

Branch Name:	Computer Engineering / Information Technology
Program Code:	ET103/ET105
Course Title:	Elective -IV Practical: Explainable Artificial Intelligence Practical
Course Code:	2ET1030705P
Prerequisite Course:	Basic of Artificial Intelligence and Machine Learning

Course Objectives: The Course Objectives for Explainable Artificial Intelligence (XAI) focus on transforming "black-box" models into transparent, interpretable, and trustworthy systems. The primary goal is to provide students with the tools to answer the "Why?" behind AI decisions.

Teaching and Examination Scheme:

Credit	Teaching Scheme Hours per week				Evaluation Scheme (Marks)								
					Theory (Marks)				Practical (Marks)				Total (Marks)
	L	T	P	Total	SEE		CCA		SEE		CCA		
					Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	
1	-	-	2	2	-	-	-	-	10	25	10	25	50

List of Experiments:

S.N.	Practical Exercise
1	Model Interpretation using LIME: Develop a classification model using a suitable dataset and apply LIME (Local Interpretable Model-agnostic Explanations) to explain individual predictions. Analyze how input features contribute to a specific prediction and interpret the model's local behavior.
2	Feature Importance using SHAP: Build a machine learning model and apply SHAP (Shapley Additive Explanations) to determine global feature importance. Visualize and interpret how each feature impacts the overall model predictions.
3	Comparison of LIME and SHAP: Implement both LIME and SHAP techniques on a trained model and compare their explanations in terms of interpretability, consistency, and usability. Evaluate which method provides better insights into model behavior.
4	Interpretable vs Black-Box Models: Train an interpretable model (e.g., Decision Tree) and a black-box model (e.g., Random Forest). Compare their performance and explainability, and analyze the trade-off between accuracy and interpretability.
5	(1) Visualization of Feature Contributions: Create visual representations of feature importance using XAI tools. Interpret how different features influence predictions and identify the most significant variables affecting the model output.
6	Local vs. Global Explanation: Demonstrate the difference between local and global explanations by applying XAI techniques. Analyze how individual predictions differ from overall model behavior.
7	Bias Detection using XAI: Use XAI techniques to analyze a dataset and identify potential bias in model predictions. Evaluate whether certain features unfairly influence outcomes.
8	Explainability in Classification Models: Build a classification model and apply XAI techniques to interpret decision boundaries and prediction logic. Assess whether the model decisions are transparent and understandable.
9	Explainability in Regression Models: Develop a regression model and use XAI methods to explain continuous output predictions. Analyze feature impact on predicted values.
10	Model Transparency Evaluation: Evaluate the transparency and reliability of a machine learning model using explainability techniques. Discuss how XAI improves trust and understanding in AI systems.

Reference Books:

1. Explainable Artificial Intelligence: An Introduction to Interpretable Machine Learning By Uday Kamath, John Liu, Springer Nature Switzerland AG (Springer)
2. Explainable AI: Foundations, Methodologies and Applications By Mayuri Mehta, Vasile Palade, Indranath Chatterjee, Springer (Intelligent Systems Reference Library Series)
3. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning By Wojciech Samek, Grégoire Montavon, et al., Springer (Lecture Notes in Computer Science)
4. Distributed Deep Learning and Explainable AI (XAI) in Industry 4.0 By Lalitha Krishnasamy, Rajesh Kumar Dhanaraj, et al., Springer

List of open-source software/learning websites:

1. https://onlinecourses.nptel.ac.in/noc26_cs01/preview
2. <https://www.coursera.org/specializations/explainable-artificial-intelligence-xai>
3. <https://www.codecademy.com/learn/learn-explainable-ai>
4. <https://alison.com/course/explainable-ai-explained>
5. <https://www.udemy.com/course/explainable-ai-xai-for-generative-ai>

Course Outcomes (CO):

Course Outcomes	Description	Bloom's Taxonomy Level
CO1	Describe the necessity of XAI in regulated domains like healthcare and finance.	Understanding (2)
CO2	Implement model-agnostic techniques for tabular and image data.	Applying (3)
CO3	Analyze deep learning models using visualization tools.	Analyze (4)
CO4	Critique the trade-offs between model accuracy and model interpretability.	Evaluate (5)
CO5	Design a transparent AI pipeline that incorporates "Human-in-the-loop" feedback.	Create (6)

Mapping of CO with PO and PSO:

COs	Program Outcomes (POs)												Program Specific Outcomes (PSOs)			
	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3	PSO 4
CO1	M	M				H		H					M			
CO2	H	H	M	M	H								H	H		
CO3	H	H	M	H	H								M	H		
CO4	M	H	H	H		M		M					M	H		
CO5	M	M	H	H	H	M		M					H	H		

H: High, M: Medium, L: Low