Predicting user motivation and retention. E.g., dropout or long pauses prediction, level checks prediction (self-supervised learning)

We plan to use users only present in the events table, because they contain more precise information about user behavior, and it is more consistent than the transactions table. We will separate user behavior in a time series to analyze user activity and dropout probability on a weekly basis.

We will decide the labels algorithmically by analyzing the input data and user activity. Then we will train a classification model: we can use a neural network for its power, but we also thought about using a decision tree to have explainable predictions.

We will feed it the first few weeks of the users and it will predict one of the categories (dropout, long pause, etc...) and compare with the assigned categories.

We are not sure about our choice of the model, we feel it would be best if the prediction can be explained, but we don't know if it will be robust enough for our task.