



GHOST

Group of Horribly Optimistic Statisticians



ML Next Step

#1 Wstęp i podstawy





GHOST

Group of Horribly Optimistic Statisticians



Agenda

1. Wstęp i podstawy
2. Dane - kluczowy element ML
3. Klasyfikacja
4. Regresja
5. Grupowanie

Deep learning model after....

Epoch 1



Epoch 10



Epoch 99





GHOST

Group of Horribly Optimistic Statisticians

Wstęp i podstawy



GHOST

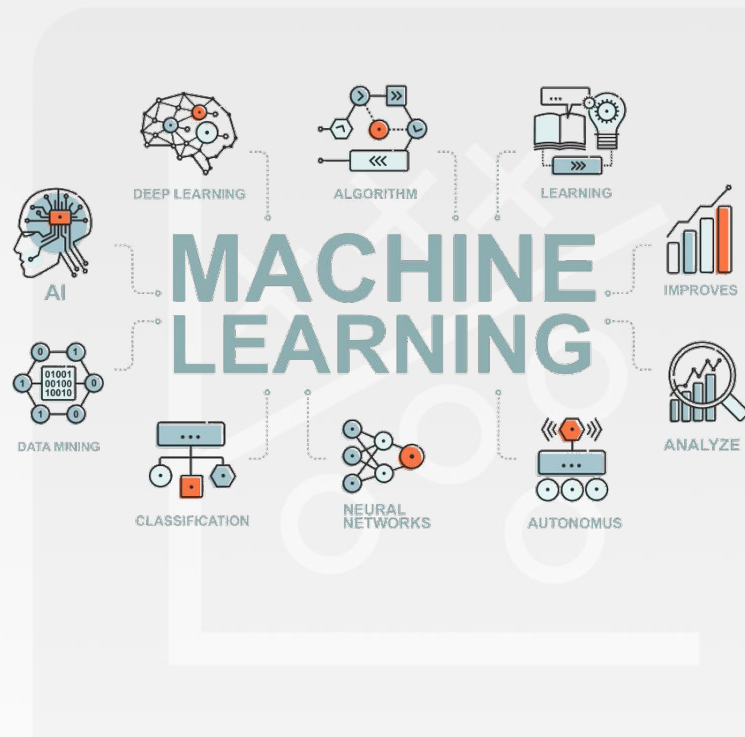
Group of Horribly Optimistic Statisticians



Czym jest Uczenie Maszynowe?

Uczenie maszynowe (ML) to rodzaj sztucznej inteligencji (AI), która pozwala komputerom uczyć się bez sprecyzowanego (explicit) programowania. Polega na wprowadzaniu danych do algorytmów, które mogą następnie identyfikować wzory w datasetach i tworzyć prognozy na podstawie nowych danych.

Główna idea: automatyczne znajdowanie wzorców w danych bez programowania "ręcznego"





GHOST

Group of Horribly Optimistic Statisticians

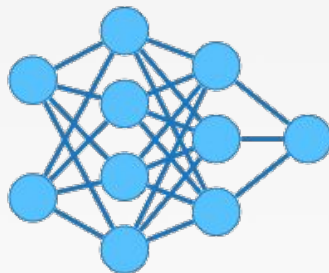


Czym jest Model?

Model ML to funkcja matematyczna, która na podstawie danych wejściowych przewiduje wynik. W procesie trenowania model uczy się odpowiednich wartości parametrów, np. współczynników w regresji.



Prepare Data



Train Model



Deploy Model



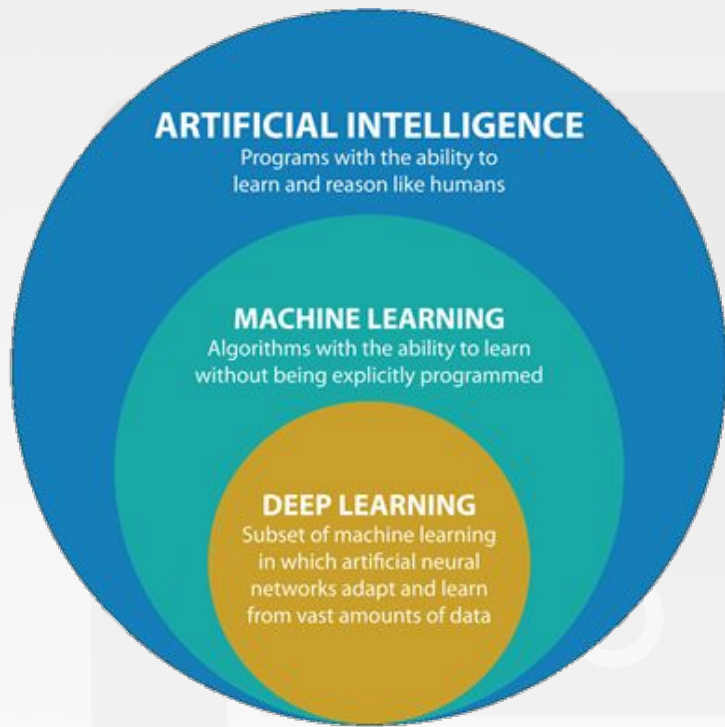
GHOST

Group of Horribly Optimistic Statisticians



Kiedy używamy ML?

- ☀ Istnieje wzór w danych
- ☀ Nie możemy opisać go matematycznie
- ☀ Mamy dane na ten temat





GHOST

Group of Horribly Optimistic Statisticians



Przykłady zastosowań ML

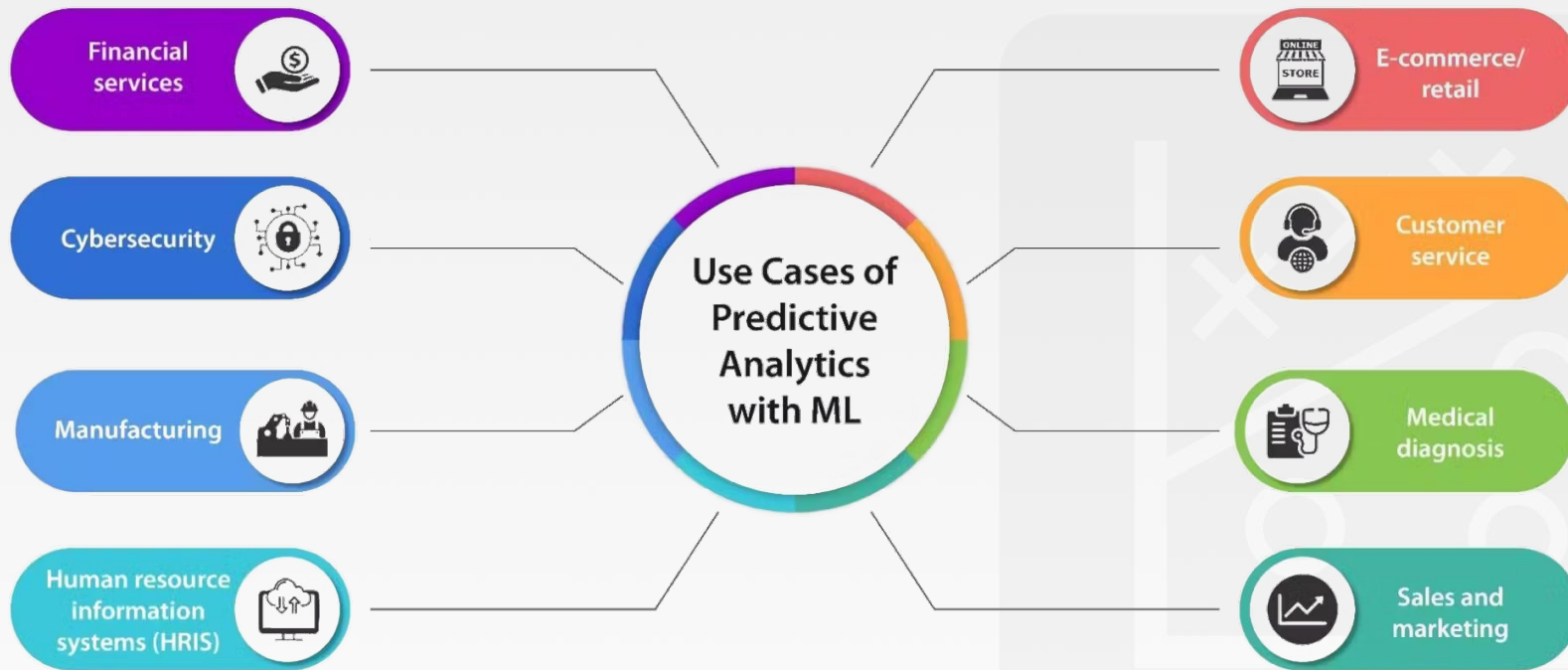
1. Filtry antyspamowe w poczcie elektronicznej
2. Rozpoznawanie obrazów i twarzy
3. Systemy rekomendacyjne
4. Autonomiczne pojazdy
5. Personalizacja reklam
6. Analiza rynku i przewidywanie trendów w finansach
7. Analiza sentymentu w mediach społecznościowych





GHOST

Group of Horribly Optimistic Statisticians





GHOST

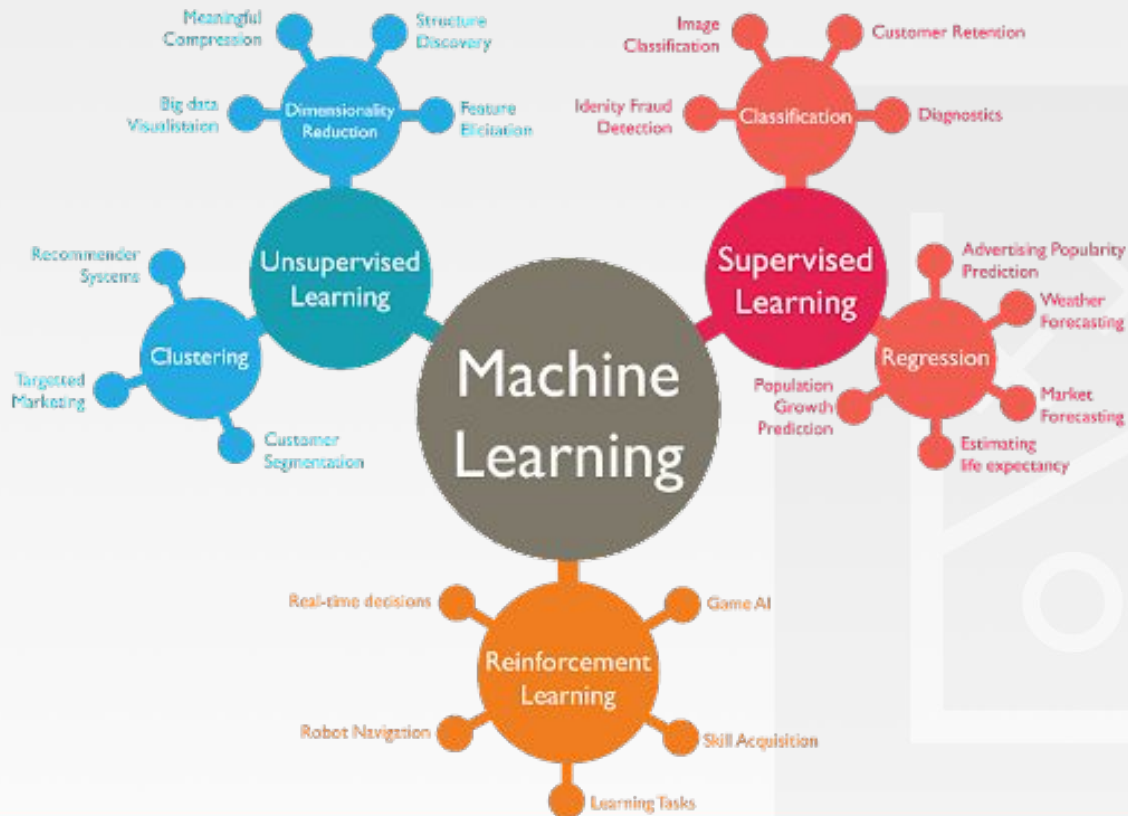
Group of Horribly Optimistic Statisticians

Rodzaje Uczenia Maszynowego



GHOST

Group of Horribly Optimistic Statisticians





GHOST

Group of Horribly Optimistic Statisticians



Rodzaje Uczenia Maszynowego

- ✦ Uczenie nadzorowane (supervised): mamy dane + etykiety
- ✦ Uczenie nienadzorowane (unsupervised): tylko dane, bez etykiet
- ✦ Uczenie półnadzorowane (semisupervised): część danych ma etykiety, część ich nie ma
- ✦ Uczenie przez wzmacnianie (reinforcement): nagrody i kary



GHOST

Group of Horribly Optimistic Statisticians

Proces Uczenia Maszynowego

1. Zbieranie danych
2. Przygotowanie danych
3. Budowanie i trenowanie modelu
4. Testowanie i optymalizacja modelu
5. Użycie modelu w rzeczywistych zastosowaniach



GHOST

Group of Horribly Optimistic Statisticians

Wyzwania w Uczeniu Maszynowym

1. Overfitting
2. Underfitting
3. Niska jakość lub brakujące dane
4. Problemy ze zróżnicowaniem danych

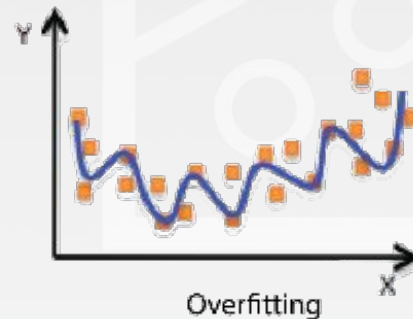
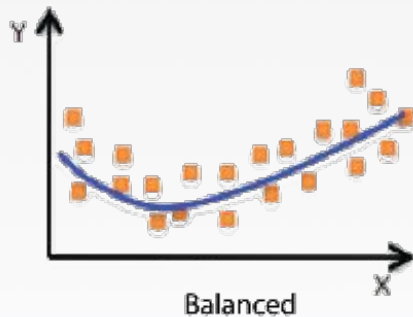
**GHOST**

Group of Horribly Optimistic Statisticians



overfitting

Overfitting pojawia się, gdy model uczy się bardzo dobrze danych treningowych, ale traci zdolność generalizacji do nowych danych. Oznacza to, że model jest "zbyt dopasowany" do szczegółów zbioru treningowego, co prowadzi do słabych wyników na danych testowych. Problem ten często rozwiązuje się poprzez techniki regularizacji lub zwiększenie ilości danych treningowych.



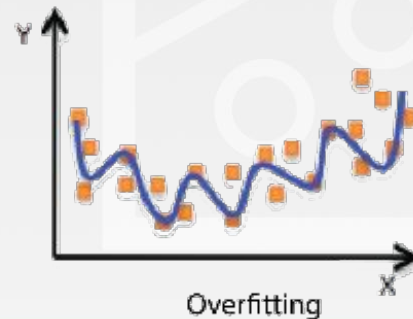
**GHOST**

Group of Horribly Optimistic Statisticians



underfitting

Underfitting pojawia się, gdy model nie jest w stanie dobrze dopasować się do danych treningowych, co skutkuje słabymi wynikami zarówno na danych treningowych, jak i testowych. Zwykle jest to efekt zbyt prostego modelu lub niewystarczającej liczby epok w procesie uczenia. Aby zmniejszyć ryzyko underfittingu, można zwiększyć złożoność modelu, dostarczyć więcej cech lub zmienić parametry modelu, tak aby lepiej uchwycił strukturę danych.





GHOST

Group of Horribly Optimistic Statisticians

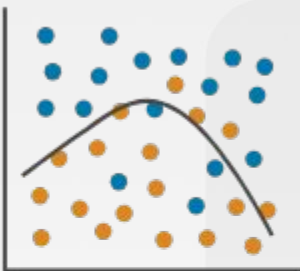


Classification

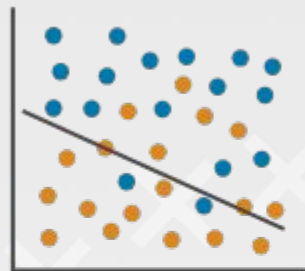
Overfitting



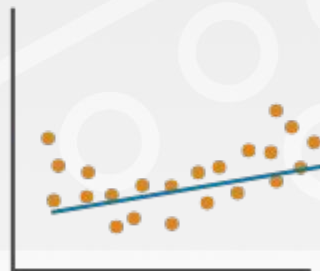
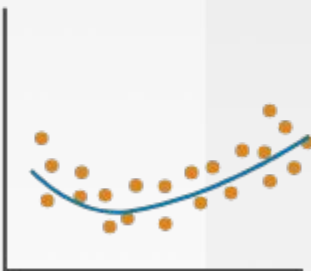
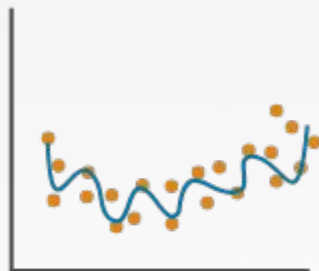
Right Fit



Underfitting



Regression



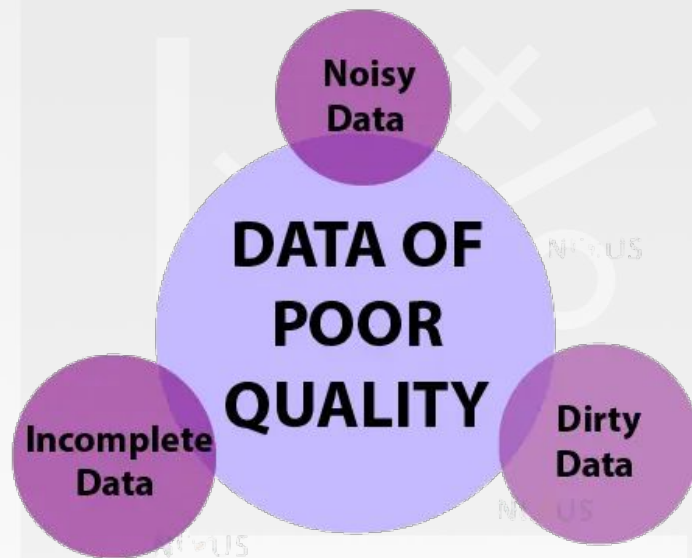
**GHOST**

Group of Horribly Optimistic Statisticians



Niska jakość lub brakujące dane

Dane o niskiej jakości lub brakujące wartości utrudniają budowę wiarygodnych modeli, gdyż niepełne lub nieprecyzyjne informacje mogą prowadzić do błędnych prognoz. Często konieczne są metody wypełniania brakujących wartości lub techniki czyszczenia danych. Praca z takimi danymi wymaga także dobrego zrozumienia źródła danych, aby zminimalizować potencjalne błędy



**GHOST**

Group of Horribly Optimistic Statisticians



Problem ze zróżnicowaniem danych

Niezerównoważone lub jednorodne dane mogą prowadzić do modeli, które nie są w stanie dobrze uogólnić. Jeśli np. pewna klasa jest nadreprezentowana, model może faworyzować tę klasę kosztem innych. W takim przypadku stosuje się metody balansowania klas lub wzbogacania zbioru danych, aby poprawić zdolność modelu do rozpoznawania rzadziej występujących wzorców.





GHOST

Group of Horribly Optimistic Statisticians

DANE

kluczowy element w uczeniu maszynowym



GHOST

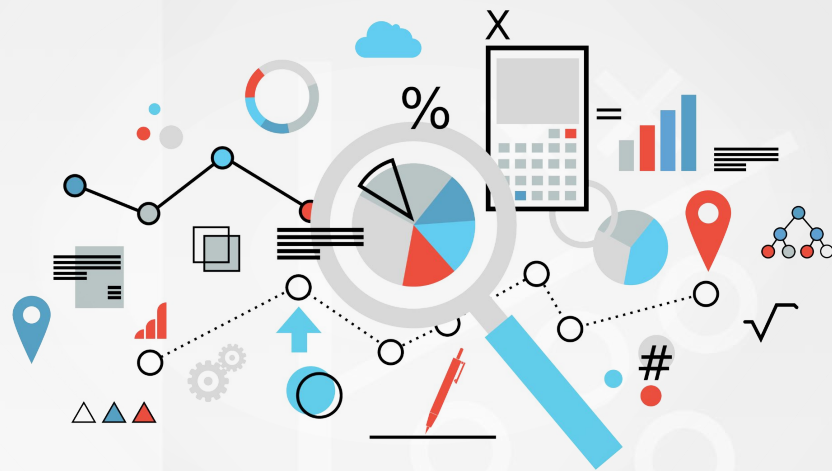
Group of Horribly Optimistic Statisticians



Czym są dane?

Zbiory wartości, które przekazują informacje, opisując ilość, jakość, fakt, statystyki, inne znaczenia lub sekwencje symboli, które mogą być dalej interpretowane i przetwarzane

eurostat 

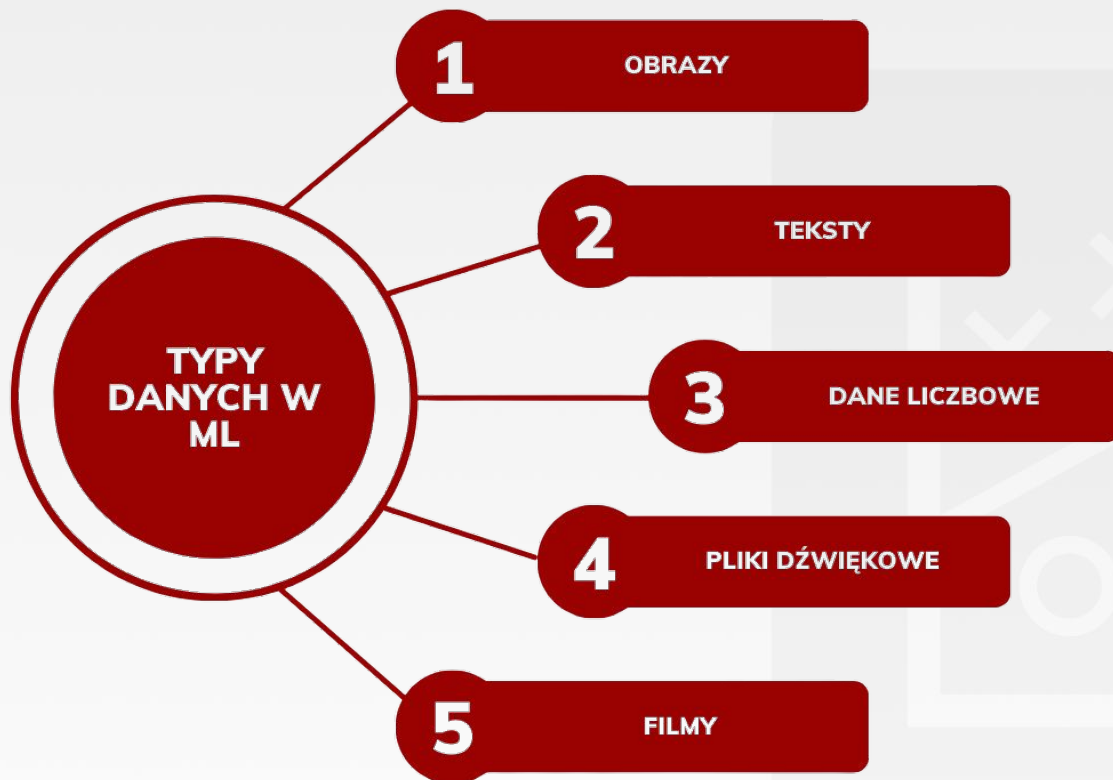




GHOST

Group of Horribly Optimistic Statisticians

Typy danych w uczeniu maszynowym

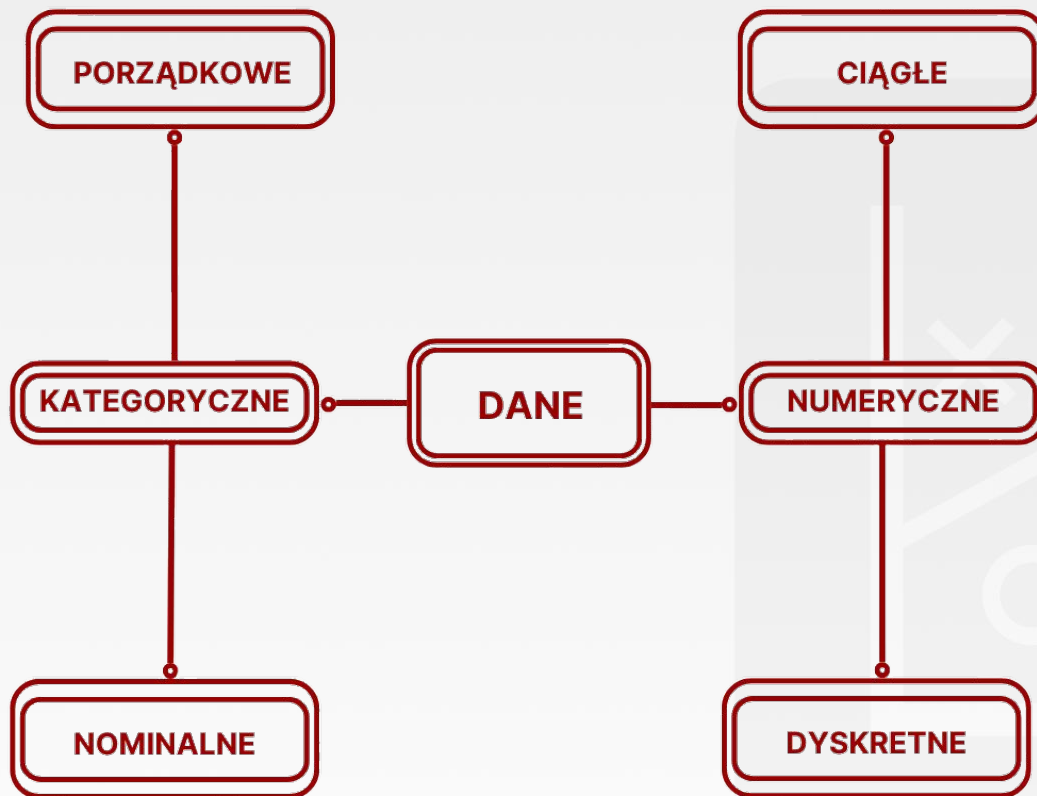




GHOST

Group of Horribly Optimistic Statisticians

Typy danych w uczeniu maszynowym





GHOST

Group of Horribly Optimistic Statisticians



OCZYSZCZANIE DANYCH

USUWANIE LUB
UZUPEŁNIANIE
BRAKUJĄCYCH
WARTOŚCI

POPRAWIANIE
LITERÓWEK,
NIEZGODNYCH NAZW
KATEGORII

USUWANIE
DUPLIKATÓW

USUWANIE DANYCH
ODSTAJĄCYCH



GHOST

Group of Horribly Optimistic Statisticians

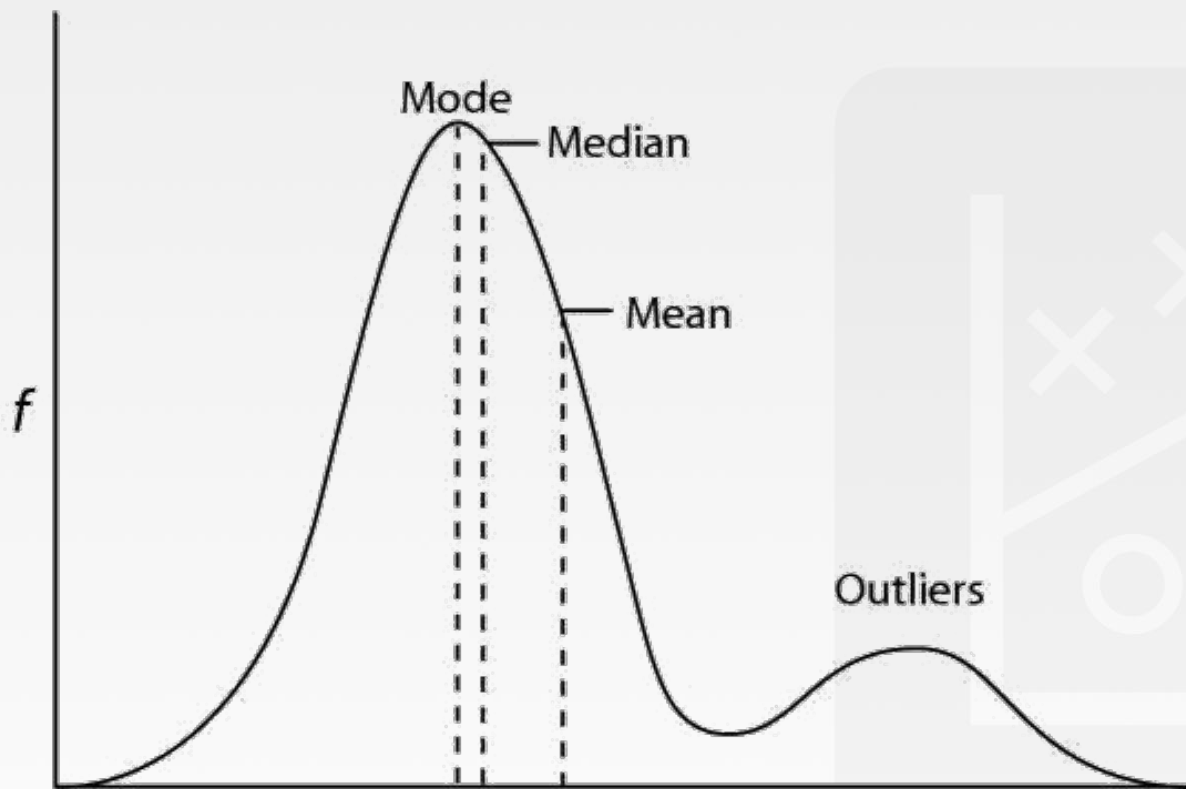




GHOST

Group of Horribly Optimistic Statisticians

Wartości odstające (outliers)





GHOST

Group of Horribly Optimistic Statisticians

Klasyfikacja



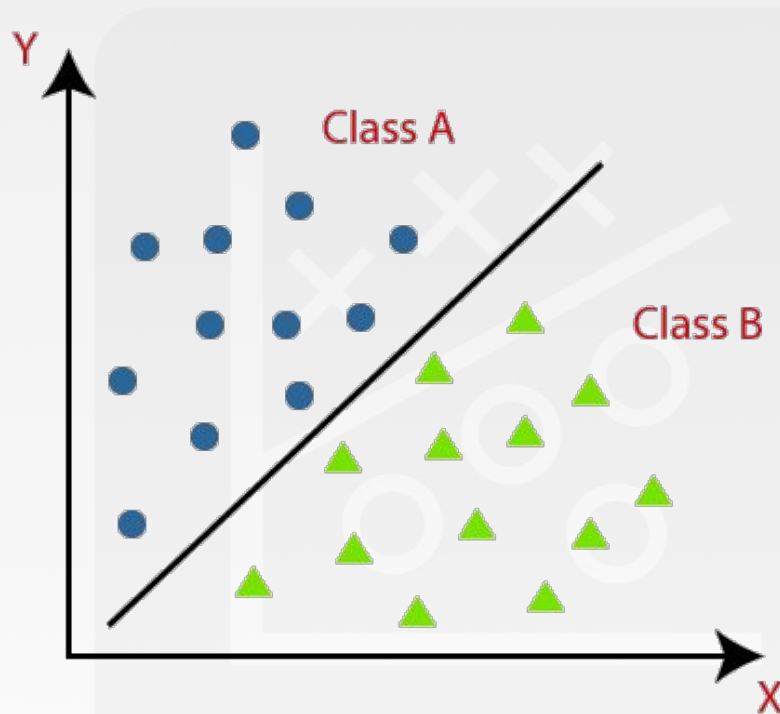
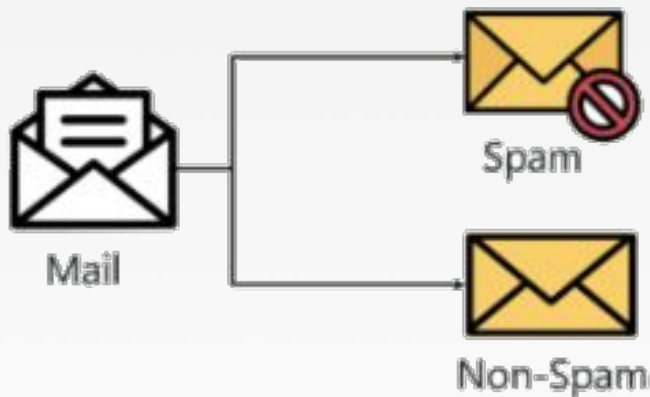
GHOST

Group of Horribly Optimistic Statisticians



Czym jest klasyfikacja?

Klasyfikacja to przypisywanie obiektów do określonych kategorii na podstawie ich cech.





GHOST

Group of Horribly Optimistic Statisticians



Przykłady zastosowań klasyfikacji

1. Bezpieczeństwo i filtrowanie

a. Filtry antyspamowe

b. Wykrywanie oszustw bankowych

2. Medycyna

a. Rozpoznawanie chorób





GHOST

Group of Horribly Optimistic Statisticians



Przykłady zastosowań klasyfikacji

3. Marketing i e-commerce

a. Segmentacja klientów

4. Technologie mobilne i autonomiczne

a. Rozpoznawanie obiektów na drodze

b. Autoryzacja użytkownika głosem lub twarzą





GHOST

Group of Horribly Optimistic Statisticians



Rodzaje klasyfikacji

Wyróżniamy klasyfikację binarną, wieloklasową i multilabel.

W klasyfikacji binarnej mamy tylko dwie klasy, np. „tak” lub „nie”. Klasyfikacja wieloklasowa dotyczy więcej niż dwóch klas – np. rozpoznawanie gatunku kwiatu (setosa, versicolor, virginica). Multilabel to przypadek, w którym jedna próbka może należeć do wielu klas jednocześnie – np. zdjęcie zawierające kota i psa. Rodzaj klasyfikacji wpływa na wybór modelu i metryk oceny.

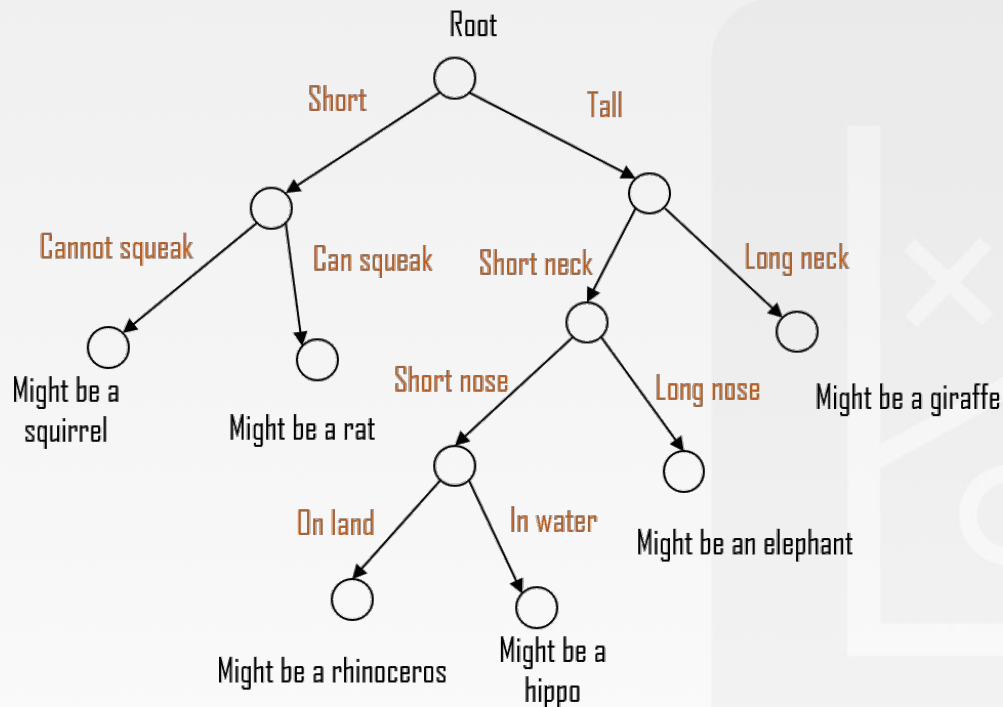


GHOST

Group of Horribly Optimistic Statisticians



Klasyczne algorytmy klasyfikacji



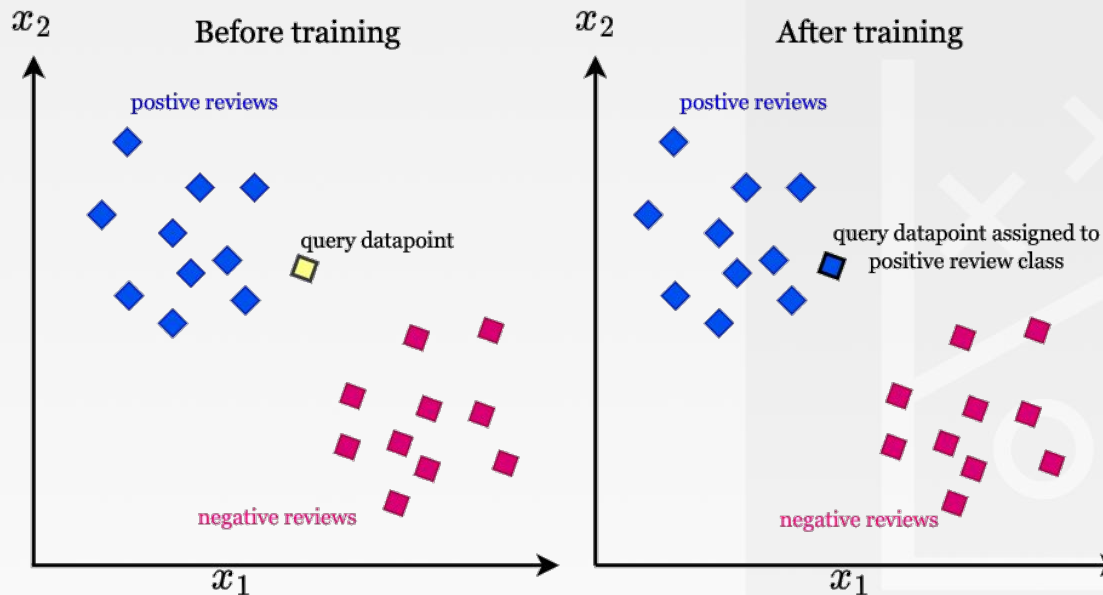


GHOST

Group of Horribly Optimistic Statisticians



Klasyczne algorytmy klasyfikacji



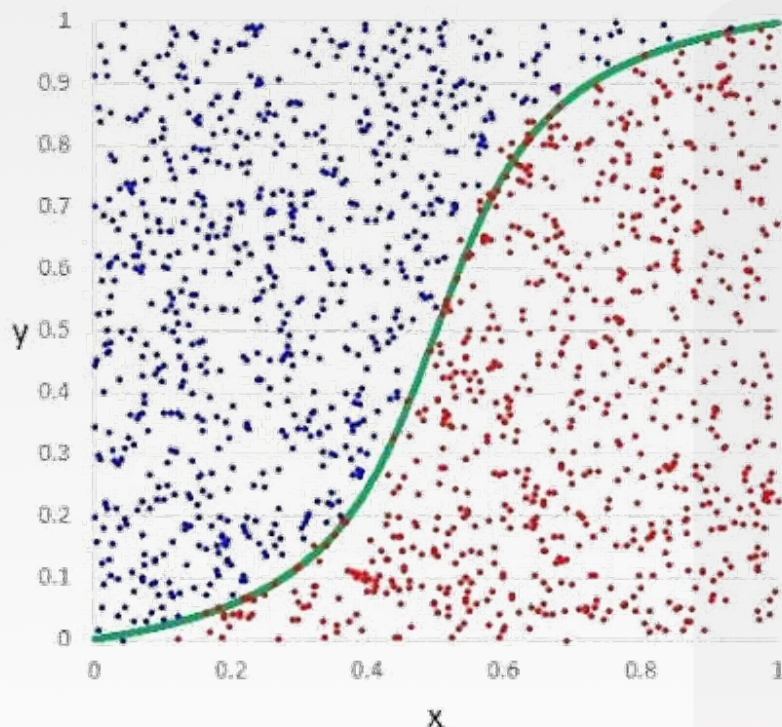


GHOST

Group of Horribly Optimistic Statisticians



Klasyczne algorytmy klasyfikacji



- Boundary
- False samples
- True samples





GHOST

Group of Horribly Optimistic Statisticians

Regresja



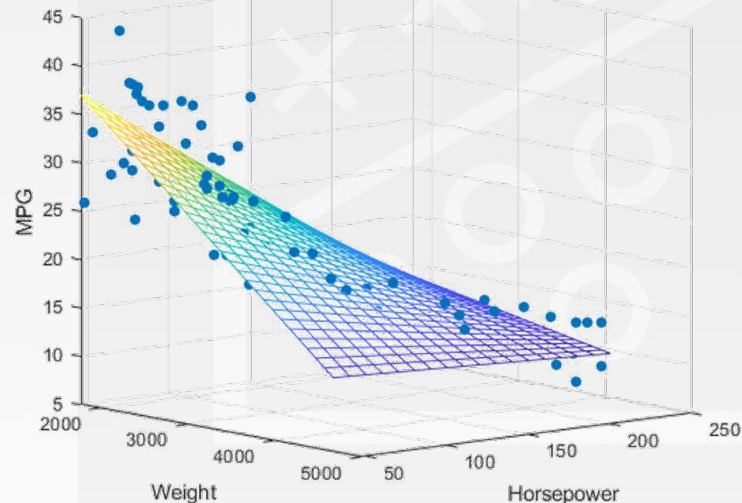
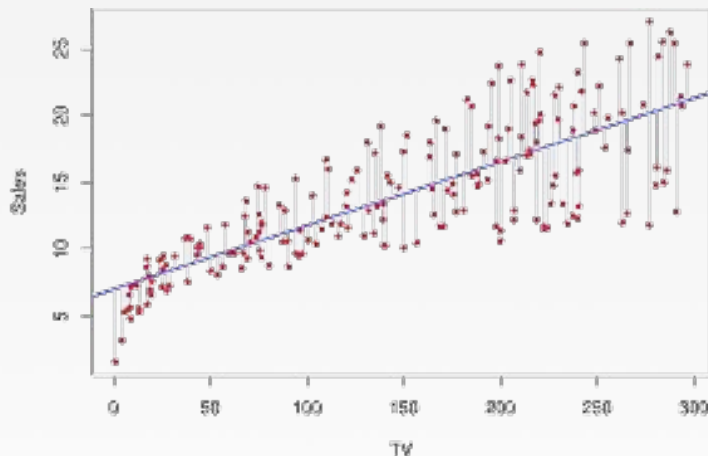
GHOST

Group of Horribly Optimistic Statisticians



Czym jest regresja?

Regresja to przewidywanie wartości liczbowej na podstawie danych wejściowych.





GHOST

Group of Horribly Optimistic Statisticians



Przykłady zastosowań regresji

1. Prognozowanie
2. Analiza trendów
3. Modelowanie zjawisk





GHOST

Group of Horribly Optimistic Statisticians



Rodzaje regresji

1. **Regresja Liniowa:** Modelowanie liniowej zależności między zmiennymi.
2. **Regresja Wielomianowa:** Modelowanie nieliniowych zależności poprzez wielomiany.
3. **Regresja Logistyczna:** Modelowanie prawdopodobieństwa wystąpienia zdarzenia binarnego.
4. **Regresja Ridge i Lasso:** Techniki regularizacji zapobiegające przeuczeniu modelu.

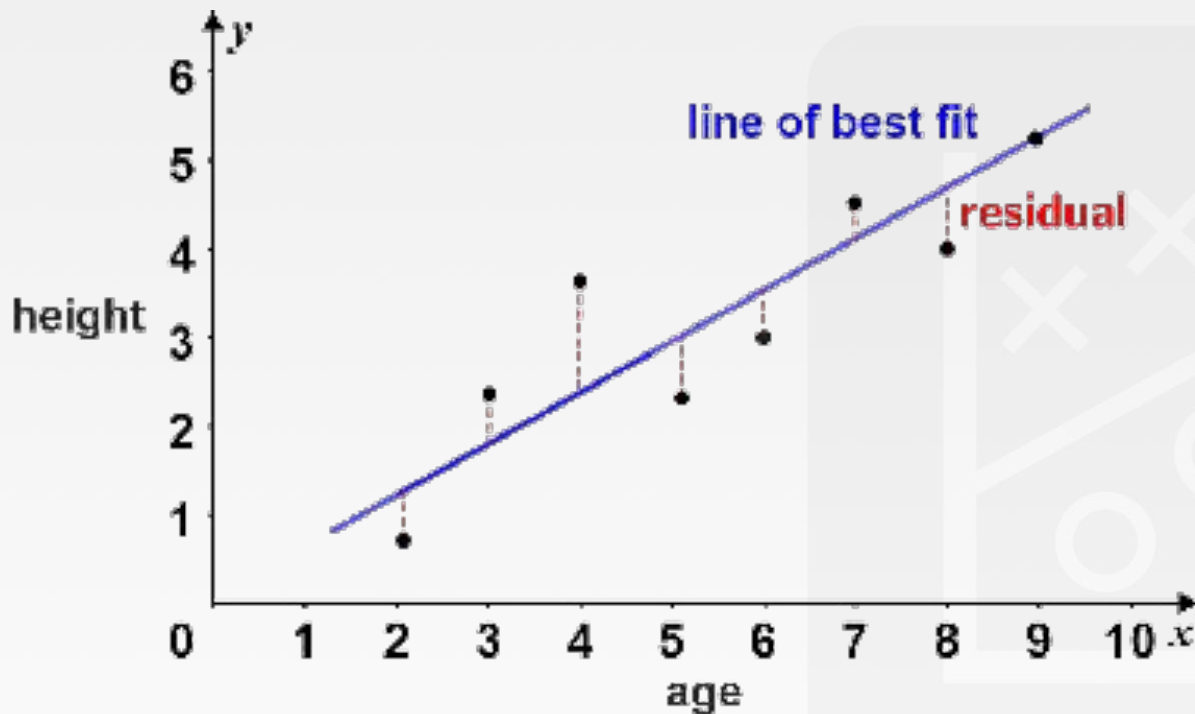


GHOST

Group of Horribly Optimistic Statisticians



Błąd losowy - Reszty





GHOST

Group of Horribly Optimistic Statisticians



Prosta Regresja Liniowa



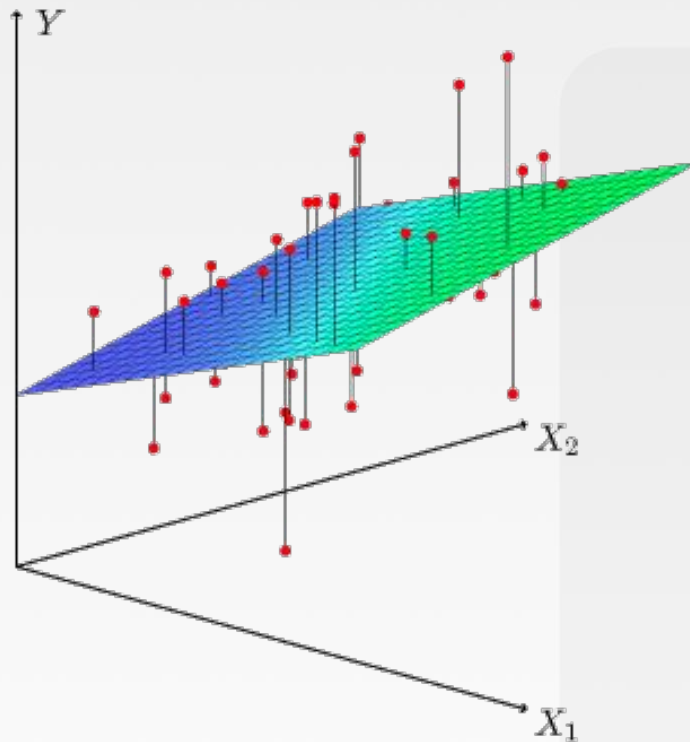


GHOST

Group of Horribly Optimistic Statisticians



Multiple Linear Regression

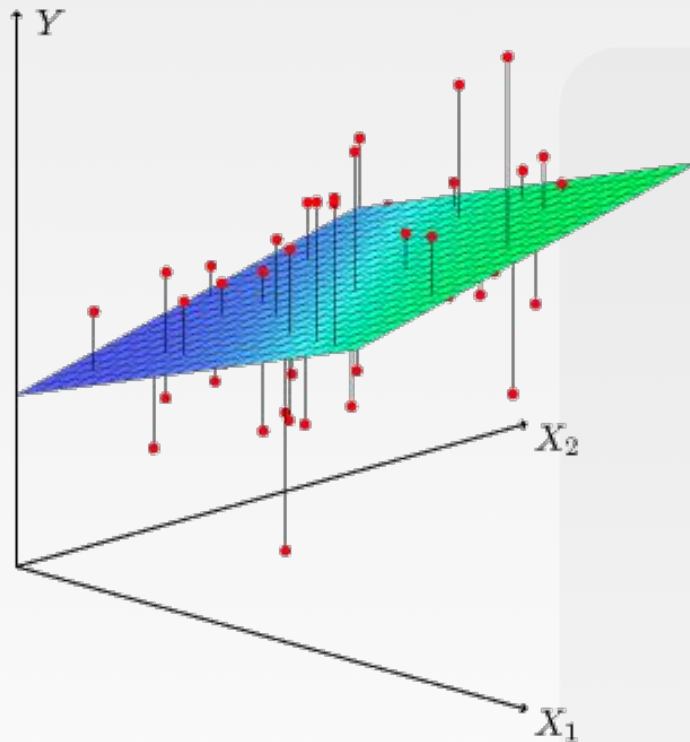




GHOST

Group of Horribly Optimistic Statisticians

Wielozmiennowa Regresja Liniowa

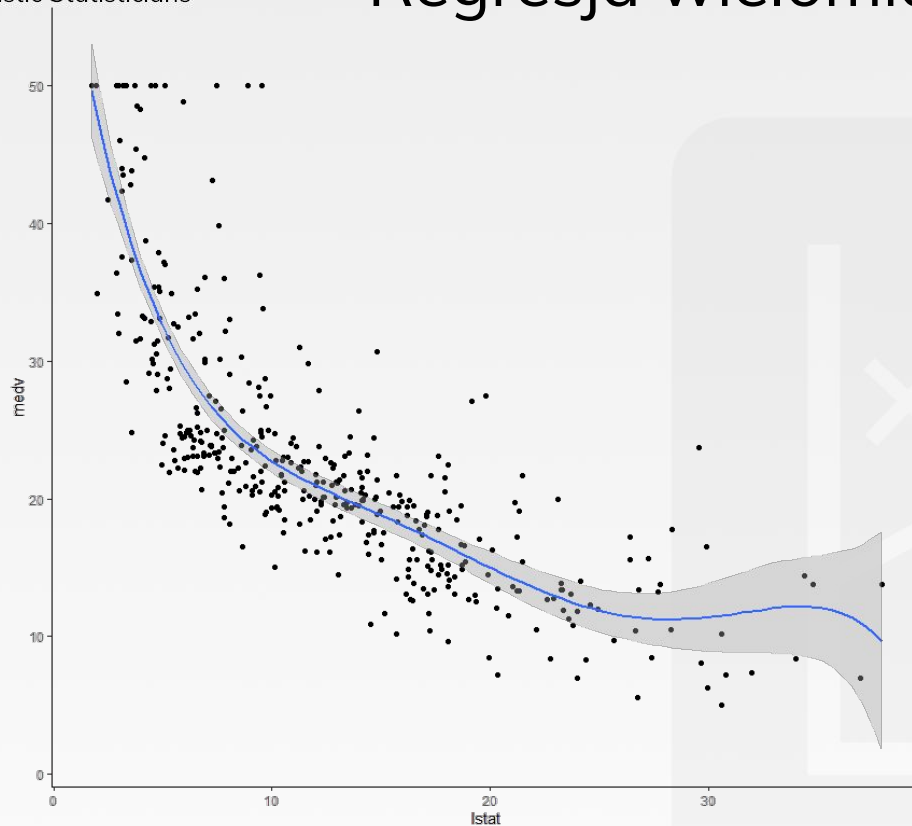




GHOST

Group of Horribly Optimistic Statisticians

Regresja wielomianowa

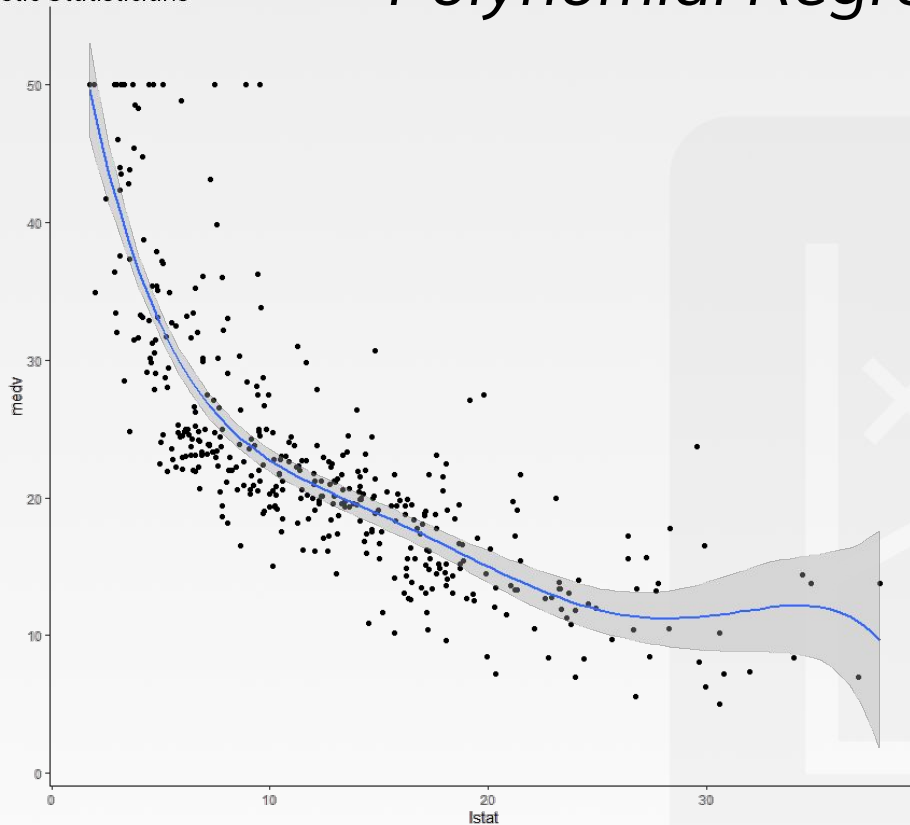




GHOST

Group of Horribly Optimistic Statisticians

Polynomial Regression





GHOST

Group of Horribly Optimistic Statisticians



Simple
Linear
Regression

$$y = b_0 + b_1 x_1$$

Multiple
Linear
Regression

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

Polynomial
Linear
Regression

$$y = b_0 + b_1 x_1 + b_2 x_1^2 + \dots + b_n x_1^n$$



GHOST

Group of Horribly Optimistic Statisticians

Klasteryzacja



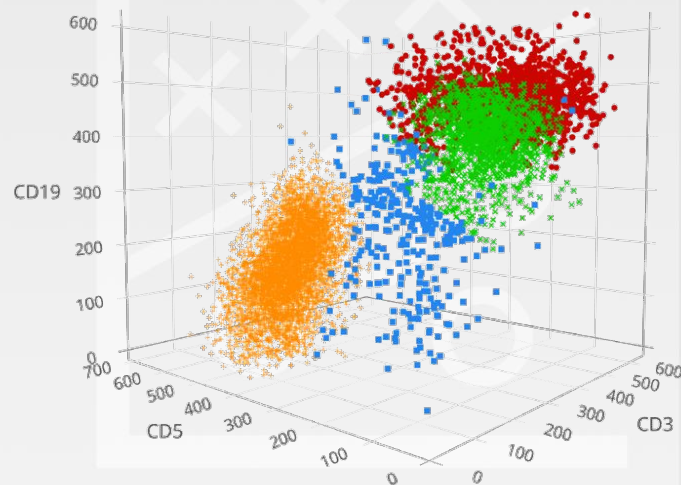
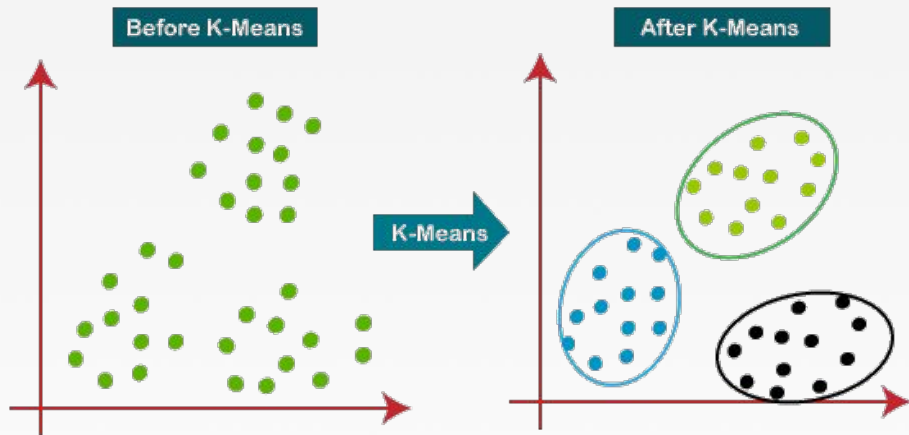
GHOST

Group of Horribly Optimistic Statisticians



Czym jest klasteryzacja?

Klasteryzacja polega na grupowaniu danych na podstawie ich podobieństw - bez znajomości etykiet.





GHOST

Group of Horribly Optimistic Statisticians



Przykłady zastosowań klasteryzacji

1. Segmentacja - segmentacja klientów
2. Analiza obrazów - rozpoznawanie wzorców
3. Biomedycyna - grupowanie genów
4. Analiza tekstu - grupowanie dokumentów





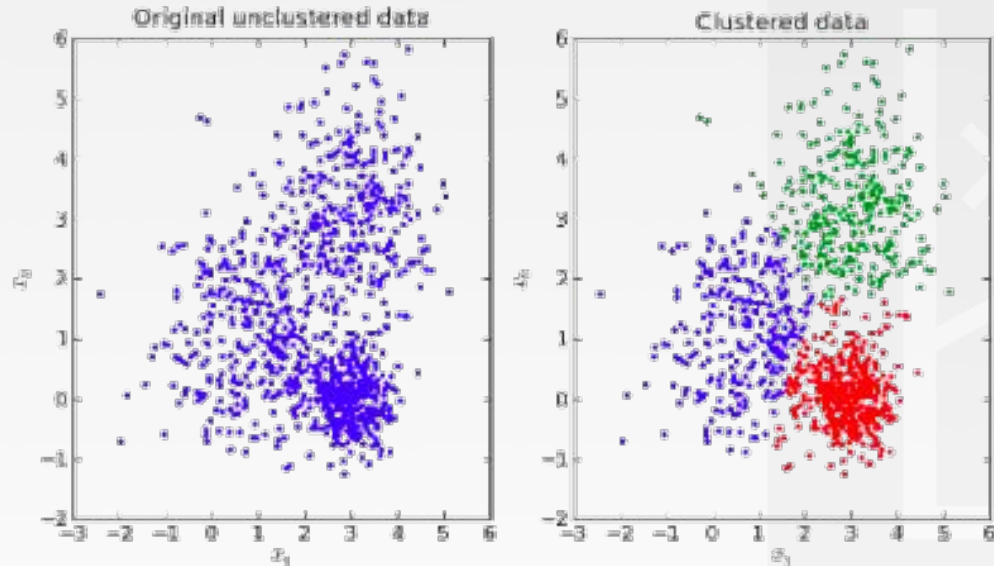
GHOST

Group of Horribly Optimistic Statisticians



Przykładowe algorytmy klasteryzacji

K-means





GHOST

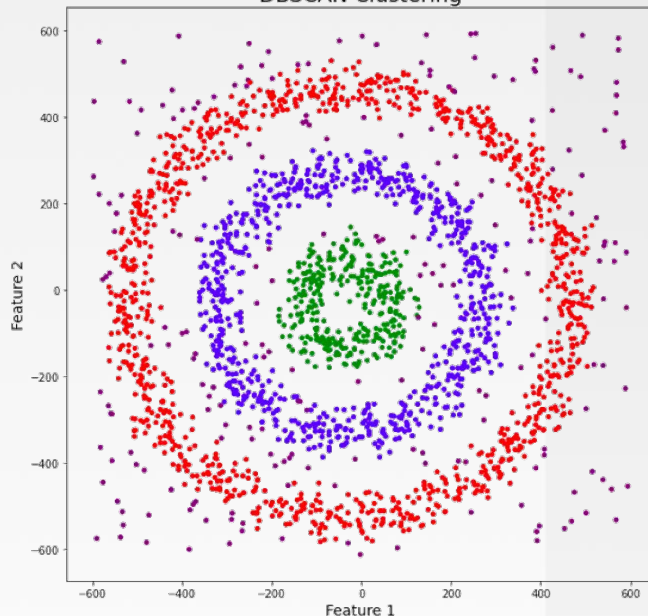
Group of Horribly Optimistic Statisticians



Przykładowe algorytmy klasteryzacji

DBSCAN

DBSCAN Clustering



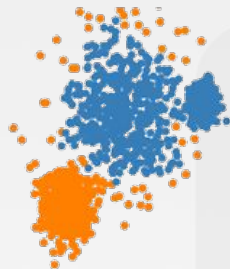
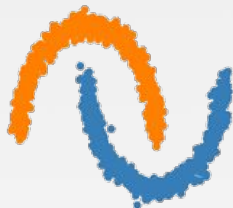
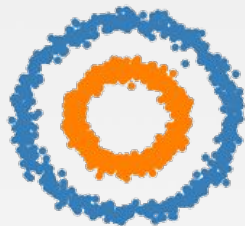


GHOST

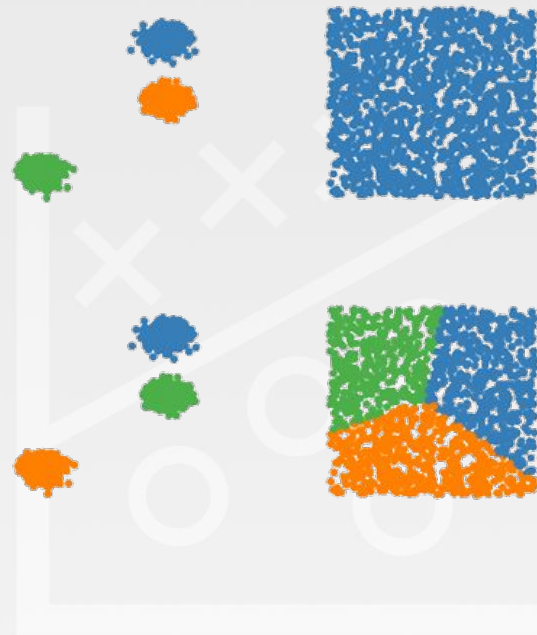
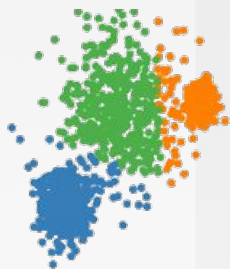
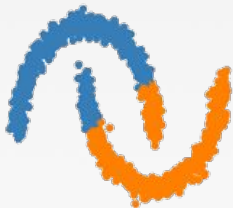
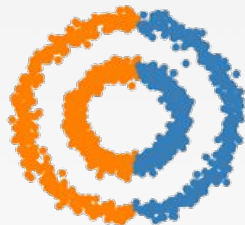
Group of Horribly Optimistic Statisticians



DBSCAN



k-means





GHOST

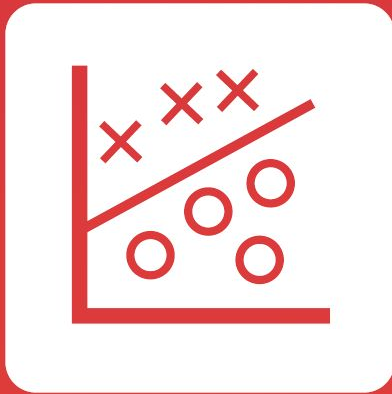
Group of Horribly Optimistic Statisticians

Dziękuję za uwagę

People with no idea
about AI, telling me my
AI will destroy the world

Me wondering why my
neural network is
classifying a cat as a dog..





GHOST

Group of Horribly Optimistic Statisticians