Reikalavimai II daliai:

1. Duomenų imties pažinimas: ydata-profiling.html, bet ir grafikų apie Y - histograma (ar yra klasių disbalansas? <15%, gal kažką daryti train duomenims), X histogramos pagal Y.
Kas buvo daryta su kategoriniais kintamaisiais, kada detektorius nepalaiko factor tipo kintamųjų? -> dummy

2. Stratified cross validation (outer loop + inner loop 4 tuning)

3. Detektorių pasirinkimas - bent 5, pvz. random forest (ntrees=200, importance=TRUE) būtinai.

4. Disbalansui: MD, CSL, SMOTE. Detektoriams: inner CV (tuning).

5. ROC (dar dailiau DET), Precision-Recall (ypač jei disbalansas), pagalvoti apie grafikus verslui - Profit (Lift, CAP), confusion matrix (bent jau geriausiam modeliui) - https://marcovanetti.com/pages/cfmatrix/

6. Išvados: kas laimėjo?, palyginti AUC (su easyROC Multiple Comparisons),
kintamųjų svarba (verslo įžvalgom) nuo RF/xgboost(importance=TRUE) iki SHAP grafikų ir t.t.