MIS 584 APPLIED MACHINE LEARNING

HOMEWORK III (8 points)

Introduction

Explainable AI (XAI) has developed as a subfield of AI, focused on exposing complex AI and ML models to humans in a systematic and interpretable manner. XAI helps to understand, visualize and interprete machine learning models.

There are two tasks in the assignment (Grading (1st task: 3 points, 2nd task: 5 points)). Please note that using XAI techniques in your term project is highly encouraged. You will have also question(s) from XAI domain in your FINAL exam.

1. XAI Literature Research Report

Prepare a short research report (or slides) for XAI. Your report template can be both in doc/docx (1-2 pages) or ppt/pptx (3-6 slides). Answer the following questions in your research report (or slides). Don't forget to state your references.

a) What are the properties / dimensions of XAI?
b) What are the popular algorithms used in XAI domain?
c) Compare the XAI algorithms (pros, cons, performance etc.)

2. XAI Algorithm Usage Demo via Jupyter Notebook

a) Describe your dataset, problem and evaluation metrics.
b) Build a Machine Learning model.
c) Use one or more XAI algorithm(s) to explain the model you built.
d) Support your solution with plot(s) and give detailed explanation about your results.

You may use any dataset(s) and any model you wish for the assignment. You may select your dataset from kaggle, UCI Machine Learning Repository or another publicly available datasets. You can find many others on the Web. Pick something that interests you.

Deliveries

Upload your research report (in doc/ docx or ppt/ pptx format) , jupyter solution notebook(s) and dataset(s). Don't forget to put code descriptions (markdown or comments), mention about your references/sources in your notebook. Each student will share their final visualization in class via jupyter notebook.

Please plan for a demo of up to 20 minutes(5 to the 10 minutes research, 10 to the 15 minutes XAI demo). Assignments not presented in the class will not be graded.

1. Research Report
2. Jupyter Demo Notebook

Some resources to check **:**

1. Explaining Explanations: An Overview of Interpretability of Machine Learning, Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, Lalana Kagal, 2019, https://arxiv.org/abs/1806.00069
2. Interpretable Machine Learning — A Guide for Making Black Box Models Explainable. https://christophm.github.io/interpretable-ml-book/
3. SHAP: A Unified Approach to Interpreting Model Predictions. arXiv:1705.07874
4. https://analyticsindiamag.com/8-explainable-ai-frameworks-driving-a-new-paradigm-for-transparency-in-ai/