

# Warsztaty Badawcze - projekt

Phase 2

# Przypomnienie zastosowanej metody

- Stworzenie 5 modeli przewidujących wartość AUC na podstawie bazy modeli.
- Predykcja na zbiorze testowym z każdym modelem
- Uśrednienie wyników
- Wybór najlepszego modelu
- Dobranie najlepszych parametrów na podstawie danych z bazy modeli

	<b>models_names</b>	<b>response</b>
<b>1</b>	classif.rotationForest	0.8828331
<b>2</b>	classif.cforest	0.8731673
<b>3</b>	classif.ranger	0.8724117
<b>4</b>	classif.bartMachine	0.8693313
<b>5</b>	classif.randomForestSRC	0.8686288
<b>6</b>	classif.ada	0.8684296

# Zmiany w fazie 2

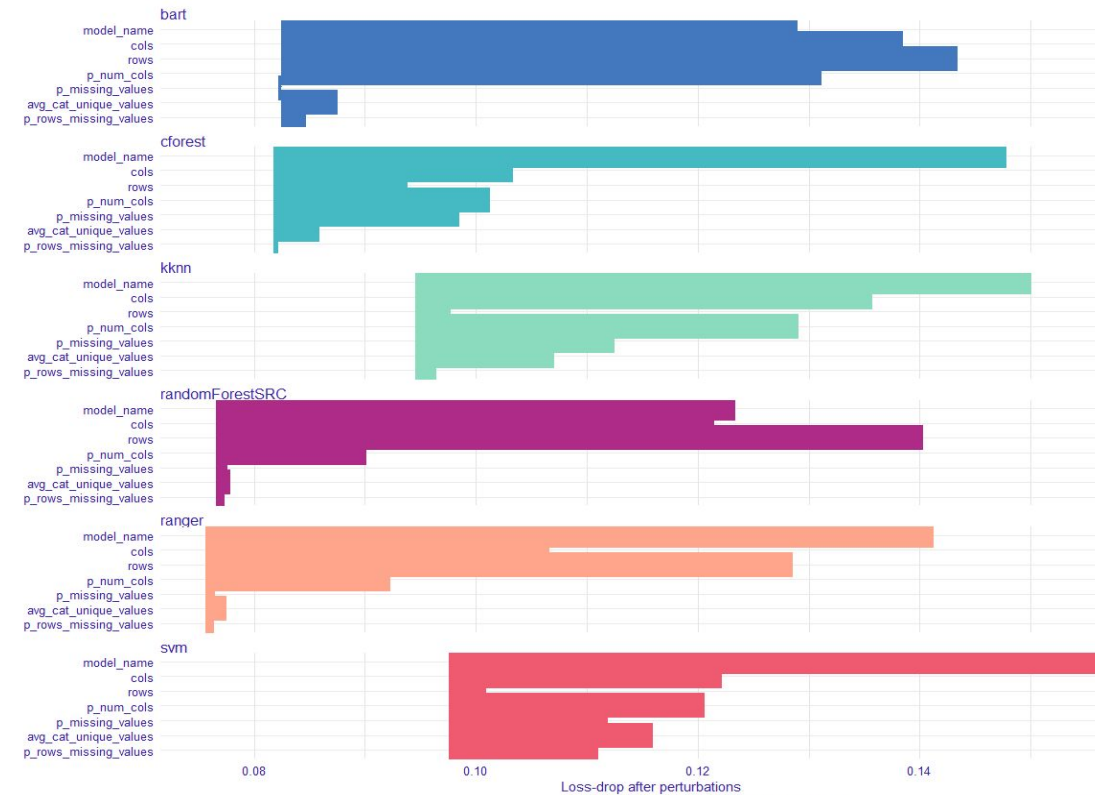
- Przerobienie zbioru testowego
- Dodanie szóstego modelu (SVM)
- Średnia ważona po mierze  $R^2$  (przeskalowanej do 0,5-1)
- Dobór hiperparametrów

# Przerobienie zbioru testowego

- Usunięcie zduplikowanych wierszy i kolumn
- Zamienienie zmiennych z małą ilością unikalnych wartości na factory

	L. wierszy	L. kolumn	% NA	% wierszy z NA	% numerycznych kolumn	Śr. liczba unikalnych wartości w kolumnach
Przed	4184	22	0	0	1	0
Po	3951	16	0	0	0.5625	5.285714

# Ważność zmiennych w używanych modelach



# Średnia ważona po mierze $R^2$

Model	ranger	cforest	bartMachine	RandomforestSRC	kkn	svm
waga	0.95	0.85	0.85	1	0.67	0.5

# Nowy optymalny model

## classif.ranger

	models_names	response
1	classif.ranger	0.8937980
2	classif.RRF	0.8789533
3	classif.ada	0.8760775
4	classif.logreg	0.8760379
5	classif.bartMachine	0.8675292
6	classif.randomForest	0.8673176

# Dobór parametrów

Tunability: Importance of Hyperparameters of Machine Learning Algorithms  
by Philipp Probst, Anne-Laure Boulesteix and Bernd Bischl

```
num.trees=983
```

```
replace=FALSE
```

```
sample.fraction=0.703
```

```
mtry=floor(ncol(data_test)*0.257)
```

```
respect.unordered.factors='ignore',
```

```
min.node.size=1
```

Mao Mo De Lee