Group Project
Machine Intelligence and Society
31. July 2020

Formalities

Form groups of **3 to 5 students**. Use RocketChat to find others. If you have problems finding a others to work with please contact us **early**. It would be ideal if every group has at least 1 engineering and 1 political/social science student.

You have to create 2 files:

1. Report your key findings in a document. There is no template, you are free to use any software you like. The only restrictions are that the document consists of **not more than 5 pages** and is a .pdf file. We would like you to focus on **meaningful plots, graphics and tables** and only write **short and reasonable comments, observations and explanations**.

2. This project requires some programming to get your results and generate plots. Write your code in the form of a Notebook (.ipynb). The programming has to be done in Python 3. You are allowed to use any Python 3 packages.

Both, the document and the Notebook must contain the names of all group members. We only need one submission from one of your group members. Appoint one group member to be in charge of the final submission. Submit one .zip file that contains your document (.pdf), your Notebook (.ipynb) and the dataset. Check all the formalities again before your submission. The deadline is **30th of August 2020 at 23:59**.

Topic

The project is about **fairness in machine learning**. In the second homework you already got to know the **COMPAS** (Correctional Offender Management Profiling for Alternative Sanctions) recidivism risk score and a dataset containing informations to analyse the algorithm behind this score.

Your tasks are the following:

1. Which of the variables in the dataset are suitable to form groups that can be checked for fairness. Calculate **independence, separation and sufficiency** for those groups. You can find the definition of those 3 fairness measures in the lecture slides. What do you notice? (Hint: The significance of high differences is linked to a high number of samples available.)

2. Try to align the fairness measures by simply using different thresholds for each group. Are you able to get a fair model? What do you observe?

3. Generate a new binary classifier using a method of your liking. Similar to the COMPAS score, the task is to classify if a convict will be recidivistic. Exclude the variables you used to form groups from the training of your classifier. Check your classifier for fairness using your groups. Is it fair? Why/why not?

You are probably not able to include all your results. Focus on the ones you find most important and interesting. Generate meaningful visualisations, e.g. confusion matrices, ROC curves (Hint: Plots from different groups can probably be plotted in one figure using different colors and a legend).