

Lab 3 -Task: Ensemble Learning and Explainable AI

Course: CSE 475 - Machine Learning

Title: Exploring Ensemble Learning and Explainable AI on Kaggle Datasets

Submission Mode: Git hub link of Google Colab Notebook (Code and Results Documentation)

Objective:

This assignment aims to deepen your understanding of ensemble learning techniques and explainable AI. You will implement bagging, boosting, stacking, and voting approaches on your assigned dataset and evaluate performance using cross-validation. Additionally, you will leverage SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to interpret your models' predictions.

Instructions:

1. Dataset Assignment:

- Each group is assigned a specific dataset based on the table below:

Group	Dataset	Link
A	Loan Approval Classification Data	https://www.kaggle.com/datasets/taweilo/loan-approval-classification-data
B	Heart Dataset	https://www.kaggle.com/datasets/mfarhaannazirkhan/heart-dataset
C	Anemia Diagnosis	https://www.kaggle.com/datasets/zeesolver/uhygtttt
D	CDC Diabetes Health Indicators	https://www.kaggle.com/datasets/abdelazizsami/cdc-diabetes-health-indicators
E	Cyber Threat Detection	https://www.kaggle.com/datasets/hussainsheikh03/cyber-threat-detection
F	Cardiology Unit Admission	https://www.kaggle.com/datasets/mansoorahmad4477/cardiology-unit-admission
G	Heart Disease Dataset	https://www.kaggle.com/datasets/krishujeniya/heart-disease

2. Ensemble Learning Techniques:

- **Bagging:** Implement at least one bagging method (e.g., Random Forest, Bagged Decision Trees) on your dataset.
- **Boosting:** Apply a boosting technique (e.g., AdaBoost, Gradient Boosting, XGBoost, or LightGBM) and analyze its performance.
- **Stacking:** Use at least two base models and a meta-model for stacking. Document the model selection process.
- **Voting:** Implement hard and soft voting classifiers by combining at least three algorithms of your choice.

3. Cross-Validation:

- Apply cross-validation to evaluate the performance of each ensemble method.
- Document and compare the cross-validation scores for each approach.

4. Explainable AI (XAI):

- **SHAP:** Use SHAP to explain the predictions of your best-performing model. Provide visualizations and explanations for significant features impacting model decisions.
- **LIME:** Apply LIME on a few test instances and illustrate how it helps interpret individual predictions.

5. Performance Analysis:

- Compare the accuracy, precision, recall, F1 score, and any other relevant metrics for each ensemble method.
- Discuss any trade-offs and insights gained by using different ensemble techniques on your dataset.

6. Documentation in Colab Notebook:

- Provide a clear, step-by-step code implementation for each ensemble method.
 - Include comments and markdown cells explaining each section of the code.
 - Display cross-validation results, confusion matrices, SHAP plots, and LIME results with analysis.
 - The notebook should be self-contained, allowing any reviewer to understand the logic and execution flow without additional context.
-

Evaluation Criteria:

1. Code Completeness (10%):

- Correct implementation of all specified ensemble methods.
- Application of cross-validation for each method.

2. Explainability (20%):

- Effective use of SHAP and LIME for model interpretation.
- Quality of visualizations and interpretations.

3. Analysis and Comparison (25%):

- Thorough performance comparison of different ensemble methods.
- Insightful discussion on results and observations.

4. Notebook Quality and Clarity (15%):

- Well-organized and documented code.
- Use of markdowns and comments for explanations.

5. Present the work in the Lab (30%)

Important Notes:

- Ensure your Google Colab notebook is accessible (public or shared with the instructor's email).
- Save and verify your results before the viva as no further modifications will be allowed post-deadline.
- Use relevant libraries for SHAP and LIME, ensuring they work seamlessly in your Colab notebook environment.

Groups:

GROUP_A

2019-3-60-019	Md. Safinur Rahman
2019-3-60-098	Al Shahriyar Shrabon
2020-1-60-062	Asif Amir Noor
2020-1-60-092	Hasnain Ahmed
2020-1-60-108	Abu Rayhan Akash

GROUP_B

2020-1-60-232	Rokeya Jahan Chowdhury Ettifa
2020-1-60-245	Sadia Mahbub Chowdhury
2020-1-64-108	Sanjoy Kumar Das
2020-2-60-009	Umme Haney
2020-2-60-057	Tamanna Sultana Tinne

GROUP_C

2020-2-60-108	Bishowjit Banik
2020-2-60-160	Sumaiya Salam
2020-3-60-082	Fahmida Islam Sumi
2020-3-60-092	Apurbo Chandra Paul
2021-1-60-038	Partho Biswas

GROUP_D

2021-1-60-060	Tansiv Jubayer
2021-1-60-088	Mahmud Bin Shafi
2021-1-60-106	Tuly Rahman
2021-1-60-108	Fatema Jalal
2021-2-60-027	Fahmida Anjum

GROUP_E

2021-2-60-051	Nirzona Binta Badal
2021-2-60-052	Ismail Mahmud Nur
2021-2-60-057	Tithi Paul
2021-2-60-065	Turjahan Islam
2021-2-60-083	Md. Ripon Al Mamun

GROUP_F

2021-2-60-102	Shamil Bin Hossain Noor
2021-2-60-133	Sabit Al Alfi
2021-3-60-053	Ishrat Jahan Momo
2021-3-60-170	Fatema Tahsin Anamika
2021-3-60-280	Halima Sadia

GROUP_G

2021-3-60-298	Shanta Singha
2022-1-60-044	Md. Murad Khan Limon
2022-1-60-108	Junnun Mohamed Karim
2022-1-60-127	Joy Kumar Oja
2022-1-60-162	Md. Yousuf Hozaiifa