Homework 3

The examination of the module *31-M29 Data Science* consists of a portfolio of four programming tasks and a final exam. This is the third programming task. Please hand in your solution via "LernraumPlus" by February 17, 2021.

Self-checkout fraud detection (22.5 Points)

Many retailers around the world offer self-checkout terminals that allow customers to scan items that they want to purchase themselves. These terminals reduce the number of required cashiers (and sometimes lead to shorter waiting times for the customers). However, they also increase the chance of customers not paying for items because they either forget to scan an item or don't do so intentionally. To identify these cases retailers can perform follow up checks. Obviously the number of checks should be kept as low as possible. Your task it to **develop and evaluate a model that classifies the scans of a purchase as fraudulent or non-fraudulent** to allow for targeted follow up checks.
This homework is based on the Data Mining Cup 2019 (`https://www.data-mining-cup.com/reviews/dmc-2019/`). Included with this task description you will find associated data sets as well as a description of all features. You should only use the training data for model training and the test data to evaluate the performance of your model. Please also note the following additional instructions:

- Hand in an *.ipython notebook* that contains the training and evaluation code. **When all cells are executed in sequential order ("Runtime" → "Restart and run all")** a new model should be trained and evaluated.
- Use the *Area Under the Receiver Operating Characteristic Curve* metric (ROC AUC) as an evaluation metric.
- Do not use the test set for parameter tuning to avoid overfitting to the test set.
- Working in groups is not permitted. Do not share your code with other students.
- Use text cells to describe and structure your code so that each step of the prediction generation is easy to understand.
- Generate *at least* 2 new features. Describe the 2 generated features that you deem the most important in 1-3 sentences.
- Based on the observed model performance argue in 3-4 sentences, if you think that that your model should be deployed in practice.

Grading will be performed based on the following criteria:

- **Pre-processing, training, evaluation: (6 Points)** Are all steps correctly and thoroughly implemented?

- **Prediction quality (5 Points):** How good are the generated predictions?
- **Generated features (4 Points):** Do the generated features make sense? Are they correctly implemented?
- **Code/Notebook quality (5.5 Points):** How well is the code/notebook structured? How difficult is it to understand the code (including whether or not there are comments)?
- **Interpretation of results (2 Points)**: Does the interpretation make sense based on the observed model performance?

Late Submission Policy

Late submissions are penalized 5% per started hour after the deadline (i.e., a solution that is handed in 122 minutes after the deadline is penalized 15%). Note that these should still be handed in through LernraumPlus.