## Klasyfikacja Artificial

Adam Majczyk I Szymon Matuszewski Praca domowa 2 AutoML

## Spis treści

01

**Selekcja** predykatorów

03

**Auto**Gluon

02

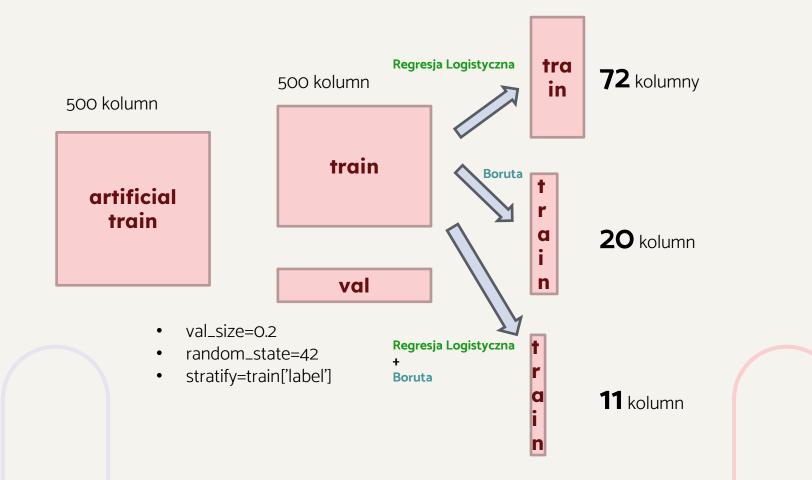
Optymalizacja

04

Wyniki

## 01

# **Selekcja** predykatorów



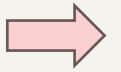
#### Regresja Logistyczna



LogisticRegression



CatBoostClassifier





RandomForestClassifier



**LGBMClassifier** 

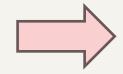


**XGBClassifier** 

500 kolumn



Dla każdej z **500** Kolumn model z jedną zmienną



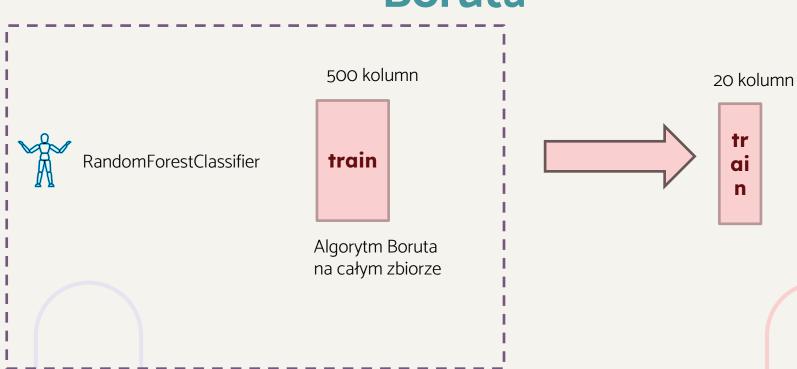
Wybór kolumn, gdzie **balanced** accurary > 0.5

72 kolumny

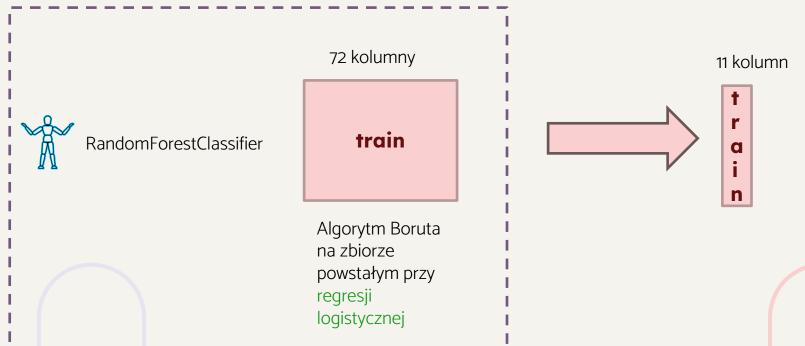


Wybór przecięcia podzbiorów kolumn dla każdego modelu

#### **Boruta**

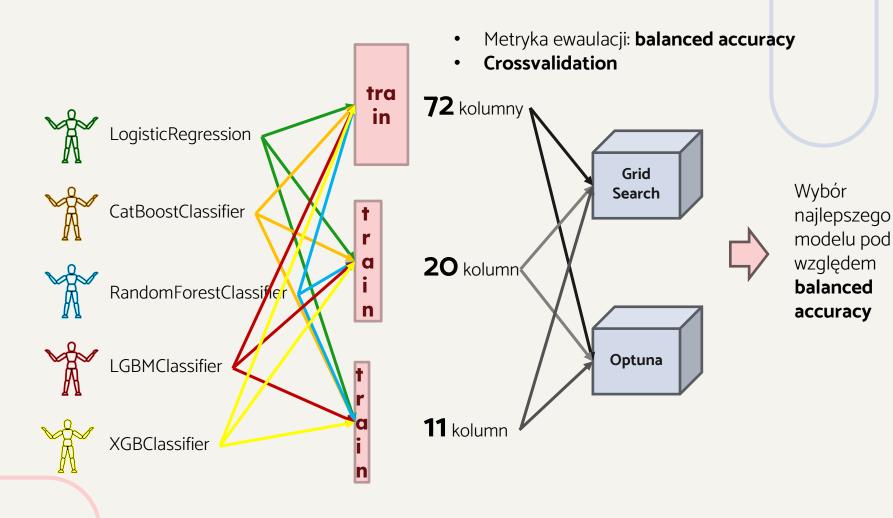


### Regresja Logistyczna + Boruta



# 02

## Optymalizacja



# 03

#### **Auto**Gluon

500 kolumn

artificial train

500 kolumn

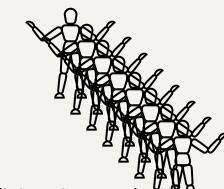
train

val

- val\_size=0.2
- random\_state=42
- stratify=train['label']

**Auto**Gluon





- Eval\_metric='balanced\_accuracy'
- Train\_data=**train**
- Tuning\_data=val
- Time\_limit=**600**

# 04 Wyniki

#### AutoGluon – zbiór walidacyjny

model	balanced_accuracy	czas trenowania [s]
WeightedEnsemble_L2	0.8525	15.7922
CatBoost	0.8475	2.6999
LightGBM	0.8350	2.7573
LightGBMLarge	0.8350	10.1192
LightGBMXT	0.7700	2.6904
KNeighborsUnif	0.7225	2.8266
KNeighborsDist	0.7225	0.0261
RandomForestEntr	0.7025	0.6494
RandomForestGini	0.6525	0.7759
ExtraTreesEntr	0.6425	0.3624
ExtraTreesGini	0.6100	0.3701
NeuralNetTorch	0.5875	4.7112
NeuralNetFastAI	0.5650	2.4301

Tabela 1: Wyniki dla automatycznego uczenia maszynowego

# Model **manualny** – zbiór walidacyjny

- model: catboost
- parameters selection: boruta
- Optimization\_method: optuna
- 'iterations': 322
- 'learning rate': 0.15542656215841885
- 'depth': 8
- '12\_leaf\_reg': 9
- balanced\_accuracy: 0.92

# Dziękujemy za uwagę!