Explainable Techniques - Machine Learning Court

60 Points Possible

9/22/2025

Attempt 1

In Progress

NEXT UP: Submit Assignment

Add Comment

Unlimited Attempts Allowed

Details

Explainable Techniques - Machine Learning Court

You will investigate a real-world AI decision and argue either for or against its validity using explainable AI (XAI) techniques such as SHAP, LIME, or Anchors. You'll work individually to analyze the model, then collaborate with peers to craft and deliver a compelling group case.

Instructions

1. **Get Your Case Assignment**
   You will be assigned:
   - One of three case domains (Loan, Medical, Recidivism)
   - A position: **Prosecution** or **Defense**
2. **Start with the Notebook**
   - Open the **starter notebook** (https://github.com/AIPI-590-XAI/Duke-AI-XAI/blob/main/assignments/machine_learning_court.ipynb)
   - Load and explore the assigned dataset and model
   - Use at least **two XAI methods** (e.g. SHAP, LIME, Anchors) to explain the focus instance
3. **Individual Analysis (Done Solo)**
   - Submit your completed notebook with:
     - Explanation results (plots + brief interpretation)
4. **Group Collaboration (In Class)**
   - Meet with teammates in your same role/case
   - Combine findings into a **5-minute argument**
   - Deliver your case to the full class (jury) on trial day
5. **Trial Format**
   - Prosecution and Defense each get 5 minutes to present
   - The rest of the class will act as a jury and vote
   - Time for debrief and open discussion follows

Submission

You must submit your individual notebook with explanations prior to class.

In class, you will present with your team and will receive a participation grade.

Online Students ONLY: You all will be graded only on your notebook, so please ensure you provide a clear case in the markdown of your notebook.

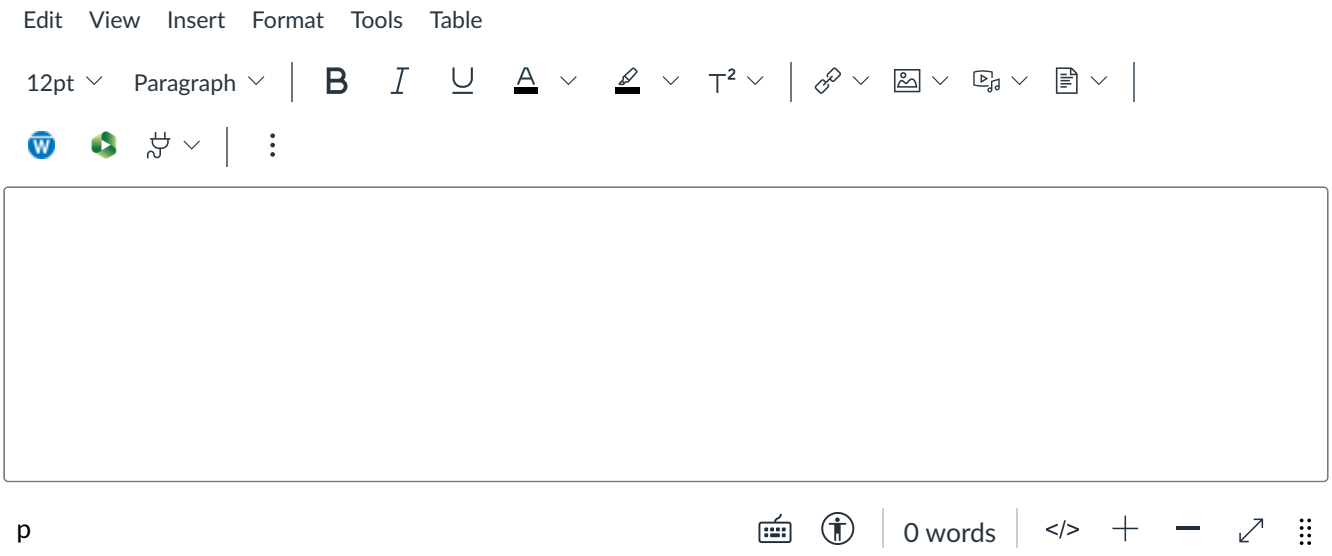
Rubric (60 Points Total)

Individual XAI Notebook (30 points)

- XAI Implementation - at least two methods with visual and written interpretation
- Insightful and well-documented explanation of results
- Uses best practices

Team Presentation (30 points)

- Presents clear case with logical reasoning
- Use of XAI evidence



(https://canvas.duke.edu/courses/62464/modules/items/669347)
Assignment
(https://canvas.duke.edu/courses/62464/modules/items/572805)