

## **CSE674: Machine Learning-1**

### **Assignment # 3**

**Deadline: May 27, 2023**

**Marks: 15**

In this assignment, you will have the opportunity to practice explainable AI (XAI) methods. Specifically, you are required to follow the steps below:

- Choose a dataset for either a classification or regression problem (you can use a public dataset from Kaggle/other sources or collect one yourself). Please ensure that your dataset has a reasonable number of instances and attributes (around 15-20, not more than that).
- Split the dataset into training and testing sets. Use appropriate preprocessing techniques, such as normalization or feature selection, to ensure high-quality input for your models.
- Train and compare the performance of four corresponding (for classification or regression) Machine Learning algorithms studied in this course. Choose the best algorithm based on the evaluation metrics (e.g., accuracy, ROC AUC, RMSE, etc.) on the test set.
- Select the best model, and use LIME, Shapley, and Counterfactual to explain how the model makes decisions on three sample scenarios (instances from the test set). Specifically, identify the most important features contributing to the model's prediction and how changing those features would affect the prediction outcome. Please be specific in your evaluation of these explainable AI methods, and make sure to highlight their strengths and limitations. For example, you could compare the uniformity of the explanation across different methods or discuss the interpretability of different features.
- Implement a programming interface using either Streamlit or Gradio that allows users to interact with your ensemble model and get explanations for the predictions made. This interface should allow users to input the feature values for a new instance and display the corresponding prediction and explanations generated by LIME, Shapley, and Counterfactual. Make sure that the interface is user-friendly and well-designed.

In your assignment report, please provide sufficient details of the following:

- The dataset you have chosen, along with its attributes and the target variable.
- The preprocessing techniques you have applied and why.
- The three standard Machine Learning algorithms you have trained and how you have compared their performance.
- The ensemble model you have built and how you have evaluated its performance.

- The three sample scenarios you have chosen, along with their feature values and the corresponding predictions from the ensemble model.
- The explanations generated by LIME, Shapley, and Counterfactual for each of the three sample scenarios.
- A detailed description of the programming interface you have implemented using either Streamlit or Gradio, along with screenshots of its different components and functionalities.