University of Ottawa

School of Electrical Engineering and Computer Science

CSI5155 - Fall 2022

Assignment 3: XAI

TOTAL MARKS 70

The aim of this to reflect on explainability (XAI), interpretability, and trustworthiness.

You are referred to the online book by Christoph Molnar [2], that introduces many relevant topics.

Instruction:

1. This is an individual assignment. Submit your assignment using BrightSpace, before the due date.
2. We cannot accept late submission.
3. For the implementation, you should either upload your code on BrightSpace or provide a link to a GitHub repository. Note that, if you choose to use GitHub, the date and time of last change to your repository should be **before** the assignment deadline.
4. Submit your report as a PDF file.

Topic: Explainable AI (XAI), interpretability, and trustworthiness

Part A [50 marks]

You are asked to explore the most accurate model obtained by the decision tree algorithm against the Drug Consumption dataset as used in assignments 1 and 2.

Answer the following questions.

Your answers should focus on feature importance, the properties of the dataset, the characteristics of the algorithm, and the usefulness of the resultant model.

1. Display/visualised the resultant model created by the decision tree. **[20 marks]**
2. Explain how, and why, the algorithm made a specific decision. **[10 marks]**
3. Explain why the algorithm didn't do something else. **[10 marks]**
4. Discuss when the algorithm succeeded and when it failed. **[10 marks]**
5. Explain how you would decide if the resultant model can be trusted. **[10 marks]**
6. Explain how the algorithm could potentially improve its predictions. **[10 marks]**

References/ Some useful links:

1. Course slides, Topic 8: linear models, slide 36.
2. Online book: https://christophm.github.io/interpretable-ml-book/
3. Showing trees:
   a. https://scikit-learn.org/stable/modules/generated/sklearn.tree.export_text.html and
   b. https://scikit-learn.org/stable/modules/generated/sklearn.tree.plot_tree.html

Other additional and optional resources:

4. Rule Fit: https://arxiv.org/abs/0811.1679
5. Useful links for visualization: https://github.com/parrt/dtreeviz or https://graphviz.org/.