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**	**23CSBTB39**
Assignment Number: 02			

Q. No.	Question	Expected Time to complete
1	**Assignment: Feature Importance Analysis using SHAP**	

Objective

To select a publicly available dataset from any domain, apply SHAP (SHapley Additive exPlanations) to identify important features, build a predictive model, and interpret the results in detail.

Dataset Selection Guidelines

Students choose datasets from the domain:

- **Health – e.g., patient diagnostics, disease prediction, medical records.**

Requirements for dataset selection:

- At least 500 rows of data.
- Minimum 5 independent variables (features).
- A clear target variable for classification or regression.
- Dataset must be publicly accessible (Kaggle, UCI Repository, government portals, etc.).

Tasks

- Data Collection & Preprocessing
- Download the chosen dataset in .csv format/ or any.
- Load it into Python using Pandas.
- Handle missing values, duplicates, and outliers.
- Encode categorical variables if needed.
- Normalize or standardize data when required.

Model Building

- Split the dataset into training (80%) and testing (20%) sets.
- Choose a suitable model (e.g., Random Forest, Logistic Regression, XGBoost).
- Train and evaluate the model using relevant metrics:
- Classification: Accuracy, Precision, Recall, F1-score, ROC.
- Regression: RMSE,MSE, MAPE,MPE, MAE, R^2 score.

SHAP Implementation

- Install and import SHAP (pip install shap).
- Select an appropriate SHAP explainer (TreeExplainer, KernelExplainer, etc.).
- Compute SHAP values for the test set.

Generate and include:

- Summary plot – overall feature importance.
- Force plot – individual prediction explanation.
- Waterfall plot – step-by-step feature contribution.

Result Interpretation

- Identify and explain the top 5 most influential features.
- Compare SHAP feature importance with the model's built-in feature importance (if available).
- Discuss whether the results are meaningful in the chosen domain.

Report Preparation

- Title Page – Assignment title, student name, roll number, date.
- Introduction – Problem statement and dataset overview.
- Dataset Description – Source, size, features, target variable.
- Preprocessing Steps – Cleaning and transformation details.
- Model & Performance – Algorithm choice, parameters, evaluation metrics.
- SHAP Analysis – Plots and explanations.

Conclusion – Key insights, limitations, and possible improvements.

Submission Requirements

- Python code file (.ipynb or .py).
- Dataset file (.csv).
- Report (.pdf) including SHAP plots and explanations.