**Predictive Model for Treatment Success**

**What?**

This project is an attempt to predict the treatment’s probability of success for a given client, based on the information he filled in during the pre-intake phase exclusively. It ultimately leads to a classification problem where the goal of the model is to correctly predict the label of a client, the label being: treatment success (yes/no).

# **What data do we have?**

# The dataset contains all questionnaire outputs, the gender, the location and two sets of labels: started treatment (yes/no) and finished treatment (yes/no).

# **What are the issues with that data?**

# For our use case, it is important to notice that a percentage of data is not labeled (~40%) since, for the people that have not started the treatment, we can’t possibly know whether they would have finished it or not. This leads to a well-known issue called selection bias. Selection bias causes a classifier to be biased towards people that were selected and generalizes poorly to (our) general population.

The dataset also contains more positive than negative classes, meaning that many people are selected (~70%) and many of those finish the treatment (~70%).

# Therefore, it is very important to take these two characteristics into account for the modeling.

# For the sake of this project I developed a bias correction framework which implements several bias correction techniques.

The first set of labels is used to model the selection process. That is, we want to predict the probability of being selected for the treatment. The probability output of that classifier is used to compute a correction factor for the second classifier which outputs a probability for a client to finish the treatment.

The second classifier is iteratively improved with a technique called self-training.

Thinking points:

* Do we want to reduce false positives (predicted finished, actual drop-out ) or false negatives (predicted drop-out, actual finished)?