

The estimate file contains a number that we estimate to be the closest to the real number of correctly predicted projects in the test dataset. We have created this number based on the certainty of the probability of the best-chosen model (see the main notebook.) The "predict_proba" function returns how certain the model is that the label is funded. The threshold lies at 0.5. Our estimate is based on the idea that the closer the probability to 0.5 is the less chance it has to be correctly classified. Hence, we decided to add a margin of error of 5%. Therefore, the estimated number is the number of rows with a prediction value above or equal to 0.53 and predictions below or equal to 0.47 divided by the number of data points (rows) of the test dataset. (the current one: 56858)