Data Science Challenge

General Remarks

In this exercise, we want to see how you deal with a data science challenge involving our customer base. We have data on 10.000 (fictitious) customers of DKB, and want to understand which of these customers we will be able to retain. We want to use the insights you generate to improve our customer retention, and identify customers at risk of leaving DKB.

Your Task

- Do an explorative analysis of the given dataset.
- Build a predictive model. Given customer data, predict whether the customer will be retained over the next 3 months.
- From the customers currently retained, find the top k customers at highest risk of leaving DKB

You may spend as much time (until the submission deadline) on the challenge as you want. However, the task is not (primarily) to build the best, most predictive model, but rather to understand how you approach such a question and how you present your results.

Often times, a graph says more than a thousand words. Instead of providing long descriptive answers you are encouraged to visualize your results in a graph or chart and give a brief summary of your results.

We expect you will need around 3 to 4 hours of effort to produce a first iteration solution.

Evaluation

Your submission will be evaluated based on the following criteria:

- Clear communication of results
- Proficiency in machine learning methods
- Reproducible and well documented experiments
- Correctness
- Good Software Engineering practices
- Clean code and knowledge of APIs
- Good evaluation of results and discussion of the methods

Conditions

You can use any language or toolset to complete the challenge. Keep in mind that the main language in the DKB Data Science team is Python, and that your reviewer may not be familiar with your chosen language and frameworks. Please make sure that any results, code etc. you produce are not accessible publicly. You may make your results available to us either through email or via a non-public repository.