# Jupyter Notebook für exemplarische Klassifizierung

## Ziel:

Das Ziel sollte sein, dass die Studenten durch einfache Anpassungen im Code das Verhalten anpassen zu können. Der eigentlich Code sollte hinter Funktionen „versteckt“ sein.

### Beispiel:

Die Studies sehen nur das und können durch die Variable *feature\_names* Steuern welche Histrogramme erzeugt werden (Bild oben). Der eigentliche Code (Bild unten) mit der function *show\_distribution* wird bspw. aus einem file *utils.py* importiert.

A screenshot of a computer

Description automatically generated

A computer screen shot of a code

Description automatically generated

## Aufbau

Notebook sollte folgende große Phasen beinhalten:

* Analyze your data:
  + Laden des Datensatzes
  + Überblick verschaffen über
    - Art der Features
    - Fehlende Werte
    - Verteilungen der Features und Ausreißer
    - Zweitrangig: Art des Labels
  + Was sind interessante Fragestellungen?
    - Bias? Zusammenhänge?

(ToDo: Countplot mit weiteres Feature zur Unterteilen -> Bspw. mistriage nach Sex (<https://seaborn.pydata.org/generated/seaborn.barplot.html>)

Scatterplot Matrix seaborn

* Feature Selection und Pre-Processing:
  + Methoden zum Ersetzen fehlender Werte -> Scikit Learn Imputer
  + Transformieren von kategorischen Features -> Scikit Learn OneHot-Encoder
  + Transformieren von numerischen Features
    - Möglichkeit 1: Skalieren -> Scikit Learn StandardScaler, MinMaxScaler
    - Möglichkeit 2: Diskretisieren -> pandas cut() function
    - …
  + Feature Selektion:
    - Möglichkeit 1: Lasso Regression
    - Möglichkeit 2: PCA (Principal Component Analysis)
    - Möglichkeit 3: Korrelationsplot (Heatmap)
    - …
  + Umgang mit „Imbalanced Data“
    - Oversampling // nur für Klassifikation, aber mistriage hier [0; 0.5; 1]
    - Undersampling
    - SMOTE
    - Bootstrapping
* Train the Model
  + Split the Data into train and test set
  + Train the Model (let them choose one model)
* Evalluate the Model
  + Evaluate the Models Performance
  + Analyze its behaviour