

UCZENIE NADZOROWANE

# DRZEWA DECYZYJNE

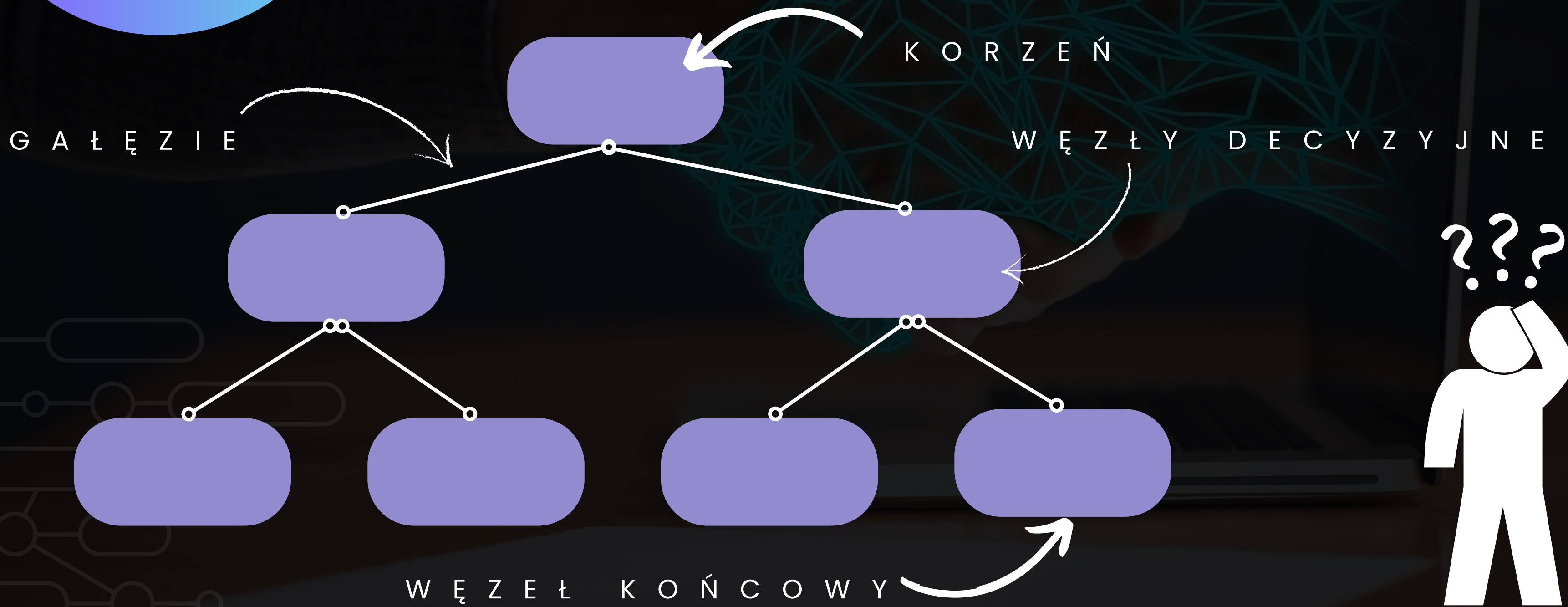
S K N   D A T A   S C I E N C E

A U T O R : M A R T Y N A   Ż U R , F I L I P   P I E T R Y G A

# AGENDA

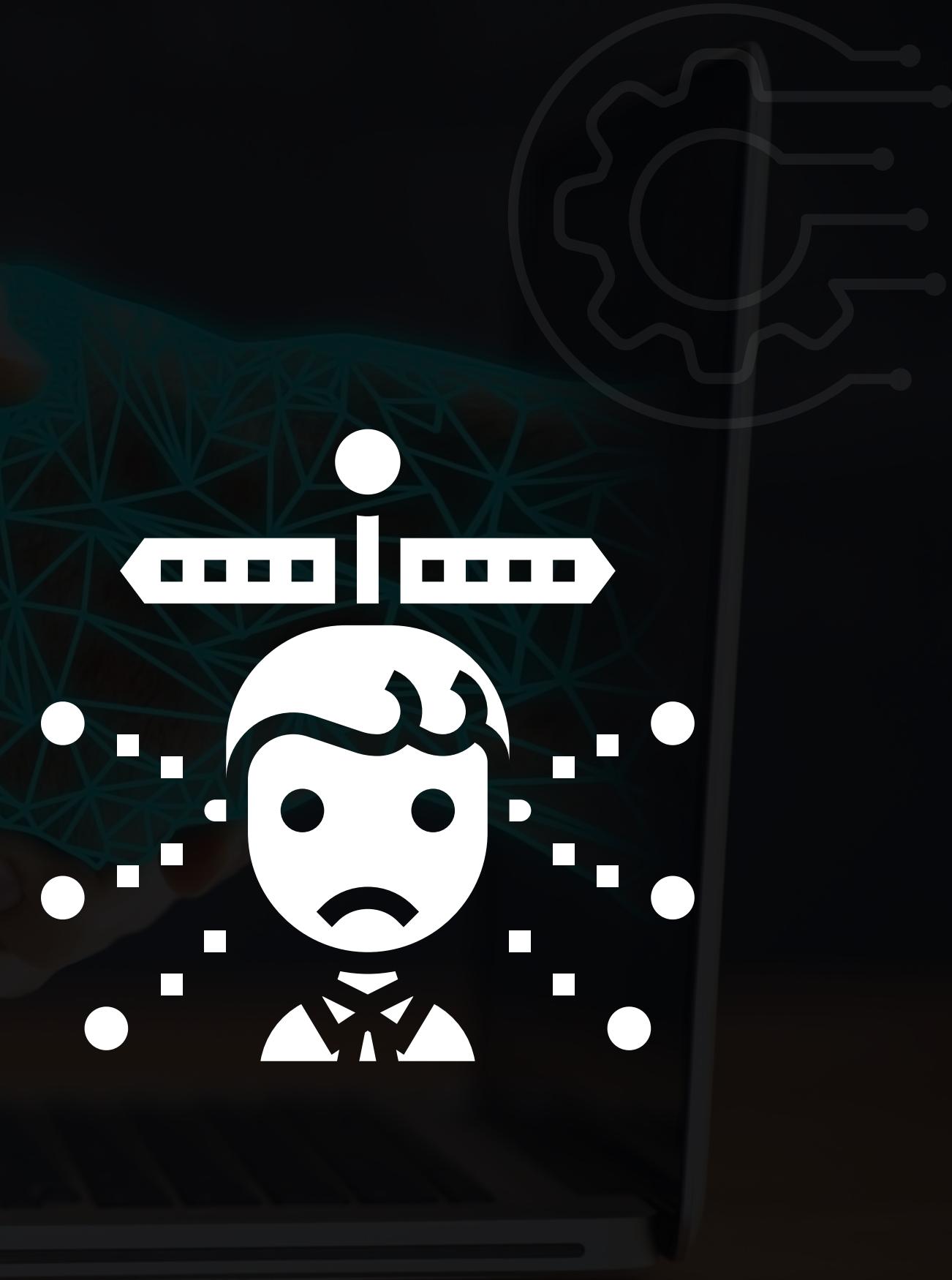
1. DRZEW A DECYZYJNE - CZYM SĄ W DATA SCIENCE?
2. WADY I ZALETY DRZEW DECYZYJNYCH
3. ALGORYTMY DRZEW DECYZYJNYCH
4. PODZIAŁ TYPÓW ALGORYTMÓW DRZEW DECYZYJNYCH
5. CART - ALGORYTM KLASYFIKACJI I REGRESJI
6. ANALIZA ALGORYTMU

# Czym są drzewa decyzyjne?



# DZREWO DECYZYJNE

to zasadniczo układ różnych wyników związanych z serią powiązanych ze sobą wyborów, jest narzędziem do modelowania predykcyjnego.



# Zalety drzewa decyzyjnego



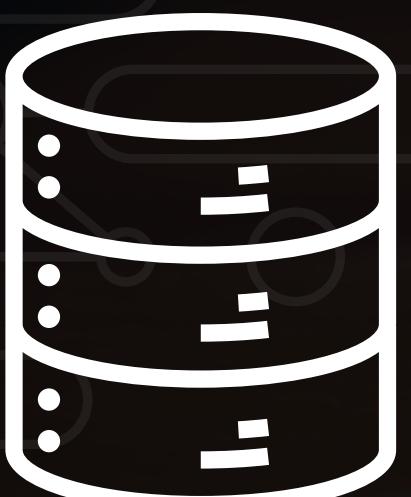
UNIwersalne



ŁATWA  
INTERPRETACJA



ŁATWA  
EDYCJA



ANALIZA  
RÓŻNYCH  
DANYCH



POMOC  
W ROZWAŻENIU  
PODEJMOWANYCH  
DECYZJI

# Wady drzewa decyzyjnego



NIESTABILNE



NIEDOKŁADNE



CIĘŻKORADZI  
SOBIE Z  
ZŁOŻONYMI  
OBliczeniami

# Algorytmy drzew decyzyjnych

C A R T - K L A S Y F I K A C J A I  
R E G R E S J A D R Z E W D E C Y Z Y J N Y C H

I T E R A T I V E D I C H O T O M I S E R I D 3

C 4 . 5 ( K O L E J N A I T E R A C J A I D 3 )

# CART

K L A S Y F I K A C Y J N E  
Z M I E N N E  
K A T E G O R Y T C Z N E

R E G R E S Y J N E  
Z M I E N N E  
C I Ą G Ł E

D Z R E W A  
D E C Y Z Y J N E



# Lasy losowe

RÓŻNICE MIĘDZYMISZKIMI:

|                  | DECISION TREE                      | RANDOM FOREST                                |
|------------------|------------------------------------|--|
| Interpretability | Easy to interpret                  | Hard to interpret                            |
| Accuracy         | Accuracy can vary                  | Highly accurate                              |
| Overfitting      | Likely to overfit data             | Unlikely to overfit data                     |
| Outliers         | Can be highly affected by outliers | Robust against outliers                      |
| Computation      | Quick to build                     | Slow to build<br>(computationally intensive) |

K I E D Y  
U Ż Y W A C ?

# Budowa drzew decyzyjnych

DRZEWO DECYZYJNE JEST BUDOWANE PODCZAS ETAPU TRENOWANIA. TO ALGORYTM OKREŚLA, JAK ZAIIMPLENTOWAĆ PODZIAŁY ORAZ JAKIE WARTOŚCI UWZGLĘDNIĆ. JEŚLI POWRÓCIMY DO PRZYKŁADU Z PŁASZCZEM, TO ALGORYTM OKREŚLI IDEALNĄ TEMPERATURĘ, NA PODSTAWIE KTÓREJ UTWORZYĆ ODGAŁĘZIENIE, A TAKŻE JAK DŁUGA POWINNA BYĆ PROGNOZA POGODY.

HIPERPARAMETRY DRZEW DECYZYJNYCH: MINIMALNĄ I MAKSYMALNĄ LICZBĘ LIŚCI, MAKSYMALNĄ GŁĘBOKOŚĆ BUDOWANEGO DRZEWA ORAZ AKCEPTOWALNY POZIOM JEDNORODNOŚCI W PODZIELONYCH ZBIORACH DANYCH.

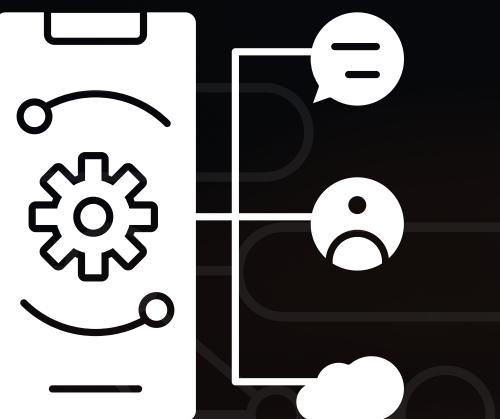
# Algorytm klasyfikacji drzewa decyzyjnego

O C E N A   N I E J E D N O R O D N O Ś C I  
W E Z Ł A

M I A R A   G I N I E G O

O C E N A   N I E S P Ó J N O Ś C I  
W E Z Ł A

M I A R A   B A Z U J Ą C A   N A  
O D C H Y L E N I U   K W A D R A T Ó W



# Źródła i polecane materiały

Materiały online:

- [1] [Python Machine Learning Tutorial #5 - Decision Trees and Random Forest Classification](#)
- [2] [Using ID3 Algorithm to build a Decision Tree to predict the weather \(opengenus.org\)](#)
- [3] [c1.10. Decision Trees — scikit-learn 1.3.2 documentation](#)
- [4] [Decision Trees in Machine Learning: Two Types \(+ Examples\) | Coursera](#)

Książki:

- [1] Dino Esposito, Francesco Esposito, Wprowadzenie do uczenia maszynowego według Esposito, Promise, Warszawa 2020
- [2] Jason Brownlee, Machine Learning Algorithms From Scratch with Python, 2020\*
- [3] Andreas Muller, Guido Sarah, Machine learning, Python i data science. Wprowadzenie, Helion, Gliwice 2023
- [4] Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana, Przetwarzanie języka naturalnego w praktyce. Przewodnik po budowie rzeczywistych systemów NLP, Helion, Gliwice 2023

\*pozycje dostępne w ramach koła