

EXPLAINABLE AI

Assignment - 1

Linear Regression with SHAP Analysis

(Book Heaven – Online Bookstore)

Name: B. Rithwik

Hall Ticket: 2303A52330

Batch: 35

Instructor: Dr. Vairachilai Shenbagavel

Introduction:

Book Heaven is an online bookstore that uses Google Ads to promote its products and tracks the effect of ad spending on weekly book sales. To understand this relationship, Linear Regression is used to model how advertising influences sales. However, prediction alone is not enough. It is also important to explain why the model gives certain results. For this purpose, SHAP (Shapley Additive Explanations) is applied to show how much ad spending contributes to each prediction compared to the baseline. This makes the model both accurate and interpretable.

Problem Statement & Requirements:

- Detailed re-write of assignment requirements in your own words.
- Each requirement broken into subpoints (Linear Regression, Baseline, SHAP, Prediction Check, Interpretation).
- Show how each requirement will be solved in the notebook.

Dataset Description:

- Present the dataset in a **nice table**.
- Highlight **Independent variable (Google Ads)** and **Dependent variable (Books Sold)**.
- Explain why this dataset is appropriate for regression.

Methodology

- Explain **Linear Regression equation** in detail.
- Show formula for baseline.
- Define **SHAP = Prediction – Baseline**.
- Explain the formula:

$$\text{Final Prediction} = \text{Baseline} + \text{SHAP}$$

- Step-by-step workflow diagram (optional, but can be added).

Results

- Regression equation with coefficients.
- Baseline value.
- SHAP Table with columns:
 - Ads, Actual, Predicted, Baseline, SHAP, Baseline+SHAP, Residual, Over/Under.
- Highlight with **interpretation row-by-row** (like: “For Ads=1, model underpredicted because actual was higher than expected”).

Model Evaluation

- R^2 , MSE, MAE with explanation of what they mean.
- Explain how high R^2 indicates model fits well even with small dataset.

SHAP Analysis

- Explain SHAP in simple words: *“how much each input (ad spend) pulls prediction above or below baseline”*.
- Example breakdown:
 - Ads = 3 → SHAP = +33.43 → means higher spend strongly increases sales.
 - Ads = 1 → SHAP = -22.29 → means low spend decreases sales below average.

Discussion

- Trend analysis (positive correlation between ads & sales).
- Over/Under prediction reasons (small dataset, real-world noise).
- Marketing insight: increasing ad spend increases sales.

Conclusion

- Recap model + SHAP insights.
- Explain how Explainable AI helps managers trust predictions.
- Future scope: using larger datasets, multiple regression with more marketing channels, etc.