GHOST
Group of Horribly Optimistic Statisticians

Intro to ML

#1 Wstęp

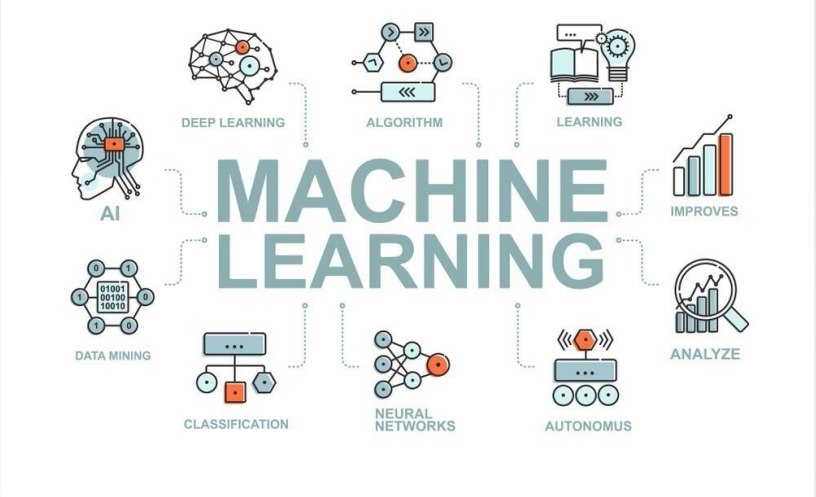
Agenda

1. Czym jest uczenie maszynowe?
2. Rodzaje Uczenia Maszynowego
3. Przegląd Modeli
4. Proces Uczenia Maszynowego
5. Zastosowania ML
6. Analiza Modelu
7. Wyzwania

Czym jest Uczenie Maszynowe?

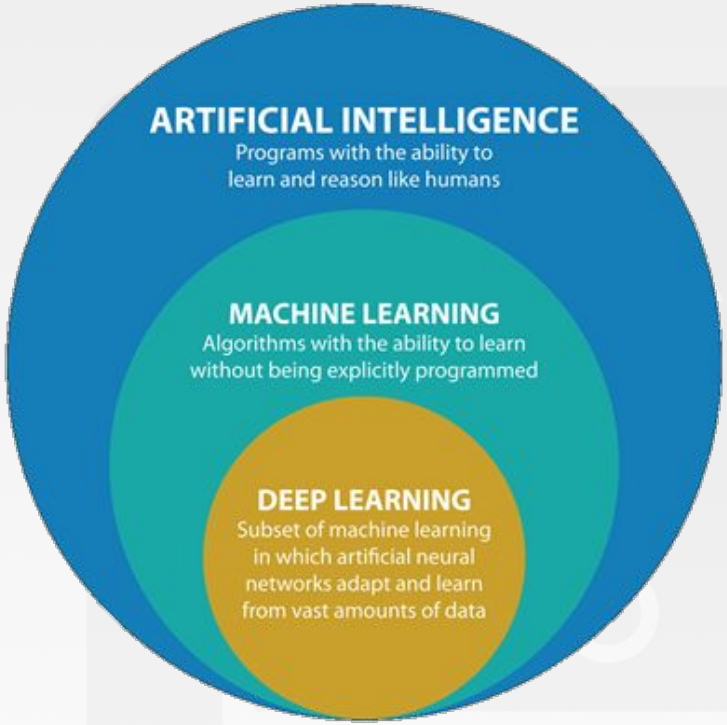
Uczenie maszynowe (ML) to rodzaj sztucznej inteligencji (AI), która pozwala komputerom uczyć się bez sprecyzowanego (explicit) programowania . Polega na wprowadzaniu danych do algorytmów, które mogą następnie identyfikować wzory w datasetach i tworzyć prognozy na podstawie nowych danych.

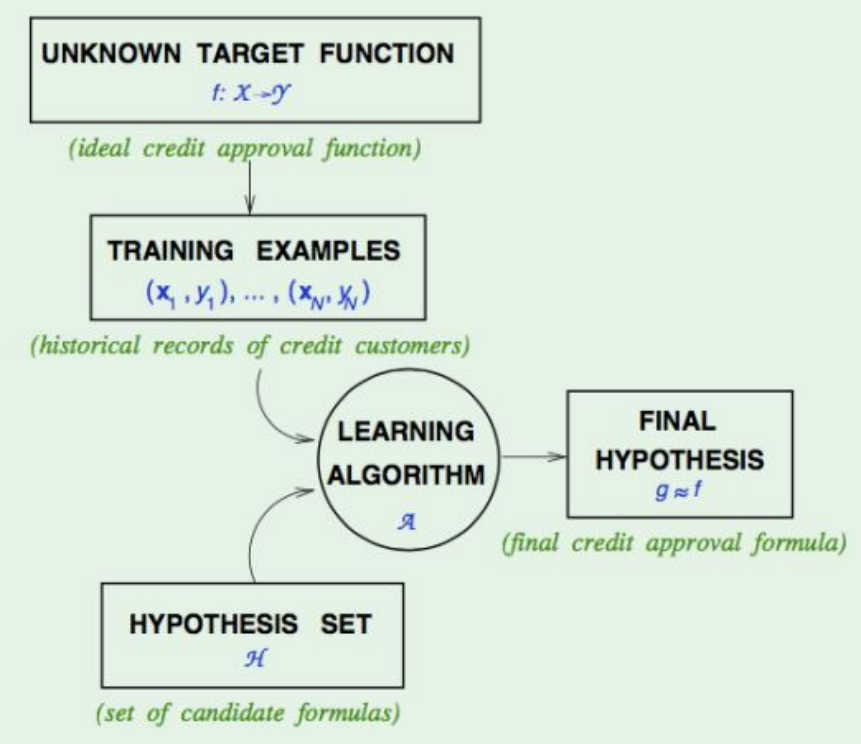
Główna idea: automatyczne znajdowanie wzorców w danych bez programowania "ręcznego"



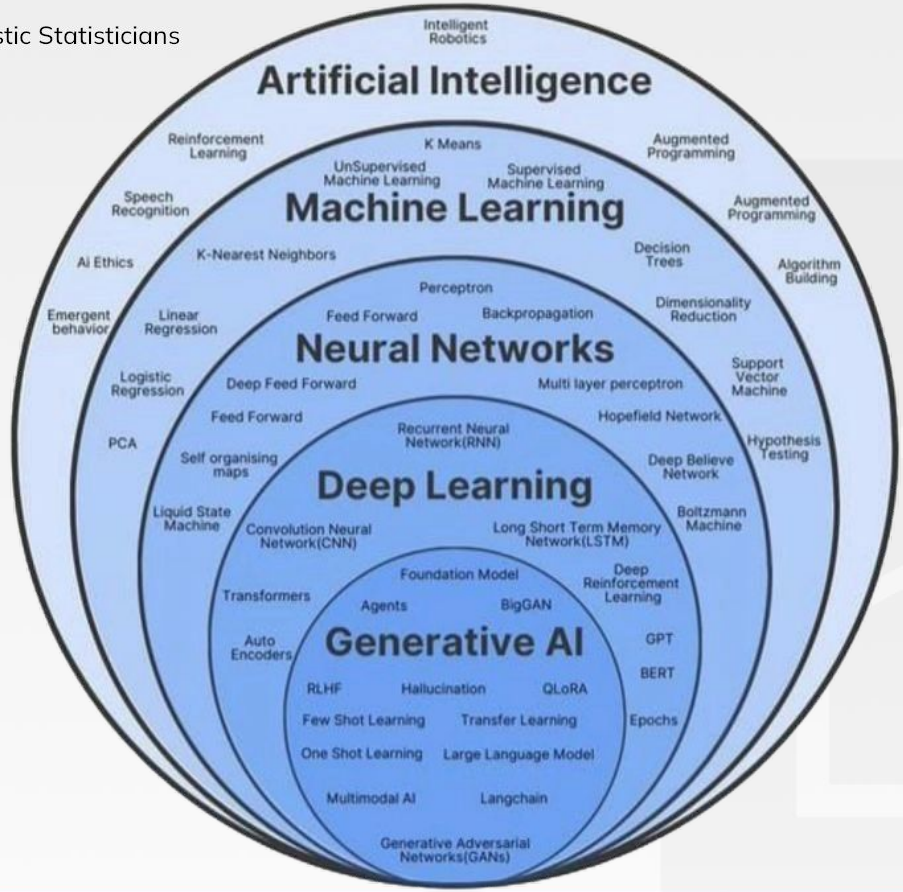
Kiedy używamy ML?

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





- Target function – wzór w danych
- Training examples – rzędy danych
- Hypothesis – oszacowana funkcja
- Learning algorithm + Hypothesis set = Learning Model

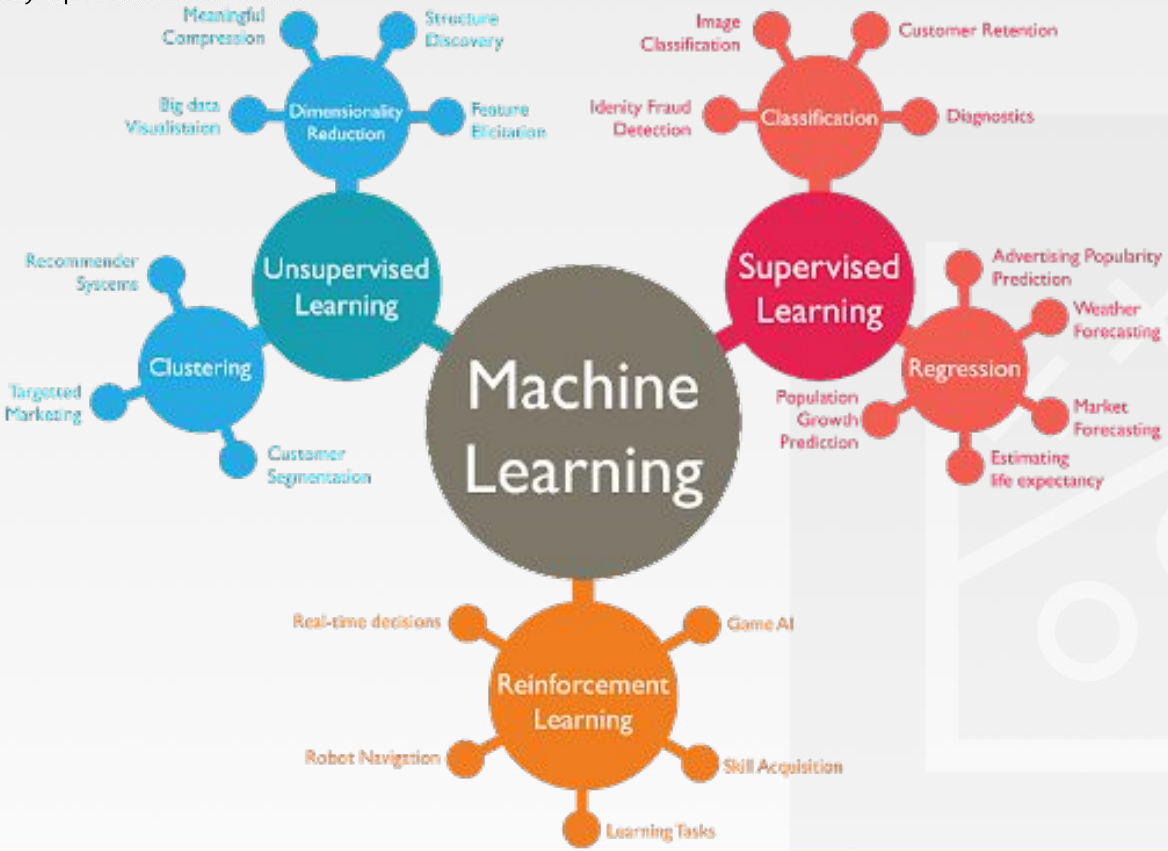

Intelligent Robotics
Artificial Intelligence
Reinforcement Learning
K Means
Augmented Programming
UnSupervised Machine Learning
Supervised Machine Learning
Speech Recognition
Machine Learning
Augmented Programming
Ai Ethics
K-Nearest Neighbors
Decision Trees
Algorithm Building
Perceptron
Emergent behavior
Linear Regression
Feed Forward
Backpropagation
Dimensionality Reduction
Neural Networks
Logistic Regression
Deep Feed Forward
Multi layer perceptron
Support Vector Machine
Feed Forward
Hopefield Network
Recurrent Neural Network(RNN)
PCA
Hypothesis Testing
Self organising maps
Deep Believe Network
Deep Learning
Liquid State Machine
Convolution Neural Network(CNN)
Long Short Term Memory Network(LSTM)
Boltzmann Machine
Foundation Model
Deep Reinforcement Learning
Transformers
Agents
BigGAN
Auto Encoders
Generative AI
GPT
BERT
RLHF
Hallucination
QLoRA
Few Shot Learning
Transfer Learning
Epochs
One Shot Learning
Large Language Model
Multimodal AI
Langchain
Generative Adversarial Networks(GANs)

Dlaczego ML jest ważne?

1. Automatyzacja zadań powtarzalnych i czasochłonnych.
2. Wzrost precyzji w diagnostyce medycznej, wykrywaniu oszustw etc.
3. Ulepszone rekomendacje np. Netflix, YouTube, e-commerce.
4. Optymalizacja procesów biznesowych i operacyjnych.

GHOST
Group of Horribly Optimistic Statisticians

Rodzaje Uczenia Maszynowego


Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualisation
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Identity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Uczenie nadzorowane (Supervised Learning)

Zbiór danych jest zbiorem labelled (oznaczonych) przykładów

Celem algorytmu uczenia nadzorowanego jest wykorzystanie zbioru danych do stworzenia modelu, który przyjmuje wektor cech x jako dane wejściowe i generuje informacje umożliwiające wyprowadzenie etykiety dla tego wektora cech. Na przykład model stworzony przy użyciu zbioru danych dotyczącego ludzi mógłby przyjmować jako dane wejściowe wektor cech opisujący daną osobę i zwracać prawdopodobieństwo, że ta osoba choruje na raka.

Uczenie nienadzorowane (Unsupervised Learning)

W uczeniu nienadzorowanym dane nie mają etykiet. Algorytmy same odkrywają ukryte wzorce.

Celem algorytmu uczenia nienadzorowanego jest stworzenie modelu, który przyjmuje wektor cech x jako dane wejściowe i przekształca go albo w inny wektor, albo w wartość, którą można wykorzystać do rozwiązania praktycznego problemu.
Przykład: Grupowanie klientów sklepu online według preferencji zakupowych.

Semi-Supervised Learning

W uczeniu półnadzorowanym zbiór danych zawiera zarówno oznaczone, jak i nieoznaczone przykłady.

Zazwyczaj liczba nieoznaczonych przykładów jest znacznie większa niż liczba przykładów oznaczonych. Celem algorytmu uczenia półnadzorowanego jest taki sam jak cel algorytmu uczenia nadzorowanego. Liczy się na to, że wykorzystanie wielu nieoznaczonych przykładów może pomóc algorytmowi w stworzeniu lepszego modelu.

Uczenie wzmacniane (Reinforcement Learning)

Reinforcement Learning (RL) to gałąź uczenia maszynowego skoncentrowana na podejmowaniu decyzji w celu maksymalizacji skumulowanych nagród w danej sytuacji. W RL agent(maszyna) uczy się osiągać cel w środowisku z którym może wejść w interakcje wykonując akcje. Każda akcja ma informacje zwrotną w postaci nagrody lub kary.

GHOST

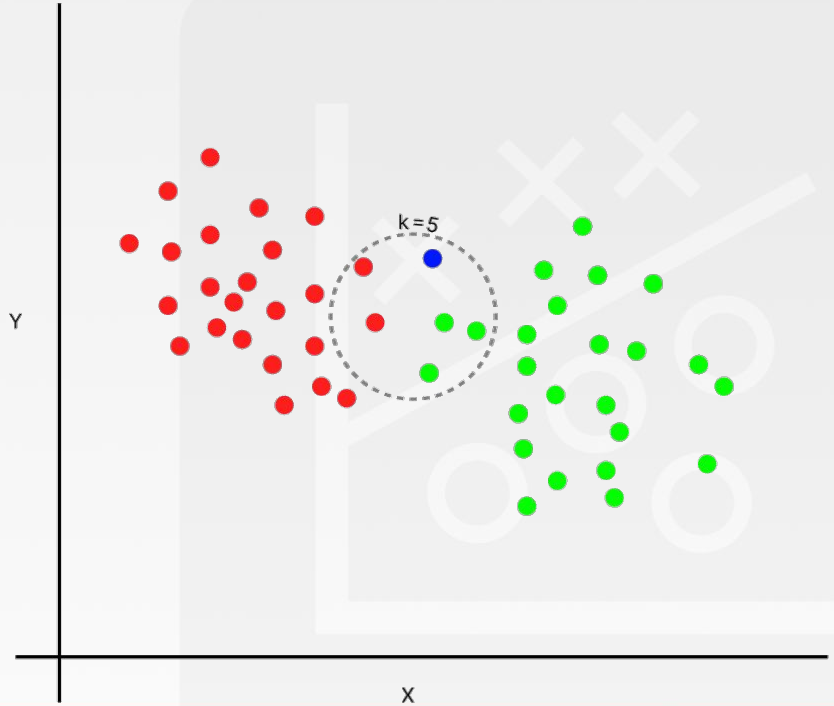
Group of Horribly Optimistic Statisticians

Przegląd Modeli

K-Nearest Neighbor – (KNN)

Algorytm KNN klasyfikuje obiekty na podstawie ich najbliższych sąsiadów.

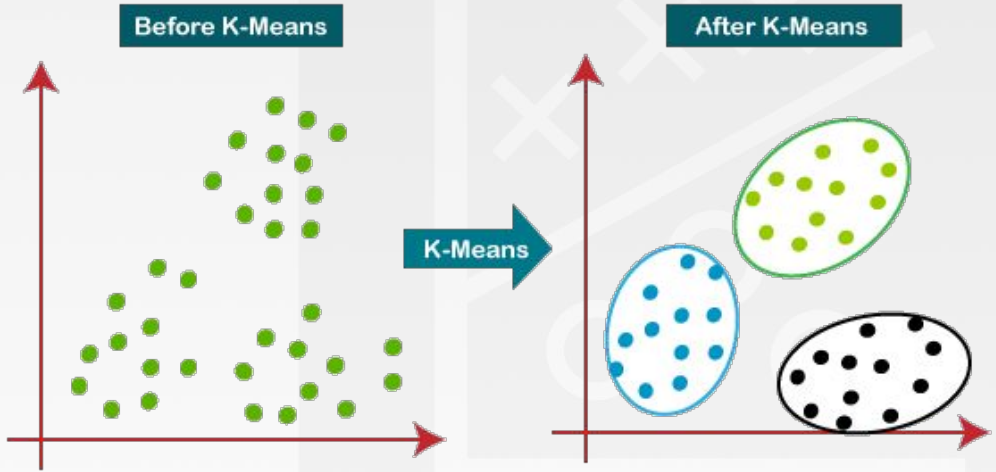
Przykład: Klasyfikacja zwierząt na podstawie ich cech, takich jak wielkość i kolor.



K-means clustering

Dzielenie danych w K różnych od siebie clusterów na podstawie podobieństwa danych.

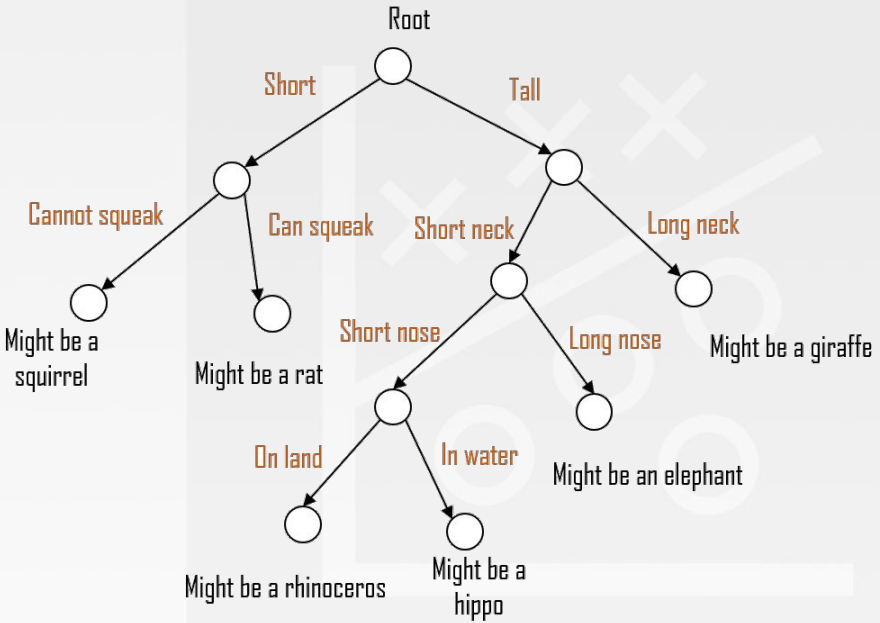
Przykład: Grupowanie typów klientów firmy i znajdowanie wspólnych cech.



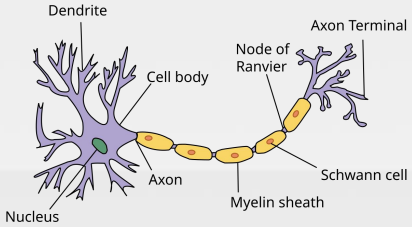
Drzewa Decyzyjne

Drzewa decyzyjne to algorytmy, które klasyfikują dane, tworząc serię decyzji opartych na cechach.

Przykład: Określanie gatunku zwierzaka na podstawie jego cech

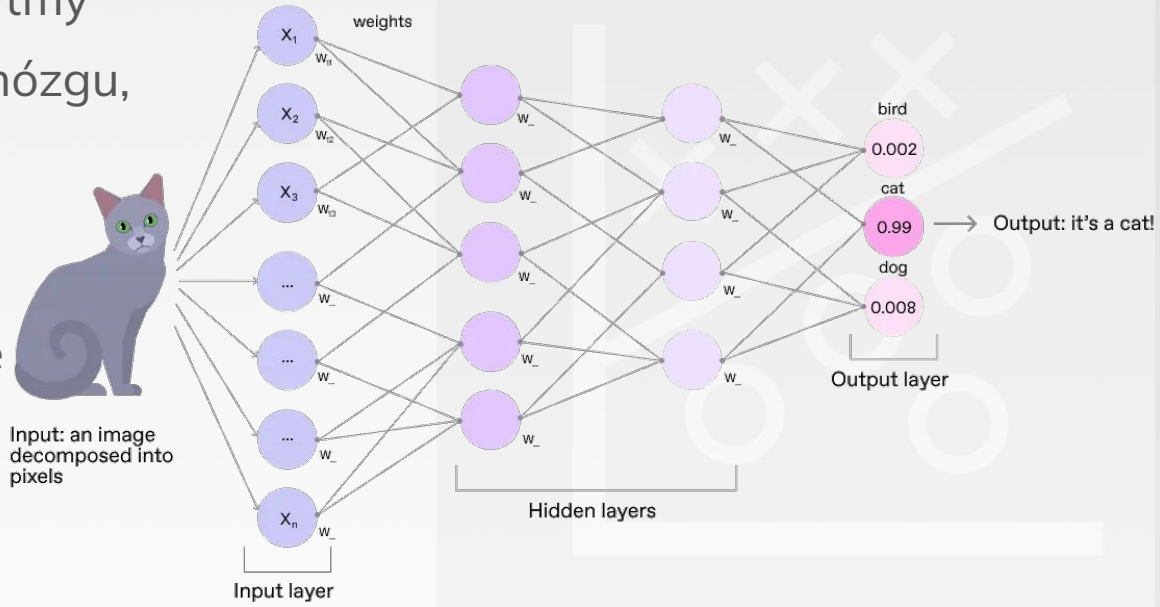


Sieci neuronowe



Sieci neuronowe to algorytmy inspirowane działaniem mózgu, zdolne do rozpoznawania złożonych wzorców.

Przykład: Rozpoznawanie gatunku zwierzecia.




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..
Dog

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

Zastosowania ML w Praktyce

1. Medycyna: Diagnostyka chorób i analiza obrazów medycznych.
2. Handel: Rekomendacje produktów i dynamiczna wycena.
3. Finanse: Wykrywanie oszustw oraz prognozowanie rynków.
4. Motoryzacja: Autonomiczne pojazdy i optymalizacja tras.
5. Produkcja: Predykcyjne utrzymanie ruchu i kontrola jakości.
6. Edukacja: Systemy adaptacyjnego nauczania i automatyczne ocenianie.

i wiele, wiele innych

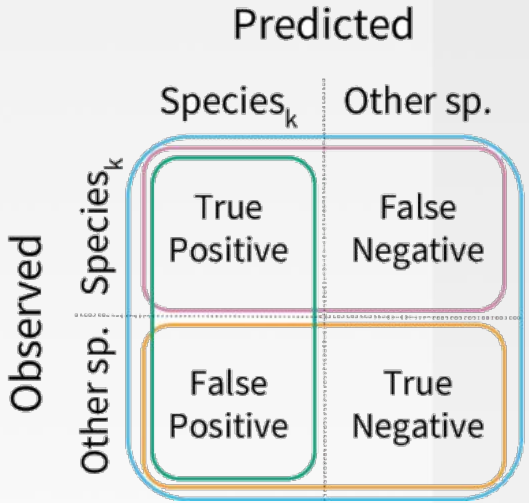
GHOST

Group of Horribly Optimistic Statisticians

Analiza Modelu

Dokładność (Accuracy)

Odsetek poprawnych przewidywań (zarówno pozytywnych, jak i negatywnych) w stosunku do wszystkich przewidywań; mierzy ogólną trafność modelu



$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

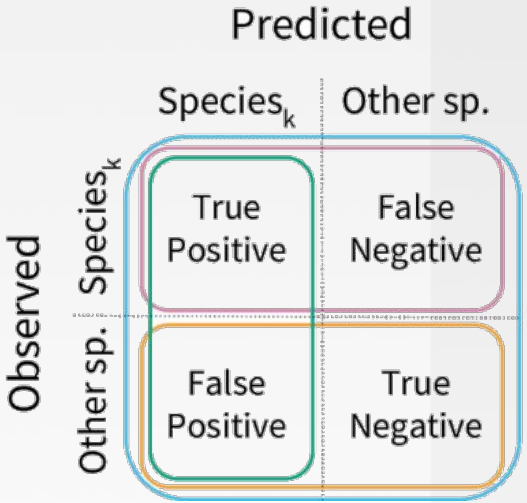
$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Precyzja (Precision)

Odsetek prawidłowo przewidzianych pozytywnych przypadków w stosunku do wszystkich przypadków przewidzianych jako pozytywne; określa, jak często model nie generuje fałszywych alarmów



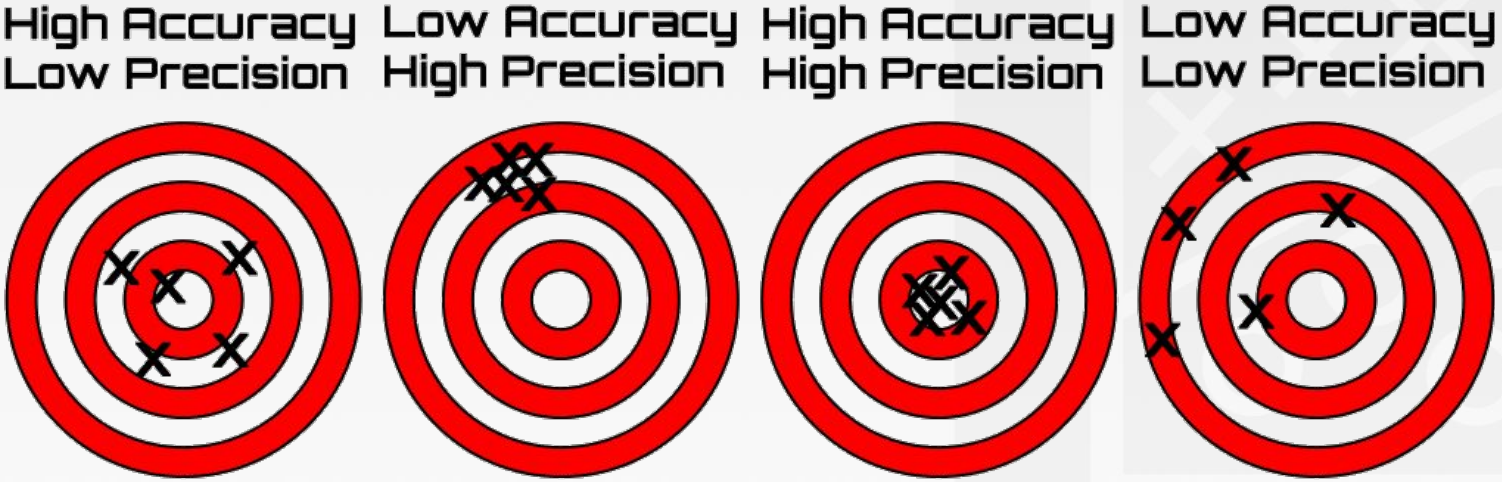
$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

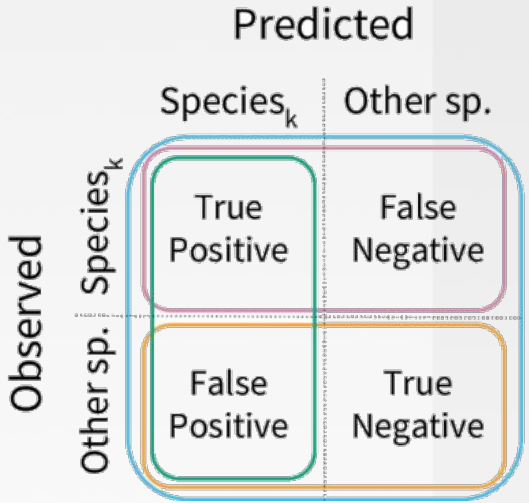
$$\text{Recall} = \frac{TP}{TP + FN}$$

Precision vs Accuracy



Czułość (Recall)

Odsetek prawidłowo przewidzianych pozytywnych przypadków w stosunku do wszystkich rzeczywistych pozytywnych przypadków; mierzy zdolność modelu do wykrywania wszystkich istotnych przypadków



$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

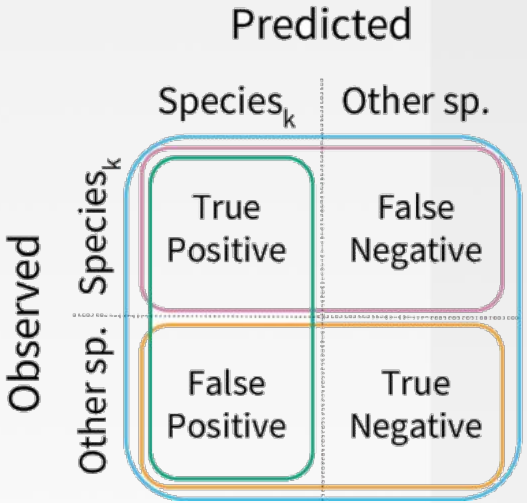
$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1 Score

Harmoniczna średnia precyzji i czułości, stanowiąca zbalansowaną miarę skuteczności modelu, zwłaszcza gdy dane są niezrównoważone (tzn. gdy występuje różnica między liczbą pozytywnych i negatywnych przykładów



$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

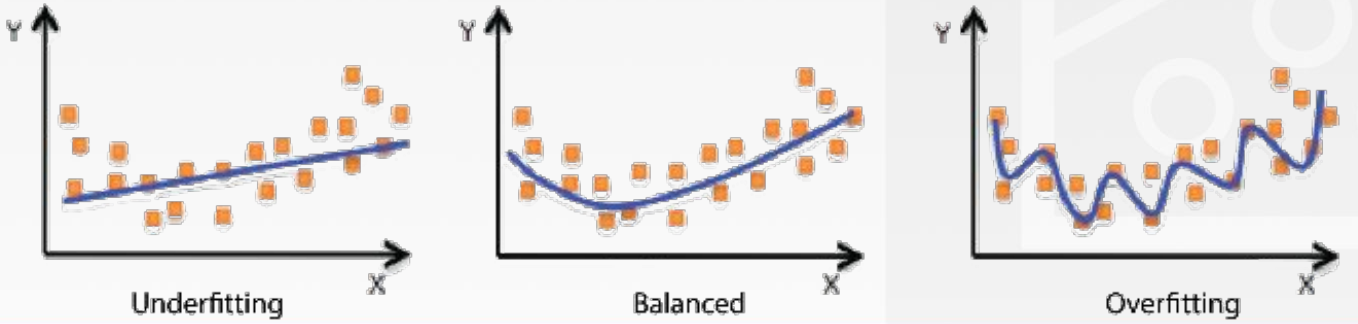
GHOST
Group of Horribly Optimistic Statisticians

Wyzwania w Uczeniu Maszynowym

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

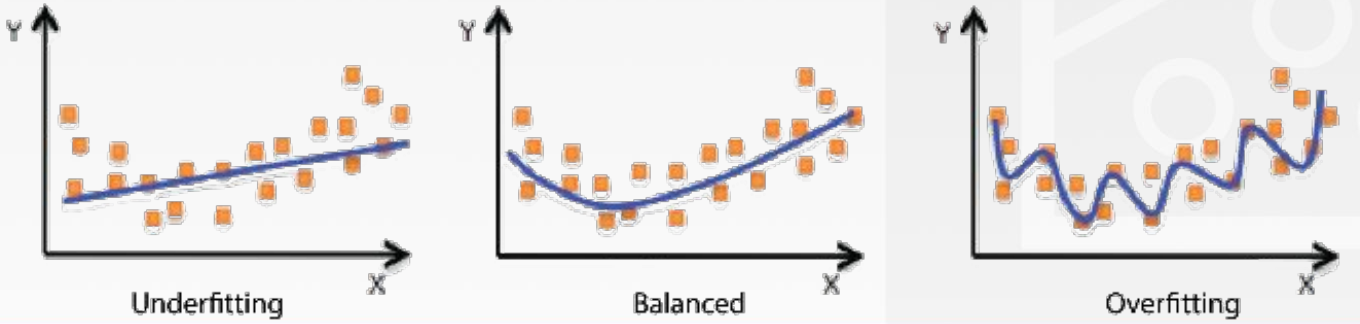
overfitting

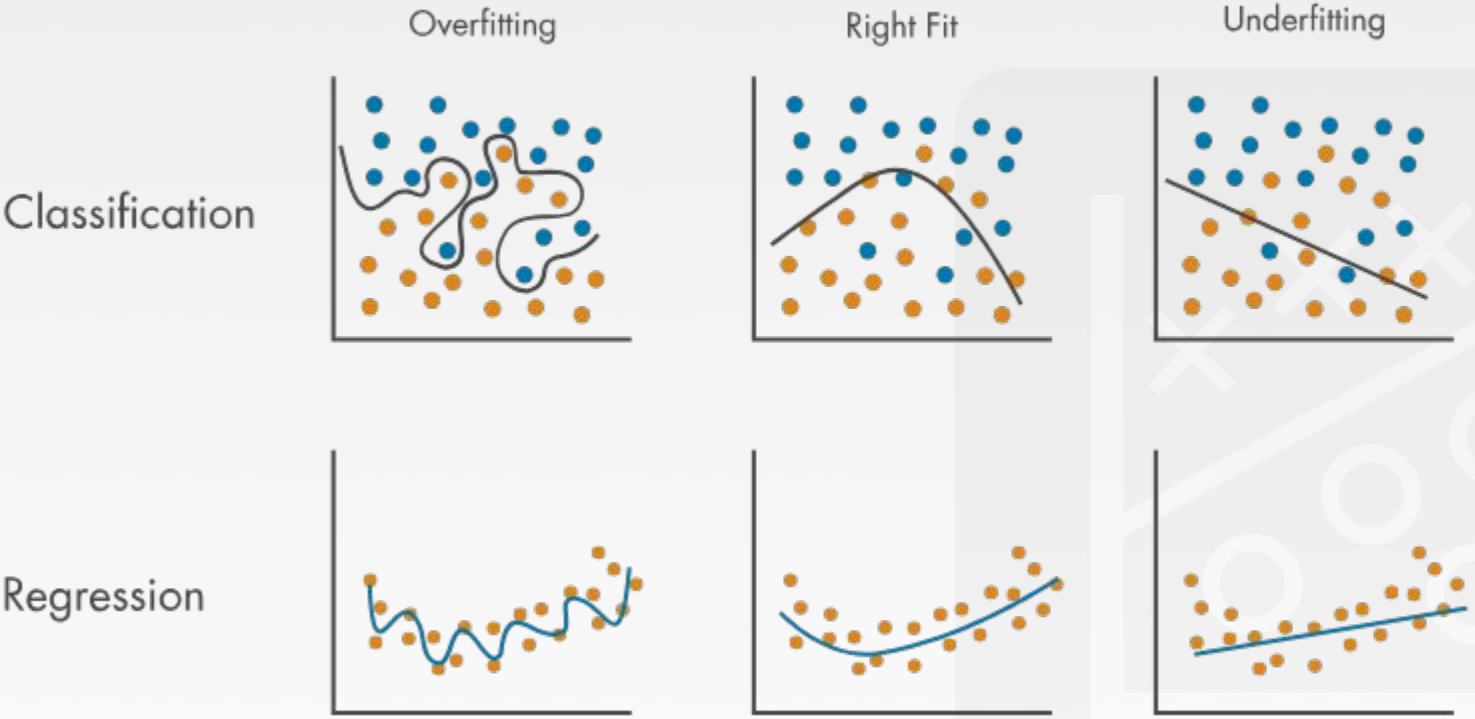
Overfitting pojawia się, gdy model uczy się bardzo dobrze danych treningowych, ale traci zdolność generalizacji do nowych danych. Oznacza to, że model jest "zbyt dopasowany" do szczegółów zbioru treningowego, co prowadzi do słabych wyników na danych testowych. Problem ten często rozwiązuje się poprzez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



underfitting

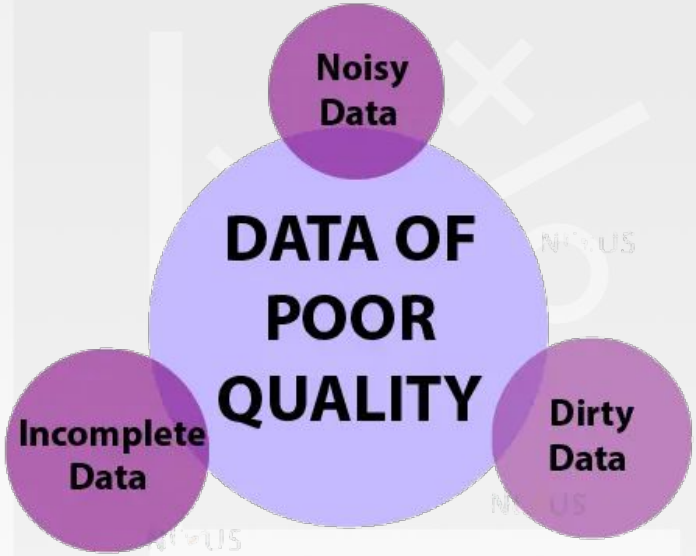
Underfitting pojawia się, gdy model nie jest w stanie dobrze dopasować się do danych treningowych, co skutkuje słabymi wynikami zarówno na danych treningowych, jak i testowych. Zwykle jest to efekt zbyt prostego modelu lub niewystarczającej liczby epok w procesie uczenia. Aby zmniejszyć ryzyko underfittingu,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




Overfitting
Right Fit
Underfitting
Classification
Regression

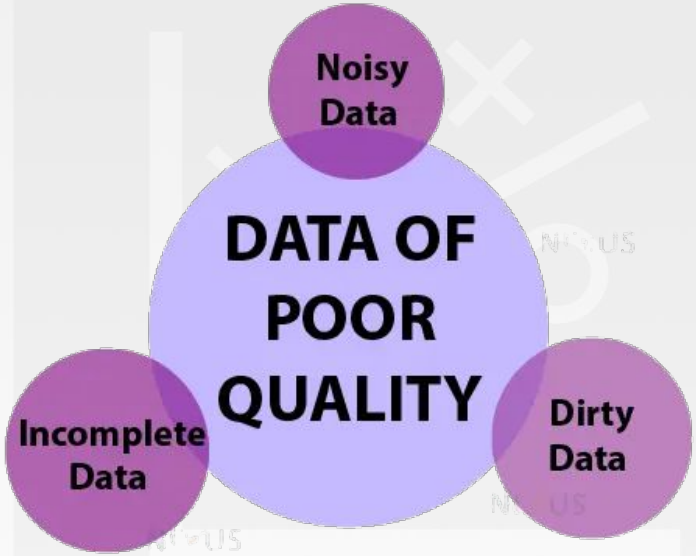
Niska jakość lub brakujące dane

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



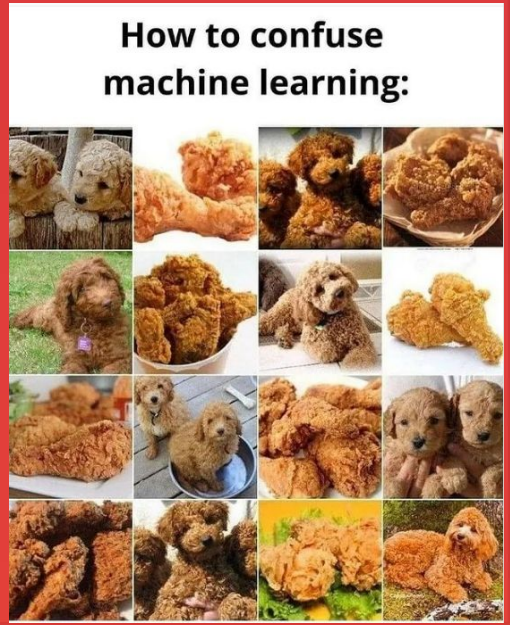
Problem ze zróżnicowaniem danych

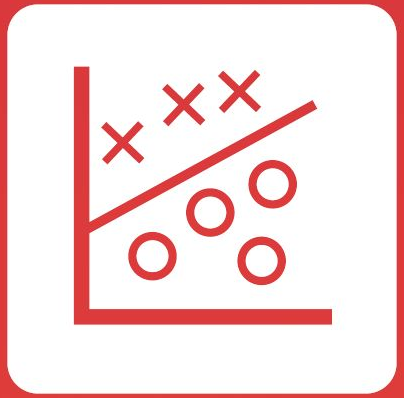
Niezrównoważone lub jednorodne dane mogą prowadzić do modeli, które nie są w stanie dobrze uogólnić. Jeśli np. pewna klasa jest nadreprezentowana, model może faworyzować tę klasę kosztem innych. W takim przypadku stosuje się metody balansowania klas lub wzbogacania zbioru danych, aby poprawić zdolność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



GHOST
Group of Horribly Optimistic Statisticians

Dziękuję za uwagę!





GHOST

Group of Horribly Optimistic Statisticians