



Rashomon ML

Katarzyna Woźnica

Modelowanie predykcyjne i uczenie maszynowe



Przykład: wiemy że cena samochodu zależy od jego marki, modelu i roku produkcji. Na podstawie tych cech możemy próbować zgadnąć cenę samochodu, który chcemy wystawić na sprzedaż. Ale ta cena zależy także od i wyposażenia, historii wypadkowej, przebiegu itd.

Modelowanie predykcyjne i uczenie maszynowe

Zjawiska które chcemy modelować mają skomplikowaną strukturę i stworzenie takich prostych reguł dobrze znajdujących wzorce może być bardzo trudne.

Tu z pomocą przychodzą komputery i algorytmy uczenia maszynowego.



<https://mirosławmamczur.pl/czym-jest-uczenie-maszynowe-i-jaki-e-sa-rodzaje/>

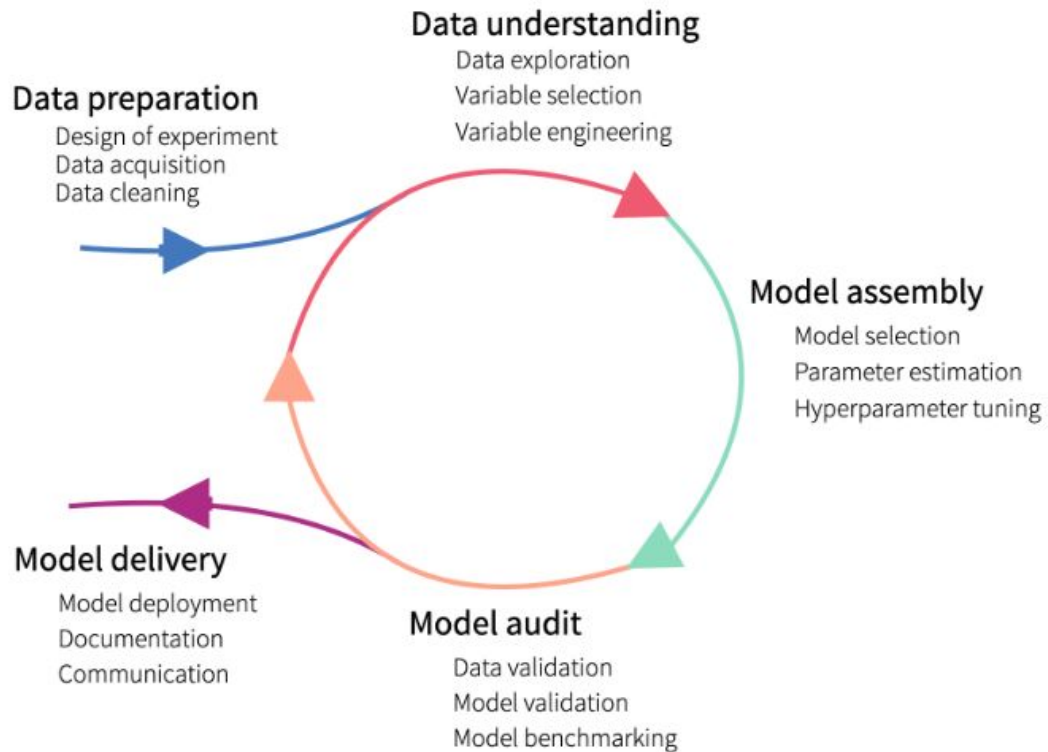
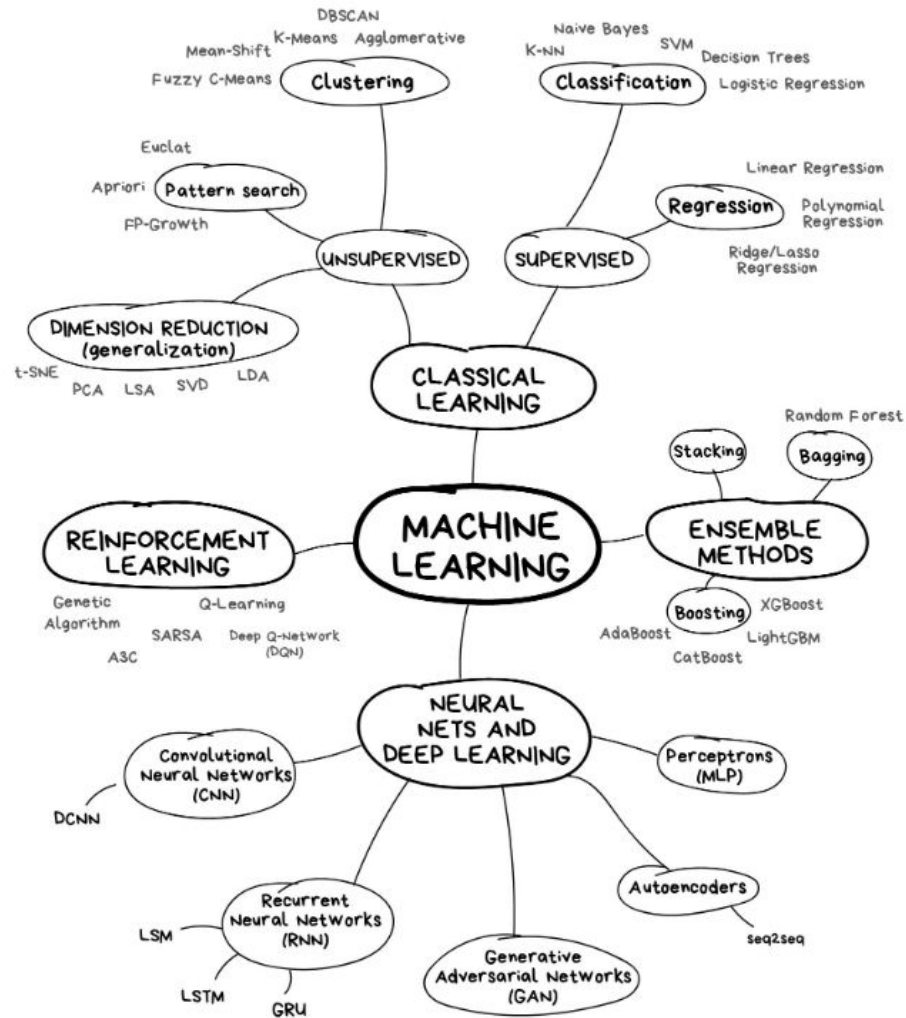
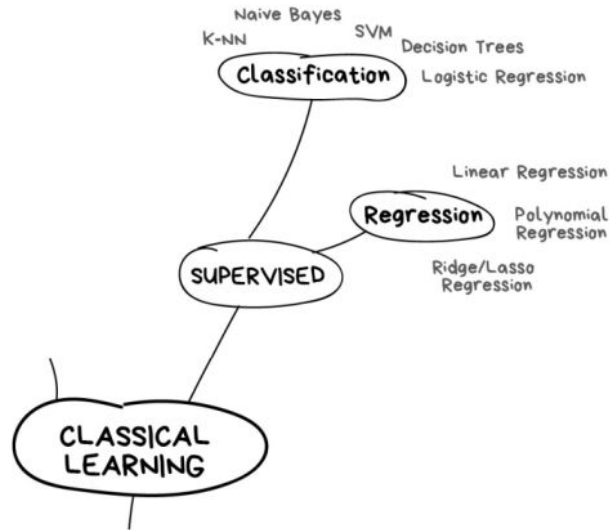


Figure 2.1: The lifecycle of a predictive model.



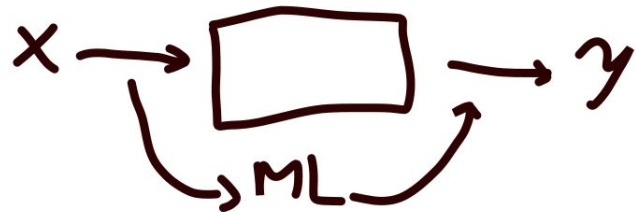
Który model wybrać?



Zwykle nie znamy prawdziwego mechanizmu, który stoi za danymi, dla których budujemy model predykcyjny.

Wybierając określony rodzaj algorytmu decydujemy się na określoną strukturę poszukiwania wzorców w danych, określony sposób *uproszczenia*.

Zwykle nie jest tak, że mamy pełną wiedzę o modelowanym zjawisku, mówimy że zjawisko jest losowe. Każdy model predykcyjny będzie popełniał błędy.



Który model wybrać?

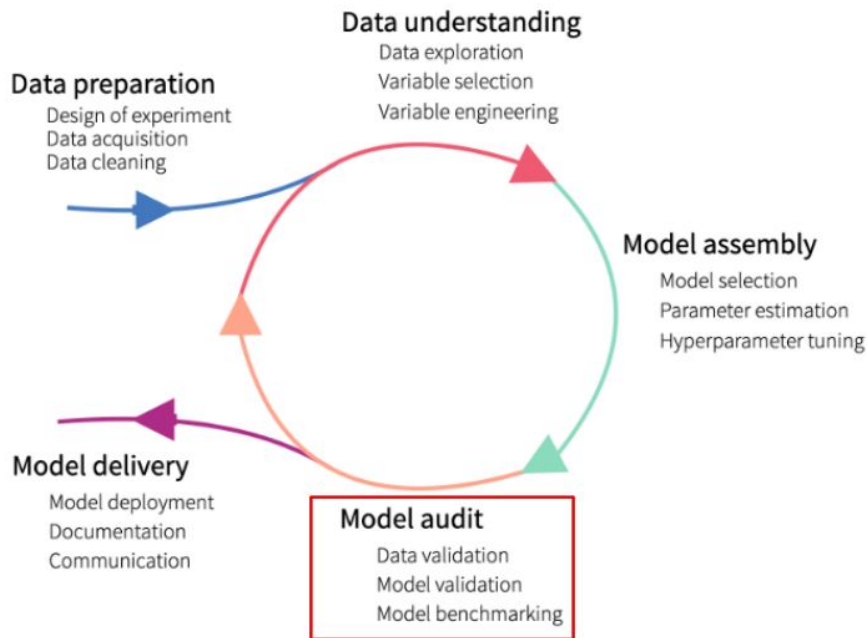


Figure 2.1: The lifecycle of a predictive model.

Ocena jakości
modelu jako miara
dokładności
predykcji (accuracy)

Ocena modeli ML w medycynie (i nie tylko)

Table 11
In-hospital mortality task on MIMIC-III using feature set B.

Method	Algorithm	Feature Set B, 24-h data		Feature Set B, 48-h data	
		AUROC score	AUPRC score	AUROC score	AUPRC score
Super Learner	SL.glm	0.7745 ± 0.0055	0.3134 ± 0.0112	0.7869 ± 0.0015	0.3103 ± 0.0212
	SL.gbm	0.8381 ± 0.0057	0.4059 ± 0.0156	0.8398 ± 0.0044	0.3932 ± 0.0155
	SL.nnet	0.8170 ± 0.0036	0.3650 ± 0.0124	0.8232 ± 0.0074	0.3591 ± 0.0135
	SL.ipredbag	0.7641 ± 0.0070	0.3127 ± 0.0085	0.7627 ± 0.0118	0.3011 ± 0.0140
	SL.randomforest	0.7582 ± 0.0080	0.3100 ± 0.0116	0.7604 ± 0.0042	0.2895 ± 0.0138
	SuperLearner-II	0.8426 ± 0.0068	0.4160 ± 0.0136	0.8471 ± 0.0036	0.4055 ± 0.0155
Deep learning	MMDL	0.8730 ± 0.0065	0.4765 ± 0.0109	0.8783 ± 0.0037	0.4706 ± 0.0178

Takie rozumowanie często prowadzi do wyboru jako najlepszy model takiego, który na wybranych danych testowych ma największą dokładność. **Ale czy najbardziej dokładny model będzie dawał nam najlepszą interpretację zjawiska?**

Rashomon Effect

So here are three possible pictures with RSS or test set error within 1.0% of each other:

Picture 1

$$y = 2.1 + 3.8x_3 - 0.6x_8 + 83.2x_{12} \\ - 2.1x_{17} + 3.2x_{27},$$

Picture 2

$$y = -8.9 + 4.6x_5 + 0.01x_6 + 12.0x_{15} \\ + 17.5x_{21} + 0.2x_{22},$$

Picture 3

$$y = -76.7 + 9.3x_2 + 22.0x_7 - 13.2x_8 \\ + 3.4x_{11} + 7.2x_{28}.$$

Which one is better? The problem is that each one tells a different story about which variables are important.

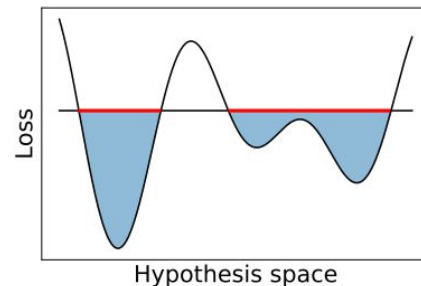
Każdy z tych modeli opisuje to samo zjawisko ale sposób w jaki je opisuje może być różny:

- może wybierać różne zmienne do modelu
- w zależności od tej samej zmiennej model 1 może być rosnący a inny nie musi być monotoniczny.

Recepta: nie musimy wybierać jednego dobrego modelu tylko wykorzystać to co wszystkie z nich mówią nam o tym zjawisku.

Rashomon set

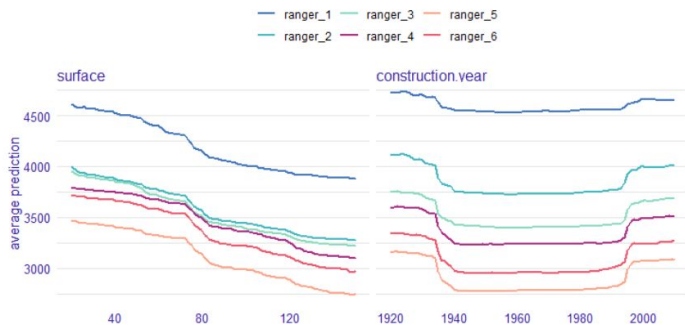
To address these concerns, we analyze **the set of prediction models that provide near-optimal accuracy**, which we refer to as a Rashomon set. (...) Additionally, even if the candidate models do not contain the true data generating process, we may hope that some of these models function in similar ways to the data generating process.



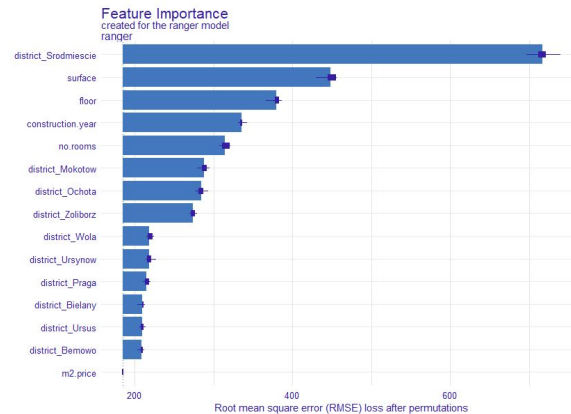
(b) The Rashomon set

Analiza struktury modelu - XAI

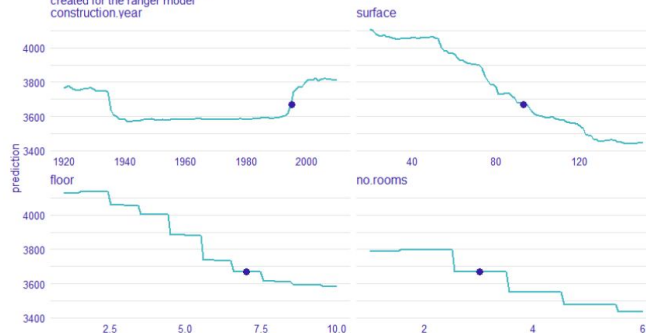
Partial Dependence profile



plot(rf_mprofile_group)



Ceteris Paribus profile
created for the ranger model
construction year



Ceteris Paribus profiles

Czym my się będziemy zajmować?

1. Dla wybranego problemu medycznego odtworzycie/ zbudujecie model predykcyjny (wybierzecie wcześniej algorytm ML)
 2. Dla wybranego problemu medycznego i algorytmu ML manipulując parametrami algorytmu zbudujecie zbiory Rashomon
-
3. Będziecie analizować otrzymane zbiory Rashomon:
 - a) dla hierarchicznej rodziny modeli zbudujecie krzywe Rashomon.
 - b) używając technik XAI sprawdzimy czy modele ze zbioru Rashomon dają nam inną interpretację opisywanego zjawiska.

Dane

MIMIC:

- <https://mimic.physionet.org/>

- MIMIC-III is a large, freely-available database comprising deidentified health-related data associated with over 40,000 patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012.

- a relational database consisting of 26 tables.

Inne dane Physionet: <https://physionet.org/>

- <https://physionet.org/content/hirid/1.1.1/>

Plan prac

Luty: request o dostęp do bazy danych (czas oczekiwania ~ 2 tyg.)

Marzec: zapoznanie się z bazą danych, artykułami i odtworzenie wybranego modelu

Kwiecień: zbudowanie zbioru Rashomon i ustalenie pod jakim kątem go będziecie analizować

I połowa maja: analizowanie zbiorów Rashomon i przygotowanie ostatecznych wyników prac

II połowa maja: przygotowanie artykułu i prezentacji