Explain and Visualize AI Algorithms using LIME and SHAP

Objective:

The objective of this assignment is to deepen students' understanding of model interpretation using XAI algorithms such as LIME and SHAP values.

Instructions:

1. Download the dataset: https://www.kaggle.com/datasets/tawfikelmetwally/employee-dataset/data
2. Build a random forest classifier by using Python sklearn library to determine whether the employee will leave or not (70% training, 30% testing).
3. Figure out how much each feature (such as education, age, city, etc.) contributes to the model's prediction on one randomly selected data point.
    a. Using LIME
    b. Using SHAP
4. Visualize the results, for example, using bar charts or other charts
    a. For LIME
    b. For SHAP
5. Write a report to analyze the results (maximum three A4 pages)
    a. Include the visualizations for the explanation from LIME and SHAP respectively
    b. Try to explain how the AI algorithm works using results from LIME and SHAP respectively
    c. Compare the computation time and the result from SHAP and LIME

Submission criteria:

1. Zip your code (.py/.ipynb) and report (.pdf/.word)
2. Use your name to name the zip file
3. Send to canvas assignment 2

Grading criteria:

1. Code: 20 points (10 points each for LIME and SHAP)
2. Visualizations in the report (20 points each for LIME and SHAP)
3. Interpretation of the results (20 points each for LIME and SHAP)
4. Comparison the results between LIME and SHAP (10 points)

Tutorials for references:

SHAP
https://www.datacamp.com/tutorial/introduction-to-shap-values-machine-learning-interpretability

LIME
https://www.kaggle.com/code/prashant111/explain-your-model-predictions-with-lime/notebook