Evaluation Document for Explainable AI Project

---

1. Introduction

Building on your previous semester projects—which involved a Machine Learning challenge and a Deep Learning challenge—this module focuses on **Explainable AI (XAI)**. While you have already tested multiple algorithms to achieve the best performance on leaderboard tasks, the next step is to **interpret and explain** these models' decisions. This document outlines the requirements, deliverables, and grading criteria for your XAI project.

---

2. Link to Previous Projects

1. **Machine Learning Challenge (First Semester):**
    - You experimented with a range of machine learning models.
    - You aimed to produce a strong leaderboard ranking based on performance metrics.
    - You delivered a short report and code illustrating your approach and findings.
2. **Deep Learning Challenge (First Semester):**
    - You explored neural network architectures for complex data (images, text, or tabular).
    - You optimized hyperparameters for accuracy and efficiency.
    - You documented your process and produced a working prototype.

These projects formed a solid foundation in model-building, hyperparameter tuning, and empirical evaluation. Now, you will **extend** these efforts by applying Explainable AI methods to better understand how your algorithms arrive at predictions.

---

3. Project Scope: Explainable AI

For this semester's project, you will:

1. **Choose one (or both) of your previous projects** (Machine Learning or Deep Learning).

2. **Integrate Explainable AI methods** to provide:
    - **Global explanations**: overall feature importance and how each feature influences the model's performance.
    - **Local explanations**: interpretability on a single-instance basis, especially focusing on **outliers** and **wrong predictions**.
3. Provide clear documentation (report), source code, and a presentation.

---

4. Deliverables

1. **Code**
    - Clean, well-structured, and commented.
    - Must integrate XAI libraries or custom interpretability functions (e.g., using methods such as SHAP, LIME, Grad-CAM for deep learning, or permutation-based importance for tree-based models).
    - Include any scripts or notebooks used for pre-processing, model training, evaluation, and interpretability.
2. **Report**
    - **Introduction**: Brief recap of the initial project (dataset, goal, main algorithms used).
    - **Methodology**: Explain the XAI techniques you chose and why (SHAP, LIME, feature importance, partial dependence plots, etc.).
    - **Results**: Present both **global** and **local** explanations. Highlight the most influential features and analyze examples of outliers or misclassifications.
    - **Discussion**: Reflect on the results, including any insights gained about model behavior, data biases, or potential improvements.
    - **Conclusion**: Summarize key takeaways from employing explainability techniques.
3. **Presentation**
    - Duration: **20 minutes** per group + **5 minutes** of Q&A.

- Groups: **2 to 4** members.
- Focus:
  - Overview of your initial model (or models).
  - Methods used for XAI and reasons for choosing them.
  - Demonstration of real examples (particularly wrong predictions and how the XAI method explains them).
  - Lessons learned & next steps.

---

5. Timeline

- **Next Practice Sessions** (29/03 and 01/04):
  - Use this time to polish your code and solidify your presentation flow.
  - Each group will present for 20 minutes, followed by 5 minutes of Q&A.
- **Final Deliverables**:
  - **Code and Report**: Submit by the end of the second session (01/04).
  - **Presentation**: Presented live during the practical sessions.

---

6. Grading Criteria

Component	Weight	Description
Code Quality	30%	- Readability (comments, structure) - Implementation correctness - Appropriate use of XAI libraries - Efficient, reproducible results
Report	30%	- Clarity of explanation - Depth of analysis - Proper justification for chosen XAI methods - Quality of examples and interpretation - Organized and coherent structure

Presentation	20%	- Organization and clarity - Ability to communicate technical aspects to peers - Quality of visuals/demonstrations - Handling of questions and audience engagement
Innovation & Insight	20%	- Novelty or creativity in approach - Insightful discussion of results - Thorough exploration of outliers and misclassifications - Reflection on limitations and future improvements

Total: 100%

---

7. Additional Recommendations

- **Emphasize Clarity**: Use plain language whenever possible, especially when explaining complex models and interpretability methods.
- **Demonstrate Practical Value**: Show real-world or business implications of your explanations (e.g., how understanding predictions can lead to better user trust or model refinement).
- **Compare Multiple Methods**: If possible, illustrate more than one XAI technique and discuss their strengths/weaknesses.
- **Be Consistent**: Use consistent variable naming and data labeling across code, report, and presentation.
- **Rehearse**: Make sure the 20-minute presentation is well-timed and that you anticipate potential questions on methodology and results.

---

8. Conclusion

This evaluation aims to foster a deeper understanding of your Machine Learning or Deep Learning models by leveraging Explainable AI. By showing **how** and **why** a model makes certain predictions, you will gain insights into data quality, model reliability, and potential areas for improvement. This is a crucial step in deploying AI responsibly and effectively.

We look forward to your final code, insightful reports, and engaging presentations!

Good luck with your projects and presentations!