



# Modelowanie

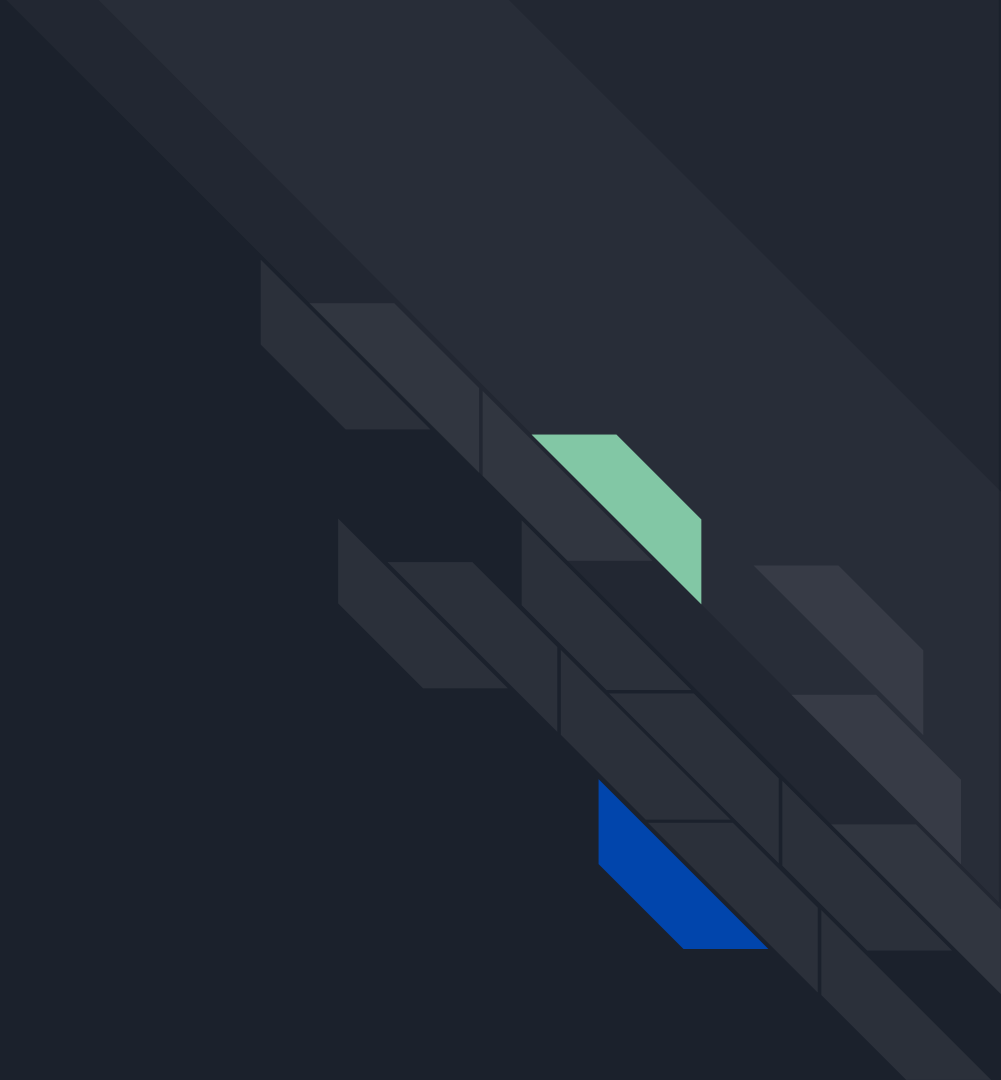
*Szymon Gut, Wiktor Jakubowski, Maciej Orłowski, Maria  
Kędzierska*



# Zastosowanie PCA

RMSE	Random Forest	XGBoost	LightGBM	CatBoost
PCA	87 740.23	360.06	74 396.47	46 023.39
Bez PCA	334 091.74	228890.90	282 358.90	246 424.10

Modelle





# Optymalizacja hiperparametrów

- Testowanie parametrów domyślnych, random search, autorska eksploracja
- Minimalizacja pierwiastka błędu średniokwadratowego
- Niwelowanie overfittingu i underfittingu



# Porównanie modeli

<i>Najlepsza wartość z każdej grupy</i>	<b>XGBoost</b>	Random Forest	LightGBM	CatBoost
RMSE test	<b>360.07</b>	87 740.23	74 396.56	46 023.39
RMSE train	<b>101.90</b>	38 723.22	75 621.56	59 582.94



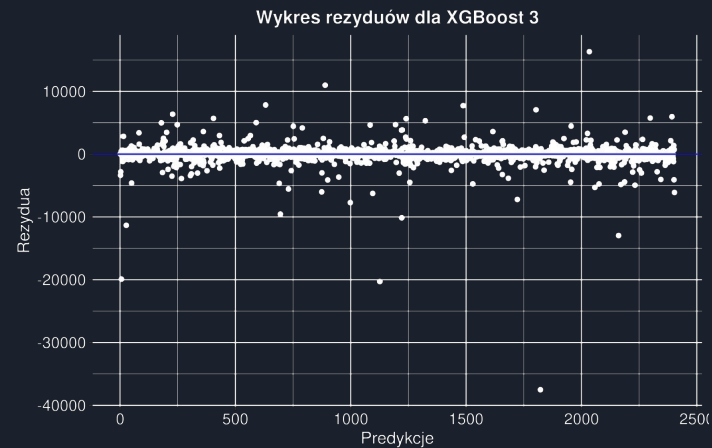
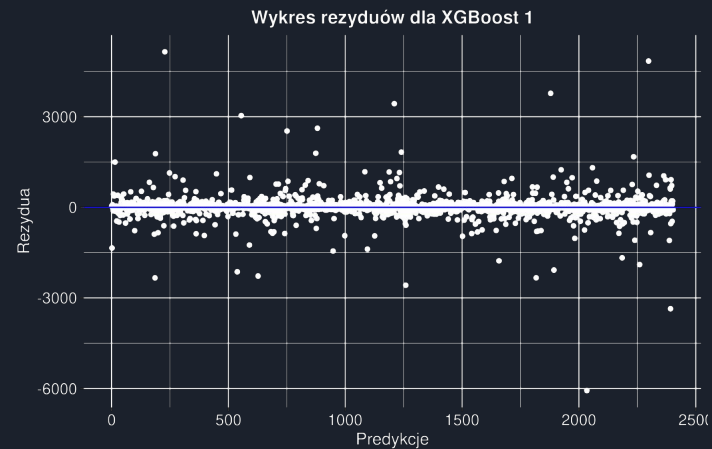
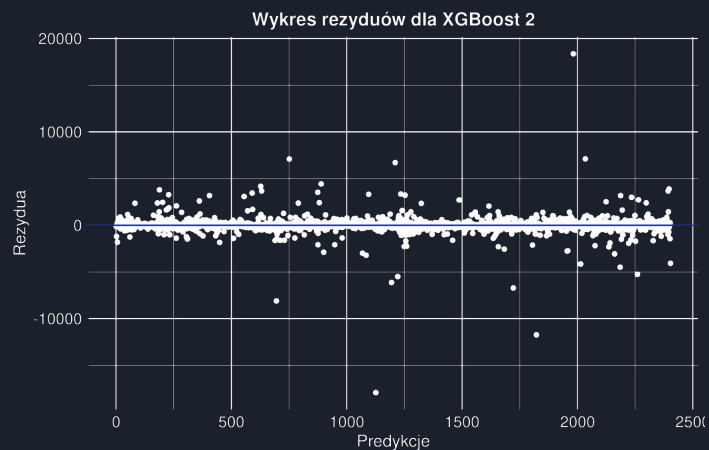
# Top 3 modele

Nazwa modelu	RMSE train	RMSE test	booster	eta	max_depth	gamma	min_child_weight	subsample	colsample_bytree
xgboost	101.90	360.07	gbtree	0.05	7	0	1	1	1
xgboost	430.97	881.18	gbtree	0.09	8	4	1	1	1
xgboost	808.91	1 477.11	gbtree	0.12	6	6	1	1	1

Diagnostyka



# Wykres rezyduów







# Metryki

Nazwa modelu	RMSE test	MAE test	$R^2$
xgboost1	360.07	135.74	0.9998625
xgboost2	881.18	289.28	0.9995856
xgboost3	1 477.11	520.25	0.9978164



Dziękujemy za uwagę!