Manual

Hélia Galinier

March 27, 2025

1 Introduction

In this document, it will be explained how to run the code to use the tool. The file containing this code is "decision_tree.py".

2 Instructions

2.1 Open the files

Open the file with any IDE like Visual Studio Code or others.

2.2 Change the dataset

To change the dataset, simply change the value of the 'dataset_name' variable at the beginning of the main function. The names of the datasets are : 'adult', 'compas', 'law_school', 'kdd', 'dutch', 'credit', 'communities_crime', and 'german'. The name directly corresponds to the dataset, so putting one of the previously cited names will take it into account. The program will automatically extract the needed information concerning the dataset using the 'info_dataset.csv' file.

2.3 Change the parameters

The main function used to run the tool is 'calculate_idi_ratio_tool'. The main parameters don't need to be touched. The only parameters that can be touched are the optional ones. There is the 'threshold' that is set to 0.05 as default, 'num_training' which is the number of training data used for the decision tree, which is set to 8000 as default, and 'num_samples' which is the number of samples extracted by the algorithm to be tested, set to 1000 as default.

2.4 How to run the code

To run the code, use the command : 'python decision_tree.py'. The results will be printed out on the terminal. You can also just click the 'run' button of the IDE. The results will be printed in the console or terminal.