliverables:

Prepare a notebook or report containing:

- **Multiple Linear Regression equation and coefficients**
- **Baseline calculation**
- **Table with SHAP values and final predictions**
- **Simple interpretation of each prediction**
- **Summary analysis of how IQ and Study Hours influenced outcomes (SHAP interpretation)**

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.