

Sick dataset analysis 2

Mateusz Polakowski

26 04 2020

Introduction & assignment 1. catch-up

Here I provide some basic data engineering, based mainly on features reduction, which explicitly had an impact on previous results (only so-called white boxes). I deliberately omit EDA chapter, as it was a part of a previous assignment.

	auc.train	auc.test	auprc.train	auprc.test
rpart_tuned	0.9855226	0.9848538	0.9573875	0.9576183

Black-boxes modelling

Below I tried to use h2o automl functionality to find the best model in black-boxes family. Results present as:

	auc.train	auc.test	auprc.train	auprc.test
rpart_tuned	1	0.9985323	0.9831461	0.9526022