Group Project
Machine Intelligence and Society
13. July 2021

Formalities

Form groups of **3 to 5 students**. Use RocketChat to find others. If you have problems finding others to work with please contact us as soon as possible. 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 **no more than 6 pages** and is provided as 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. The Notebook should be **well structured**.

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) and your Notebook (.ipynb). Check all the formalities again before your submission. The deadline is **31st of August 2021 at 23:59**.

Topic

The project is about **basic machine learning methods** and **fairness in machine learning**. We would like to see the whole process of building a model: preprocessing, training, cross-validation and testing. The provided dataset is a subset of the [North Carolina Policing Dataset](#). Get familiar with the dataset but only use the subset we provide.

Your tasks are the following:

1. The provided dataset has not been cleaned for you. It is your job to figure out which samples or features to drop, which to use and how to encode them.
2. The task is to classify if a person will be arrested (stop_outcome). Generate two binary classifiers using two different methods of your liking and compare their performances against a dummy classifier which always predicts not arrested.
3. Which of the features in the dataset are suitable to form groups that can be checked for fairness. Use the results of one of your classifiers to calculate **independence, separation and sufficiency** for those groups. What do you notice?
4. Exclude the features you used as sensitive characteristics from the training of the model. Check the new model for fairness. Is it fairer now? Why/why not?
5. 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?

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. density plots, confusion matrices, ROC curves (Hint: some plots from different groups can probably be plotted in one figure using different colors and a legend).