



Uczenie maszynowe w medycynie

Warsztaty PowerR 2022 x deepsense.ai

22.03.2022

Agenda

1. Kim jesteśmy?
2. O deepsense.ai
3. Przykłady uczenia maszynowego w medycynie
4. Czym jest uczenie maszynowe?
5. Wyzwania uczenia maszynowego w analizie obrazów medycznych
6. Odpowiedzialność
7. Case Study

Kim jesteśmy?



Maciej Domagała



Adam Maciaszek

We are data science experts

delivering AI-driven competitive edge
for global leaders across industries

>100 commercial AI projects in the US and Europe.

>20 R&D projects i.a. with Intel and Google Brain on topics involving reinforcement learning and generative adversarial networks.

Winning team of over 70 world class AI/ML experts - data scientists and data engineers supported by software engineers.

Top notch capabilities in: predictive analytics, computer vision and natural language processing.



Allianz

Santander

StateFarm

TAXUS • IT
INNOVATION

Google



bmc



L'ORÉAL

JUNIPER
NETWORKS



nielsen

BCG
BOSTON
CONSULTING
GROUP

NVIDIA

IBM



HEXAGON



HITACHI
Inspire the Next

Global tyre manufacturer

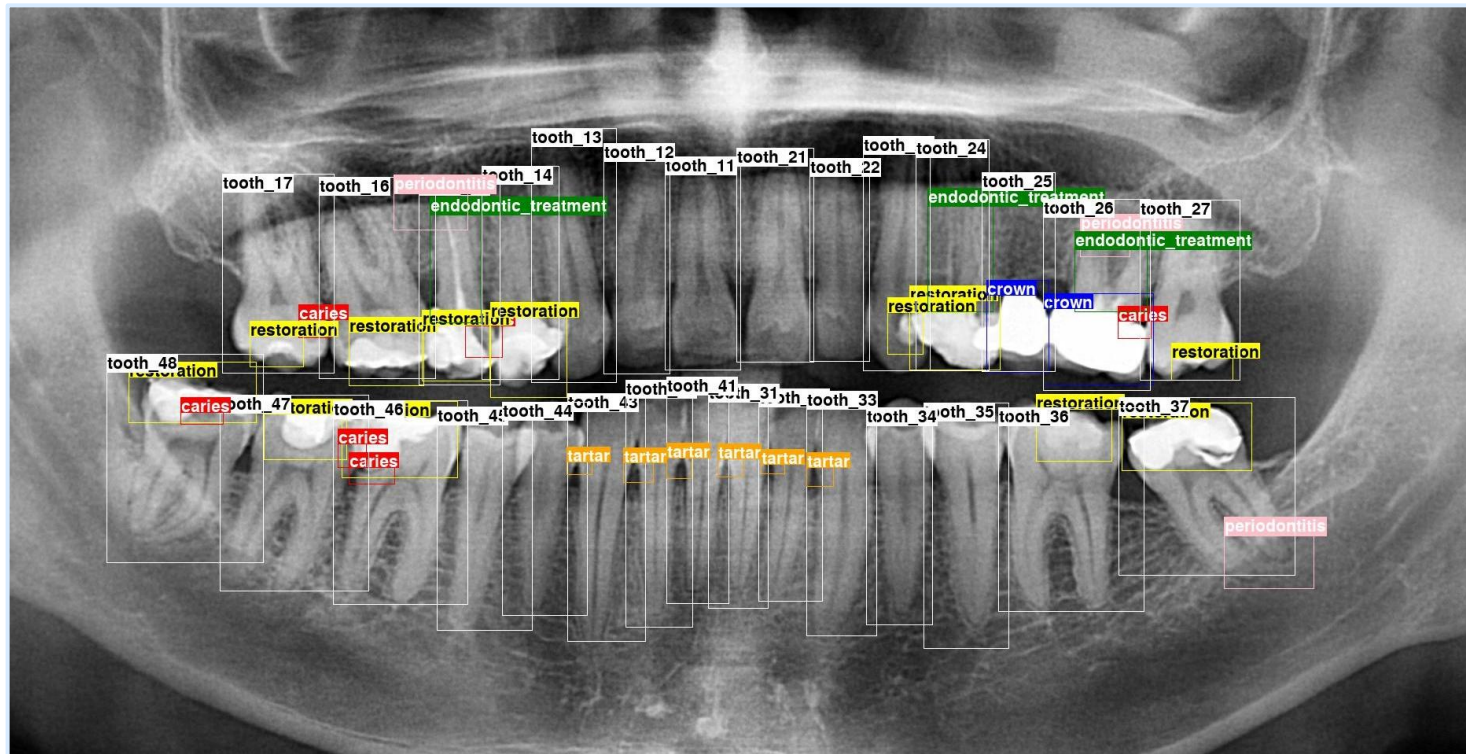
Global industrial player

Leading CEE fashion retailer

Leading CEE CPG player

Przykłady uczenia maszynowego w medycynie

Detekcja zębów oraz zmian chorobowych na zdjęciach pantomograficznych



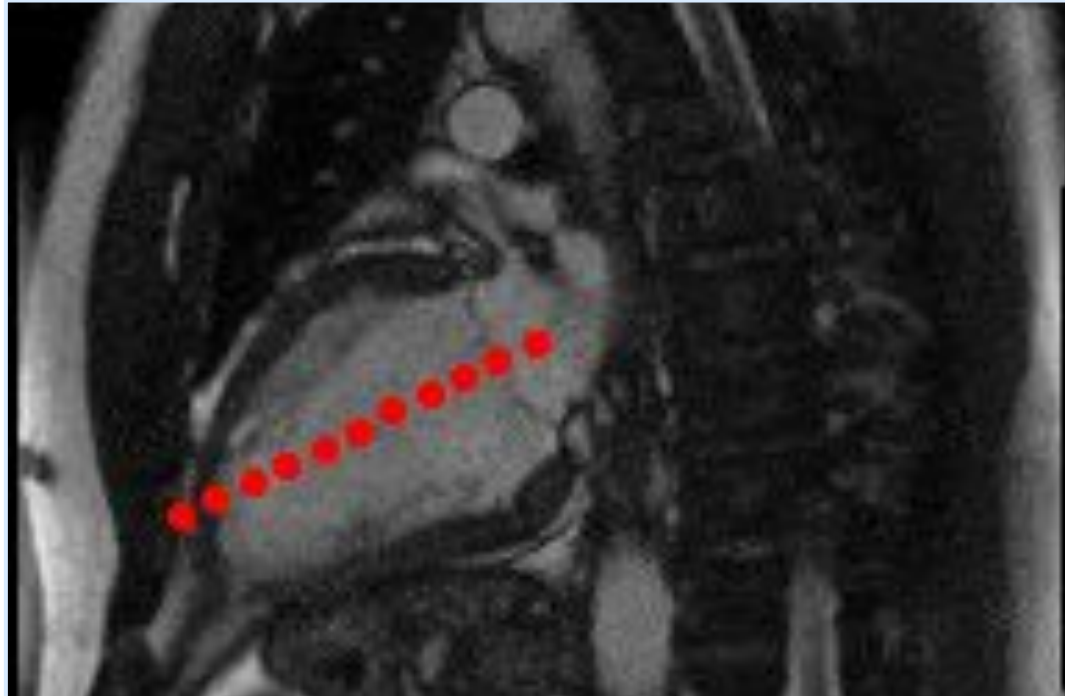
Wykrywanie retinopatii cukrzycowej na zdjęciach dna oka



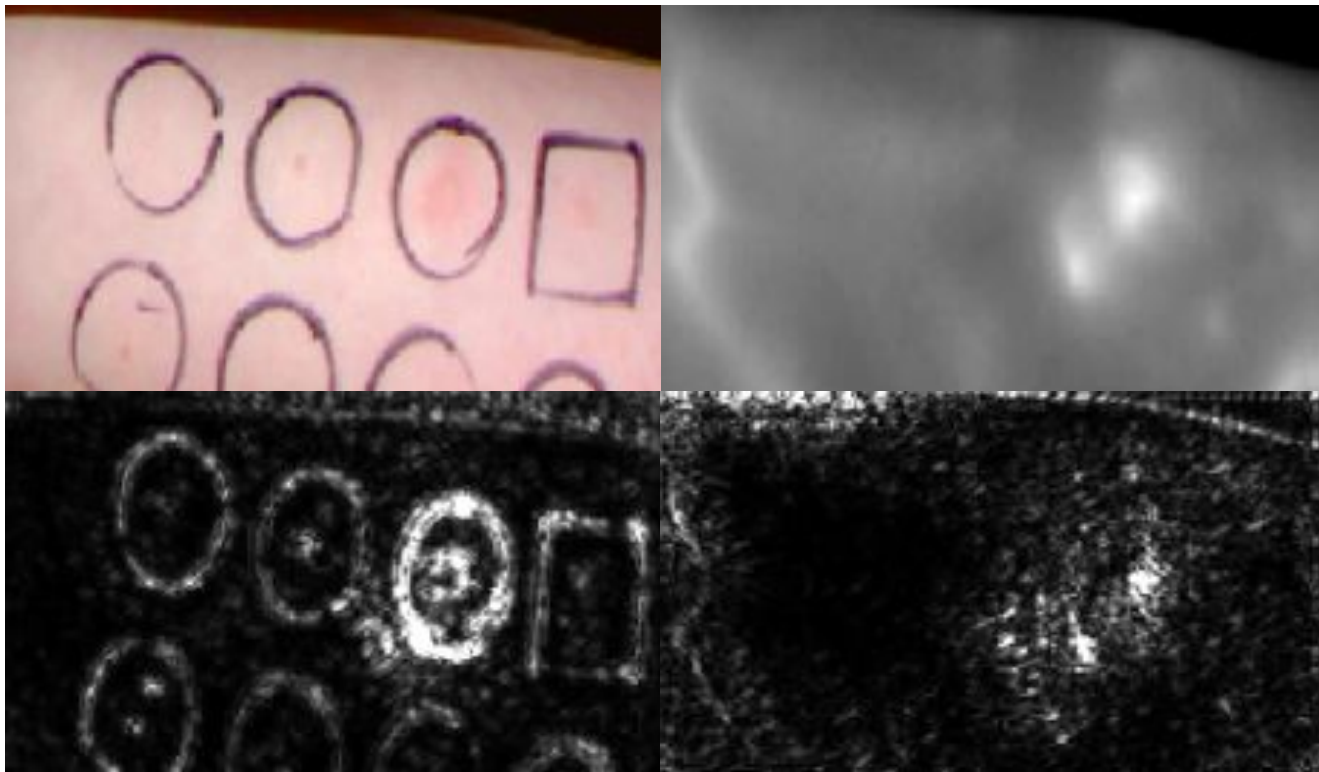
brak zmian retinopatycznych

zmiany retinopatyczne

Obliczanie objętości końcoworozkurczowej i końcowoskurczowej na animacjach z obrazowania rezonansem magnetycznym



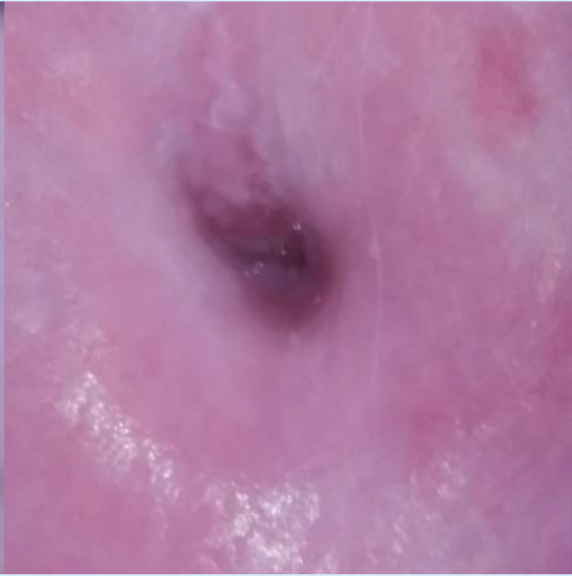
Ocena odczynu alergicznego w punktowych testach skórnych w paśmie widzialnym i podczerwieni



Klasyfikacja ujścia szyjki macicy na podstawie zdjęć z kolposkopii



Typ 1



Typ 2



Typ 3

Czym jest uczenie maszynowe?

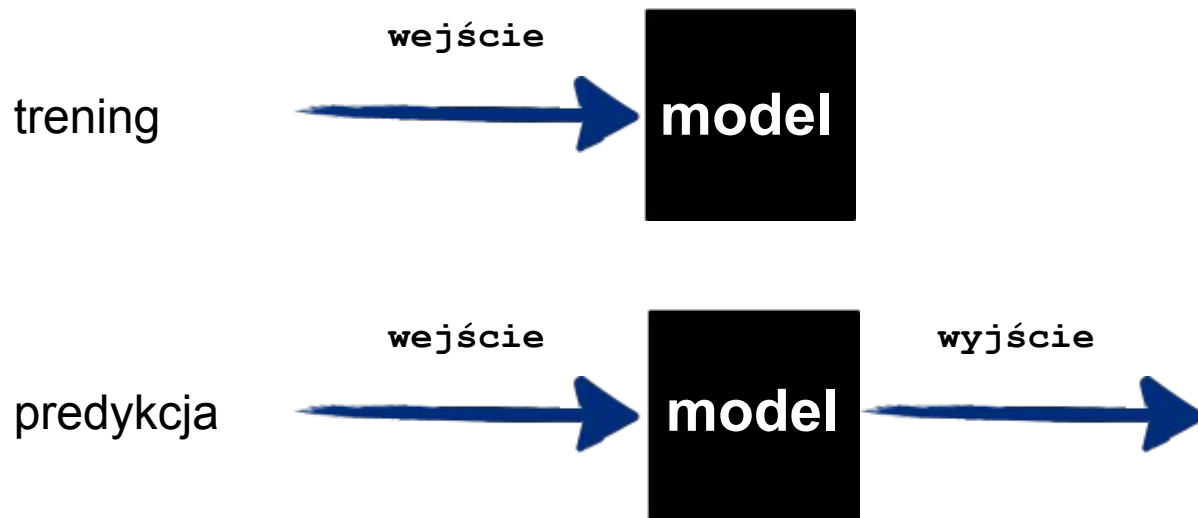
Tradycyjny algorytm



Inżynierowie decydują jak przekształcić wejście w wyjście.

Uczenie maszynowe

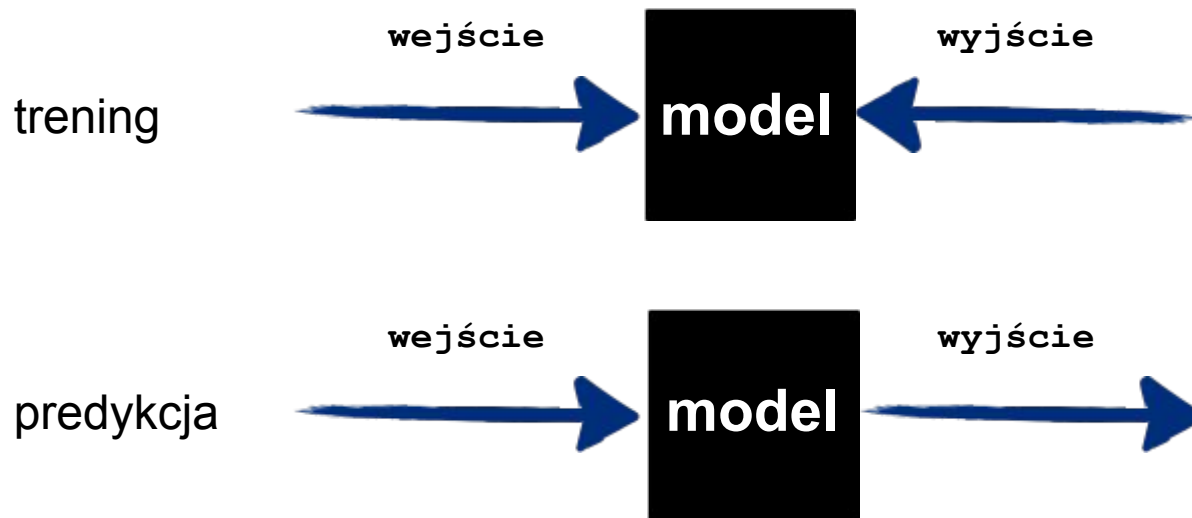
Schemat uczenia bez nadzoru



Trenujemy model korzystając ze zbioru uczącego

Uczenie maszynowe

Schemat uczenia z nadzorem



Trenujemy model korzystając ze zbioru uczącego oraz etykiet

Zalety i wady uczenia maszynowego

Zalety:

- + bardzo duża dokładność rozwiązania
- + modele nie są ograniczone naszą wiedzą
- + stosowane gdy danych jest za dużo do ręcznego przetwarzania
- + oszczędność czasu i pieniędzy

Wady:

- potrzeba dużo dobrej jakości danych
- proces etykietowania jest kosztowny
- skomplikowany model może działać wolno
- bardzo często dobre modele są trudne do zinterpretowania



Wyzwania dla uczenia maszynowego w analizie obrazów medycznych

Wyzwania

1. Mała ilość danych / występowanie “uprzedzeń” (ang. bias) w danych
2. RODO i wyzwania dotyczące ochrony danych osobowych
3. Proces etykietowania danych jest kosztowny, trwa długo i jest **trudny**
4. Każdy człowiek ma inne ciało
5. Kosztowny i złożony prawnie proces wprowadzenia produktu na rynek
6. Brak zaufania ze strony środowiska medycznego
7. Medycyna to ryzykowna dziedzina - reperkusje spowodowane błędem modelu są bardziej dotkliwe

Odpowiedzialność

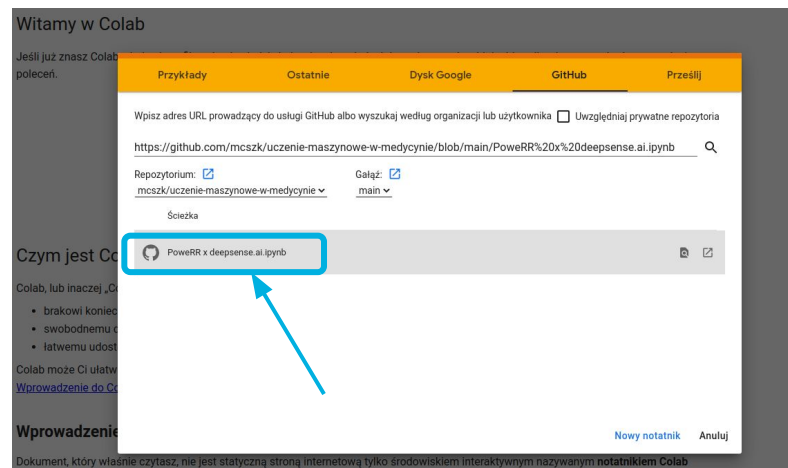
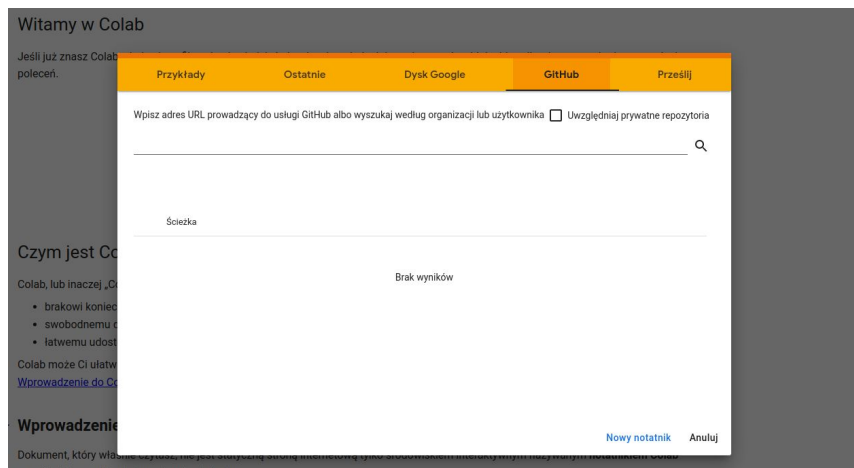
Założmy, że rozwiązanie oparte na sztucznej inteligencji dokonuje błędnej predykcji. Kto jest odpowiedzialny za ten błąd?

1. Ten, kto zapewnił dane do trenowania algorytmu?
2. Autor modelu?
3. Autor produktu?
4. Właściciel praw autorskich produktu?
5. Lekarz stosujący model?
6. Jednostka w której lekarz pracuje?
7. Pacjent?

Warsztaty - wykrywanie ryzyka udaru

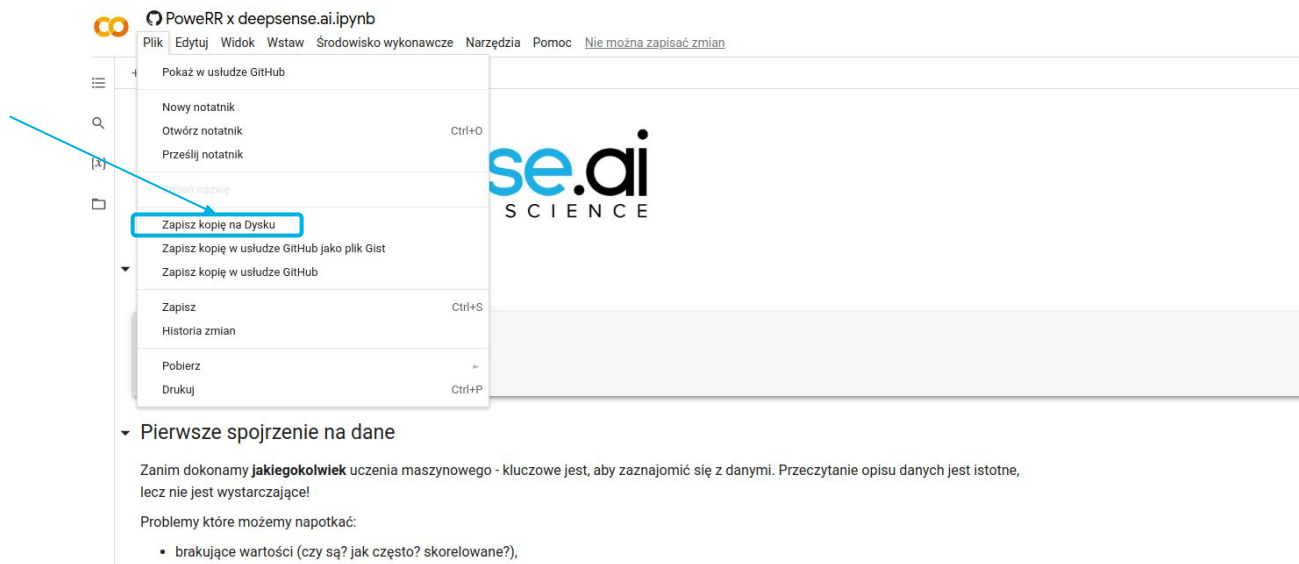
Stworzenie środowiska:

- Wejdź w link: <https://colab.research.google.com>
- Jeśli nie jesteś zalogowany na koncie google - zaloguj się!
- Otwórz zakładkę *GitHub* i wklej podany adres. Następnie otwórz notebook:
https://github.com/mcszk/uczenie-maszynowe-w-medycynie/blob/main/PowerR_x_deepsense_ai.ipynb



Warsztaty - wykrywanie ryzyka udaru

- Po otwarciu notebooka, utwórz kopię na swoim Dysku Google, żeby móc zapisywać wprowadzone przez siebie zmiany:





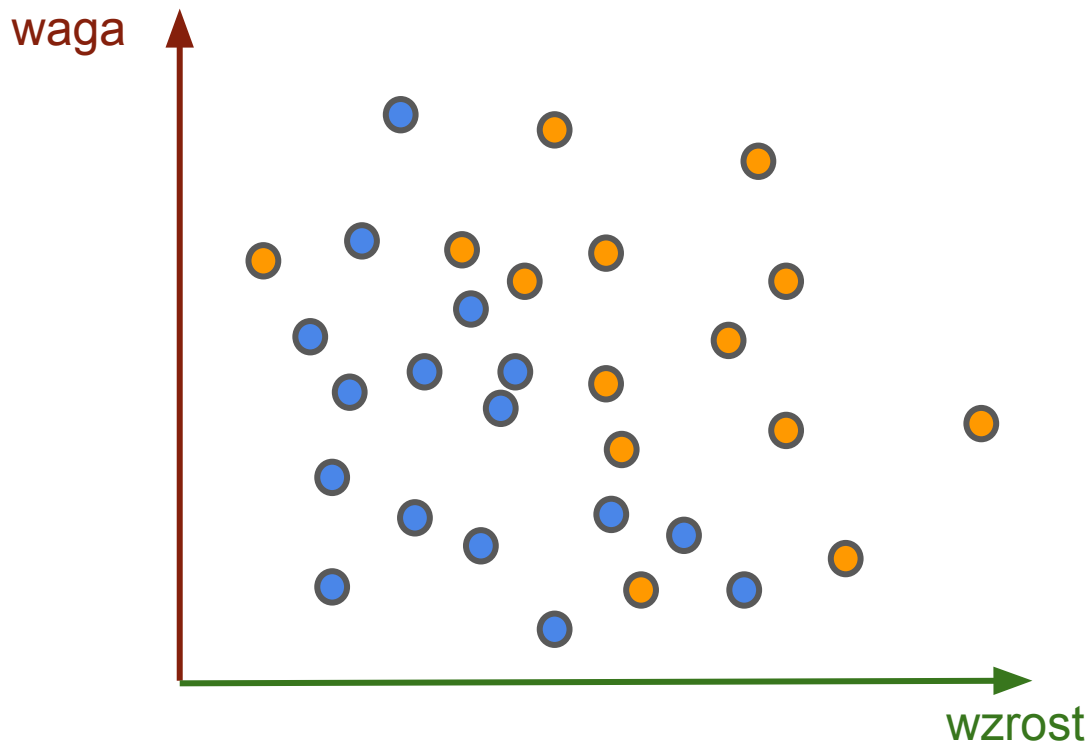
deepsense.ai
BIG DATA SCIENCE

Drzewo decyzyjne

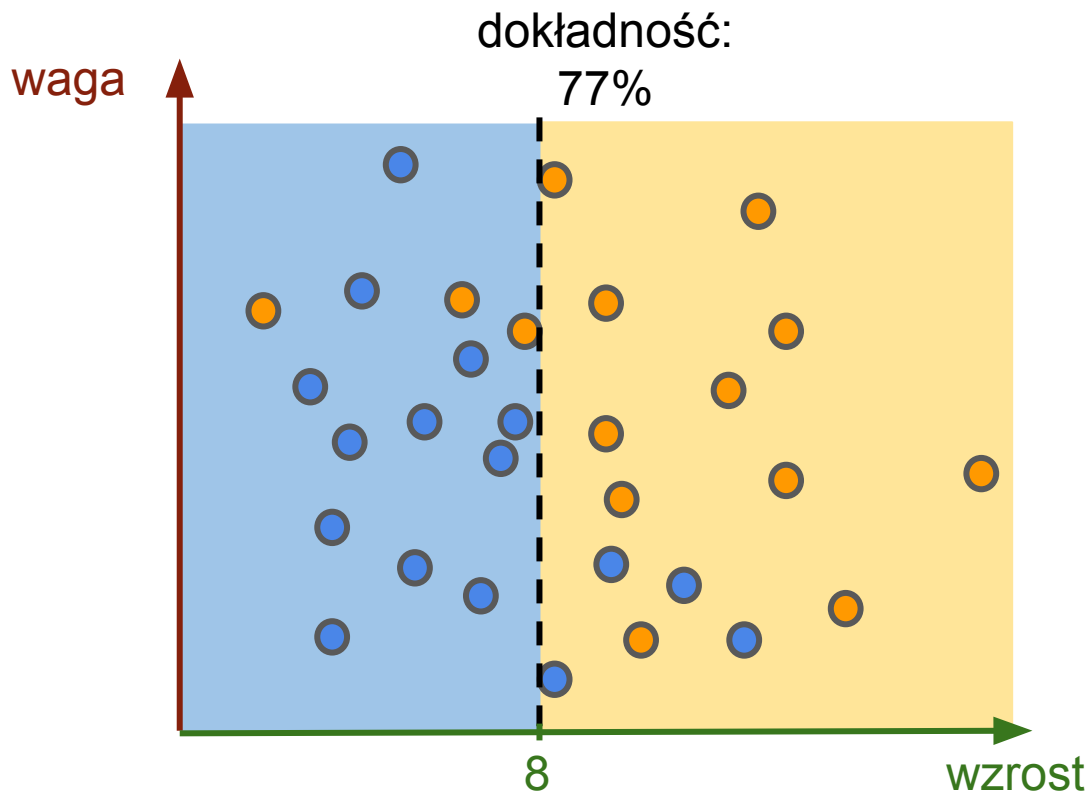
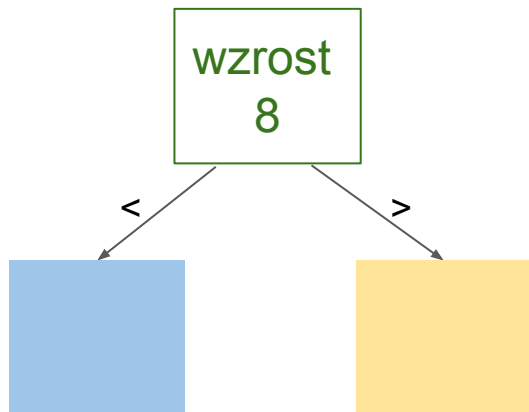
Drzewo decyzyjne



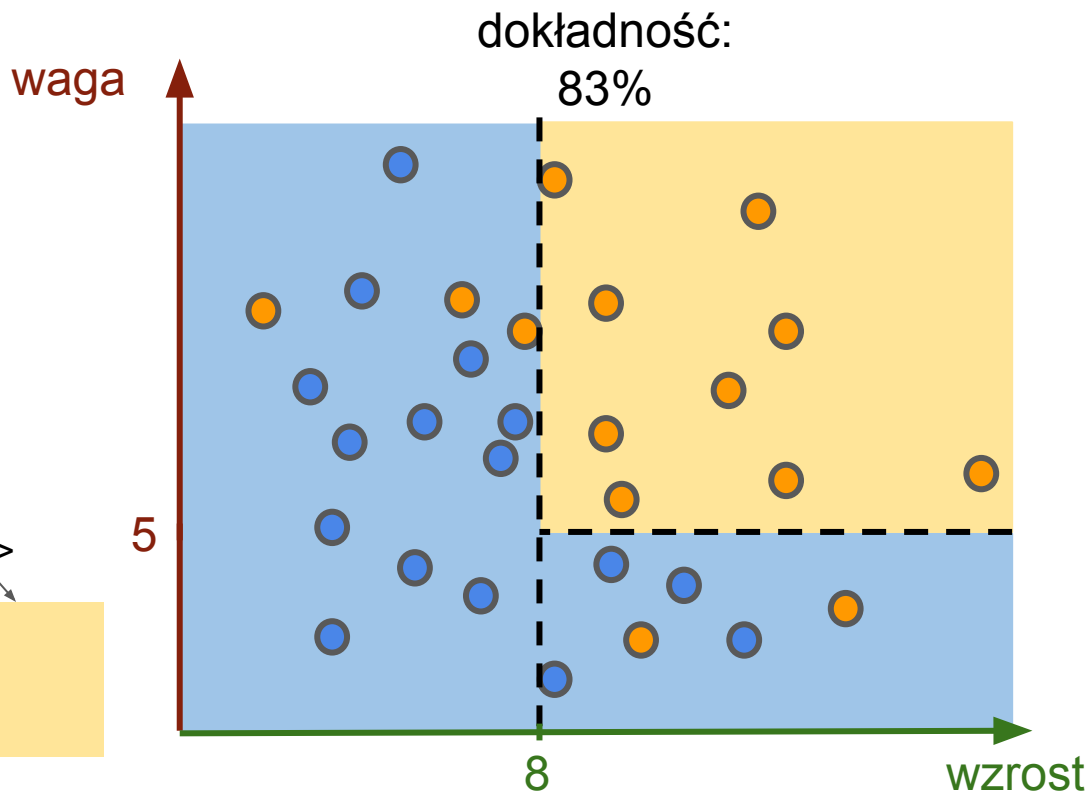
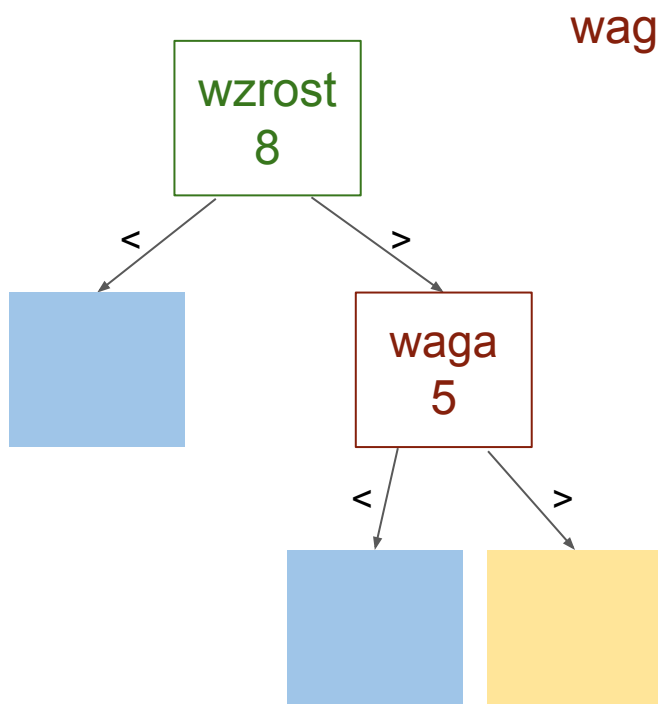
Drzewo decyzyjne



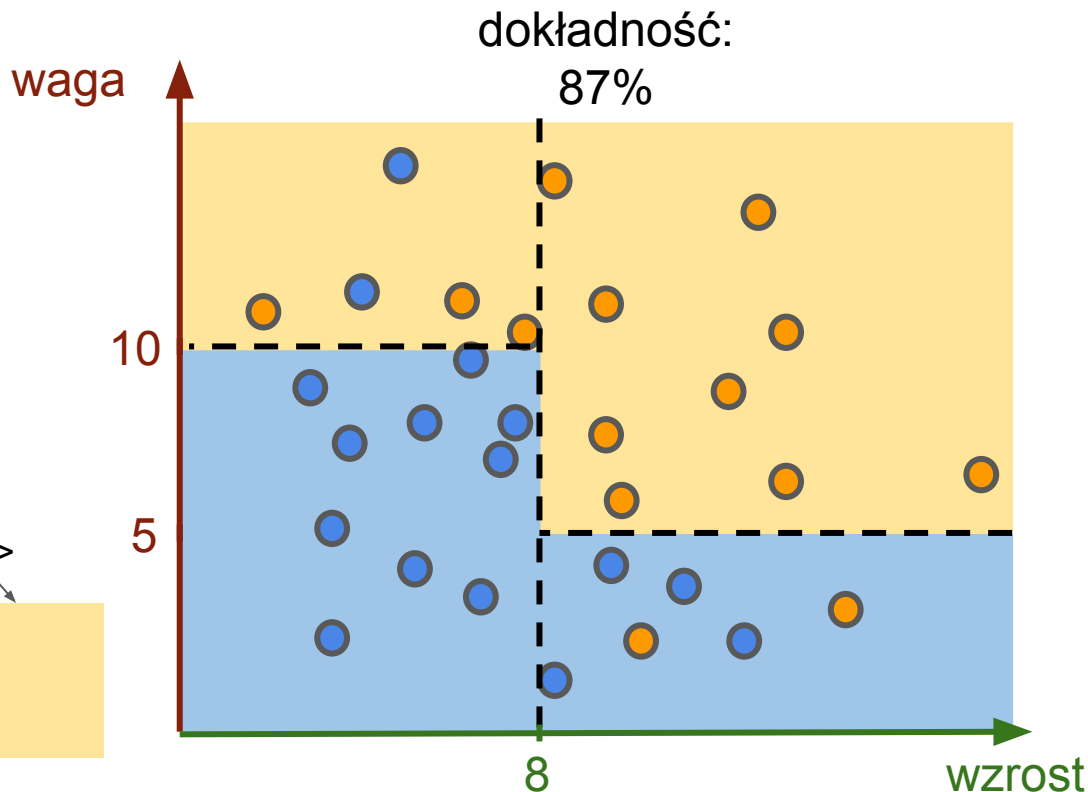
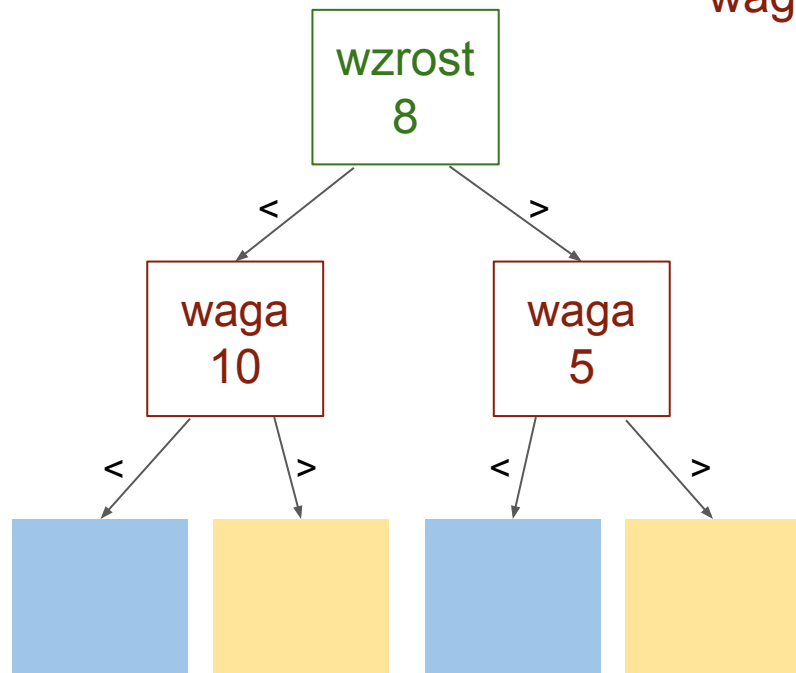
Drzewo decyzyjne



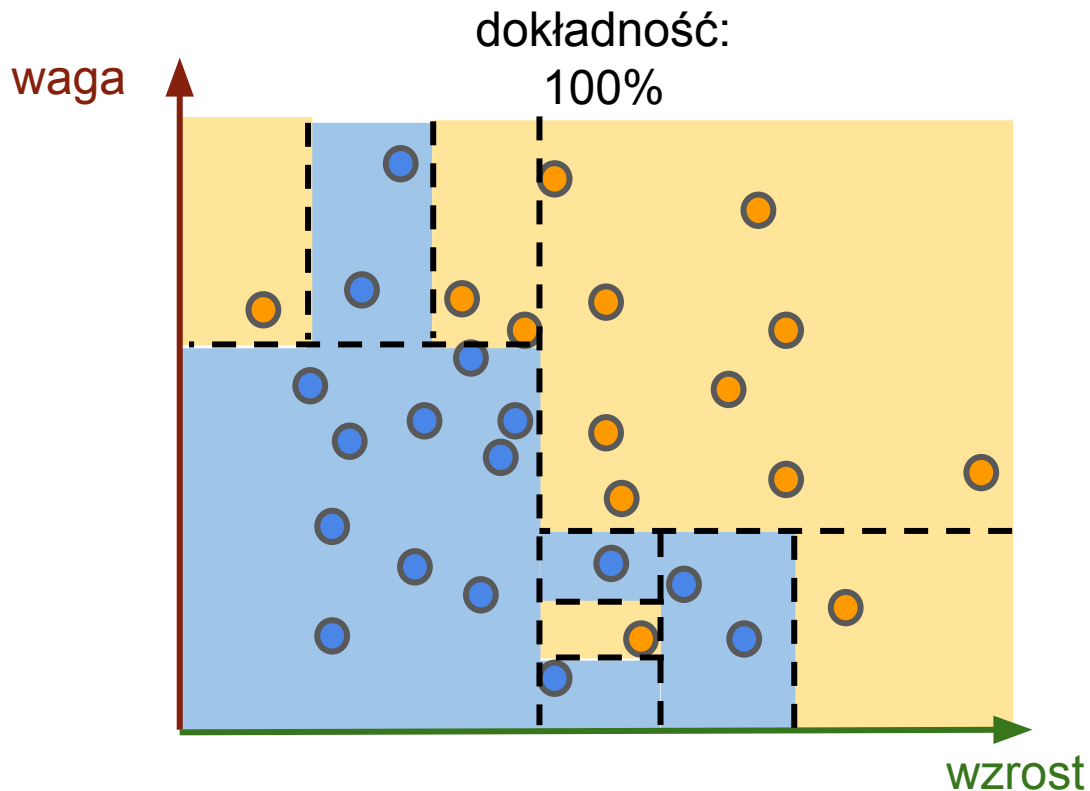
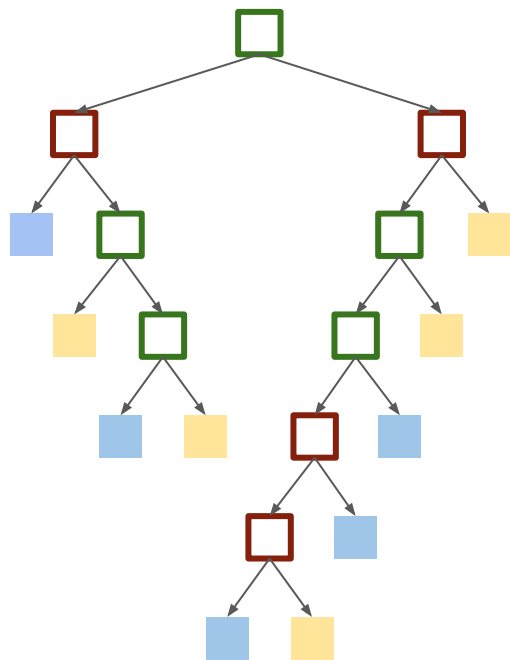
Drzewo decyzyjne



Drzewo decyzyjne



Drzewo decyzyjne



Drzewo decyzyjne

- Nie musimy normalizować danych
- W zasadzie potrzebujemy tylko, żeby dane były w dobrej kolejności, nie interesują nas konkretne wartości
- Stosując je do regresji, nie możemy dokonać predykcji wartości spoza zbioru treningowego

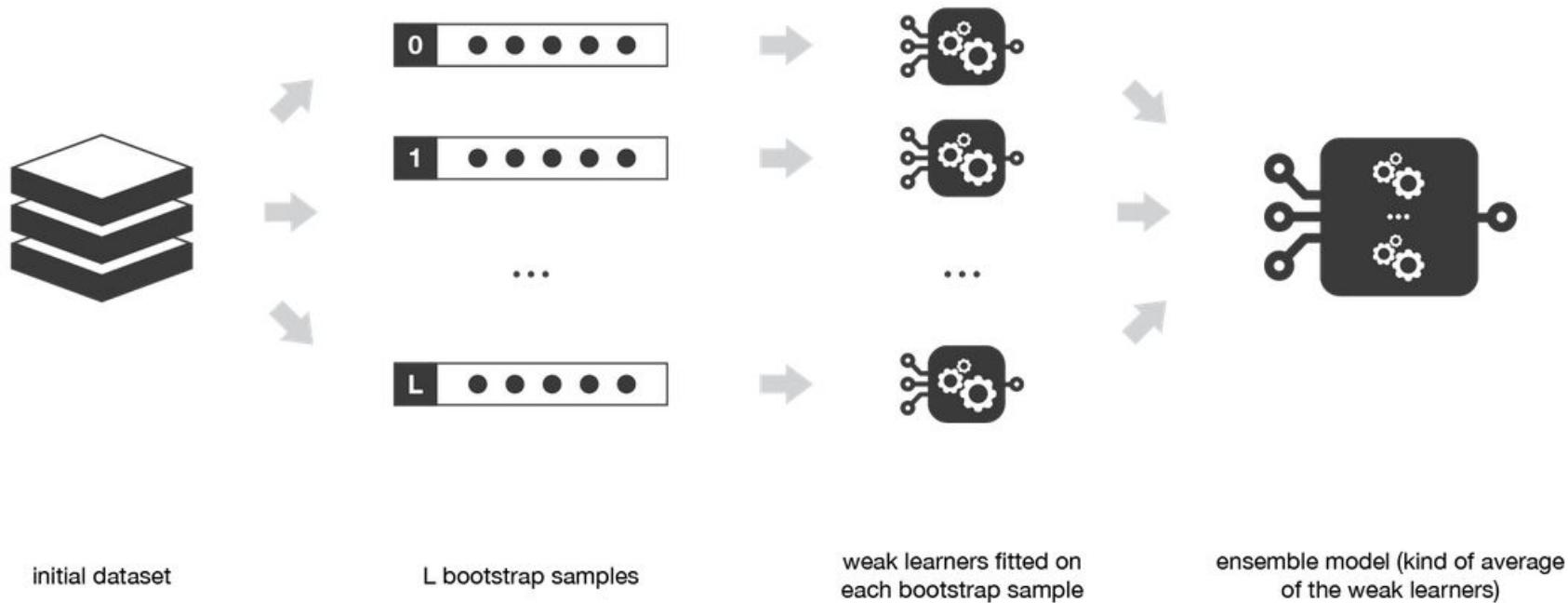




deepsense.ai
BIG DATA SCIENCE

Bagging

Bagging





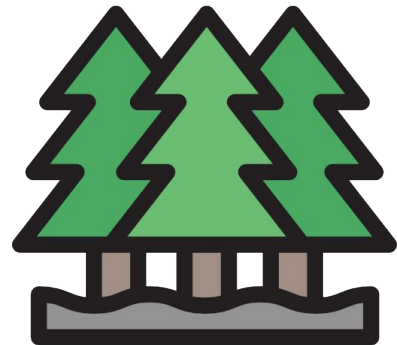
deepsense.ai
BIG DATA SCIENCE

Las losowy

Las losowy

Dużo drzew decyzyjnych

- każde drzewo widzi podzbiór cech
- drzewa są ustalonej głębokości
- każde drzewo głośuje



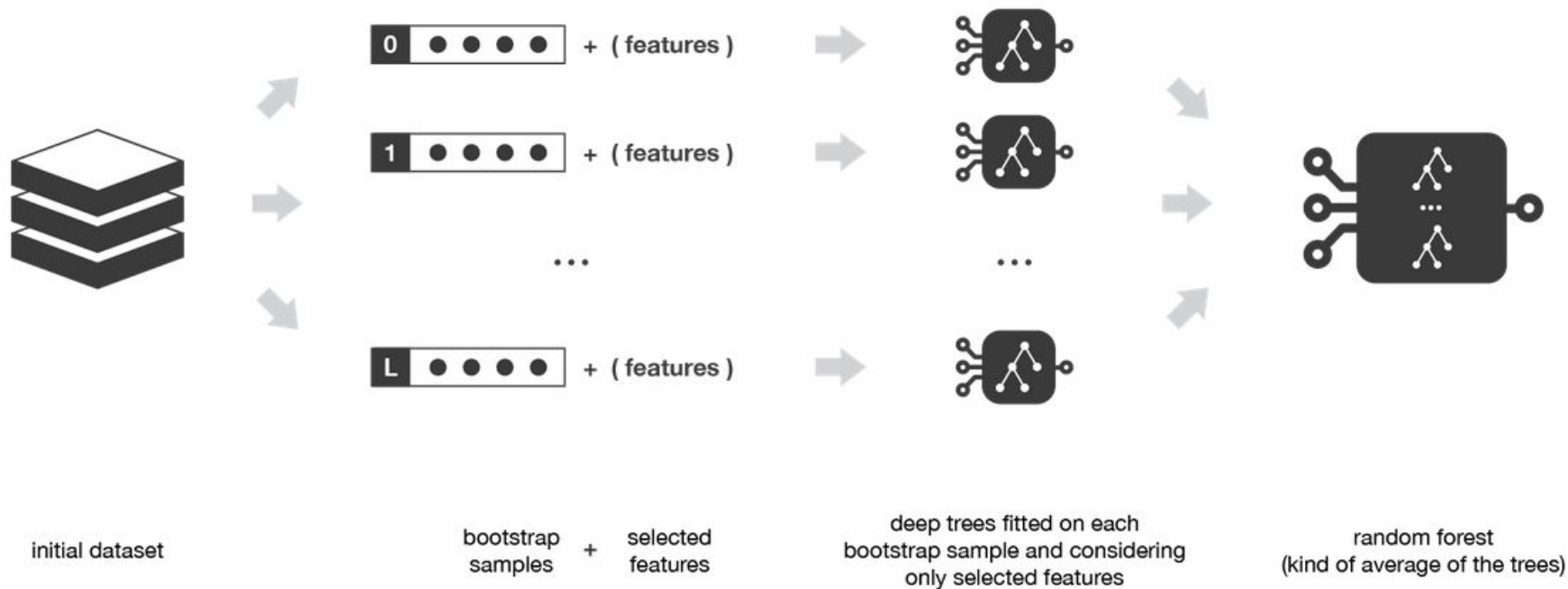
Zalety

- obserwacje odstające nie są problemem
- stosowane do regresji i klasyfikacji
- nie musimy przetwarzać danych

Wady

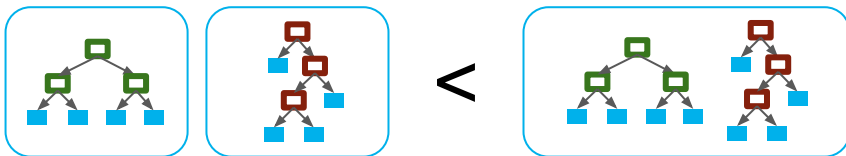
- nie może dokonywać predykcji spoza zbioru treningowego
- nieużyteczne dla szeregów czasowych

Las losowy

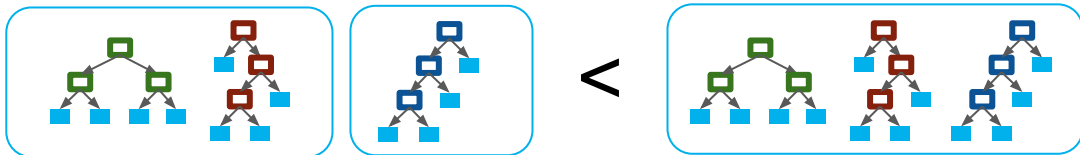


Im więcej drzew tym lepiej?

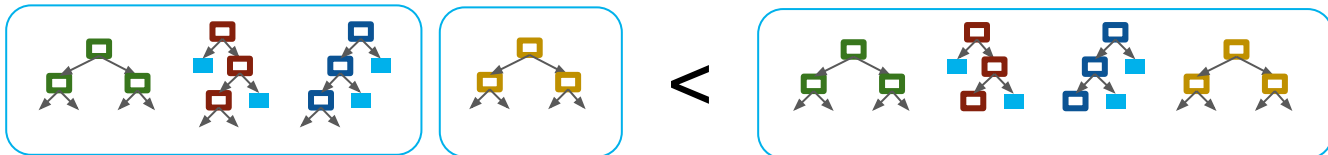
niezależne



niezależne



niezależne



Im więcej drzew tym lepiej?





deepsense.ai
BIG DATA SCIENCE

Boosting

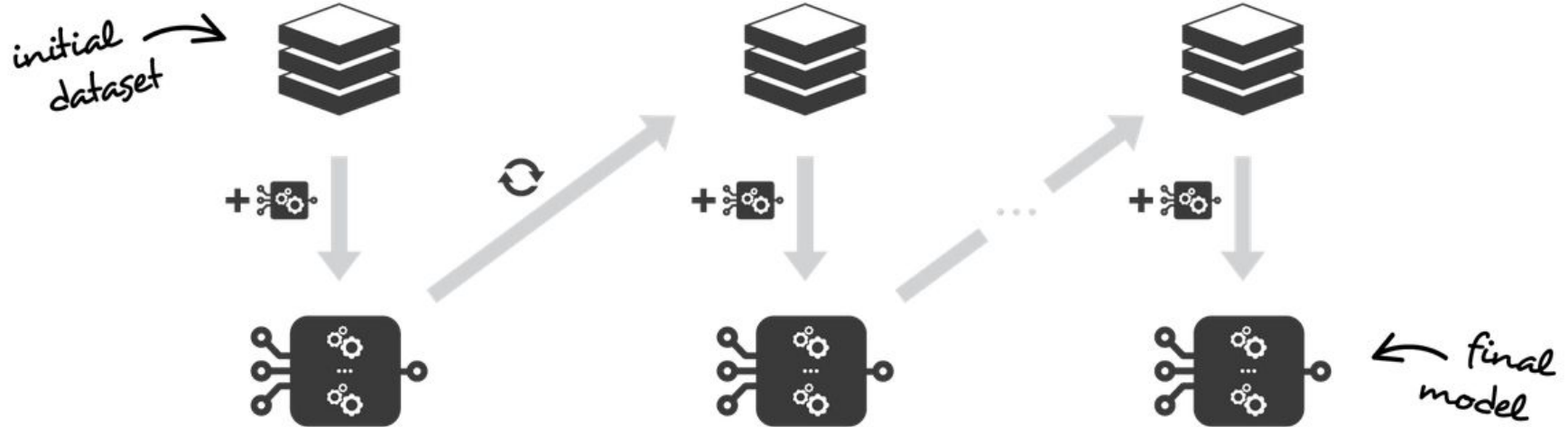
Boosting



train a weak model
and aggregate it to
the ensemble model



update the training dataset
(values or weights) based on the
current ensemble model results

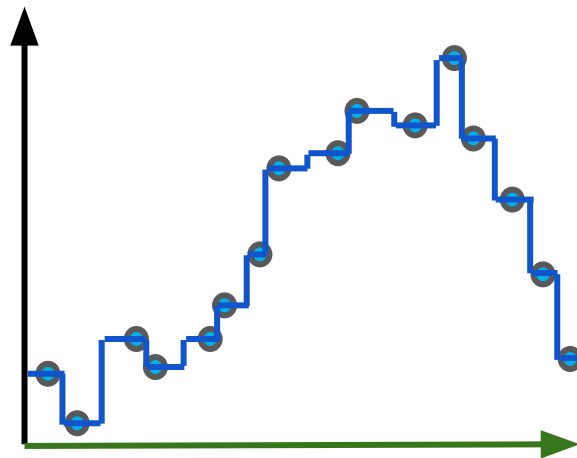


XGBoost

- szybki i dobry - oryginalnie napisany w C++, szybszy niż inne metody oparte o tzw. ensembling
- świetnie się spisuje - mnóstwo wygranych konkursów na Kaggle,
- mnóstwo sposobów tuningu - walidacja krzyżowa, regularyzacja, parametry drzew etc.

Im więcej drzew tym lepiej?

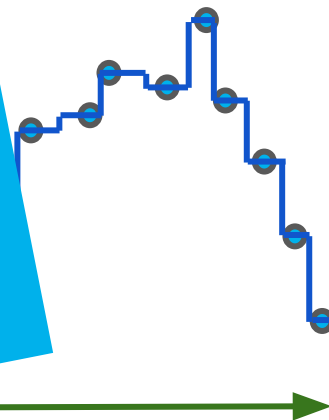
- Więcej drzew - lepsze dopasowanie do danych treningowych.
- Za dużo drzew - przeuczenie!
- Liczba drzew jest określana przez hiperparametr.



Im więcej drzew tym lepiej?

- Więcej drzew
danych tren
- Za dużo drzew
- Liczba drzew
hiperparamet

NIE





Kontakt

Formularz kontaktowy: <https://lnkd.in/dBSg-gzQ>

Program stażowy: <https://deepsense.ai/intern/>

Dziękujemy!



deepsense.ai Sp. z o.o.

Al. Jerozolimskie 162A
02-342 Warsaw
Poland

deepsense.ai, Inc.

2100 Geng Road, Suite 210
Palo Alto, CA 94303
United States of America

Contact us:

Maciej Domagała

maciej.domagala@deepsense.ai

Adam Maciaszek

adam.maciaszek@deepsense.ai

Our offer

contact@deepsense.ai

Media relations & marketing

media@deepsense.ai