

Trustworthy machine learning 2024

Project (Task 3)

Fairness-aware machine learning

Project

- Form a group of 2 students (recommended) or 1 person group
- Select a dataset (or bring your own), come up with a story for predictive modeling/automated decision making.
- Measure discrimination in data, justify the selection of the measuring method
- **Step 1:** Test a simple predictive model that has no fairness-awareness strategy, measure discrimination (PRESENTATION I, 5 min per group).

The presentation slides are to be submitted through Moodle by 14:00 Dec 2 and presented in the practical session on the same day.

(max. 5 points)

- **Step 2:** Implement a fairness-aware predictive model, measure the effect, discuss (PRESENTATION II, 10-15 min per group).

The presentation slides are to be submitted through Moodle by 12:00 Dec 9 and presented in our practical session on the same day.

(max. 10 points)

- **Step 3:** Write up a report: ~5 pages of text + any figures, one report per group, summarizing your case study. Please use UH Master thesis template. Abstract and references do not count towards the page limit. Please include an explicit statement describing any use of large language models in the assignment or reporting. Using of large language models is permitted at own risk.

- **The final report is to be submitted through Moodle by 16:00 Dec 16.**

(max. 15 points towards the final grade)

Possible themes and datasets

Census income (Adult) dataset

Theme: records level of income of individuals and demographic information

Possible prediction task: predict the level of income (e.g. for taxation)

Data size: about 50 000 records

Data available from: <https://archive.ics.uci.edu/ml/datasets/adult>

COMPAS Dataset

Theme: jail and prison times, demographics, criminal histories

Possible prediction task: predict whom to release on parole (e.g. application in courts)

Data size: about 20 000 records

Data available from: <https://github.com/propublica/compas-analysis/>

German credit

Theme: demographic information, credit history

Possible prediction task: to give a credit or not

Data size: 1000 records

Data available from:

[https://archive.ics.uci.edu/ml/datasets/statlog+\(german+credit+data\)](https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data))

Communities and crimes dataset

Theme: communities in the United States, their descriptors and common crimes such as robberies, murders or rapes.

Possible prediction task: level of crimes per community

Data size: about 2000 records

Data available from: <https://archive.ics.uci.edu/ml/datasets/Communities+and+Crime>

Firefighter promotion (small dataset, for one person only)

Theme: firefighter promotion based on oral and written examination scores

Possible prediction task: who will get a promotion

Data size: 120 records

Data description: https://en.wikipedia.org/wiki/Ricci_v._DeStefano

Dataset: available on Moodle