Introduction to scikit-learn

EuroSciPy 2025
18.08.2025, Kraków

Justyna Szydłowska-Samsel, University of Szczecin

Co-financed by the Minister of Science under the “Regional Excellence Initiative”

Regionalna Inicjatywa Doskonałości

Minister of Science and Higher Education
Republic of Poland

Tutorial agenda

What are we going to talk about?

1. Introduction
2. Why Python?
3. What is Scikit-learn?
4. What is machine learning?
5. Tutorial

Who am I?

- Research Assistant at University of Szczecin, Institute of Management
- Research focus : machine learning in business application and methods of teaching programming

Why Python?

Easy to learn

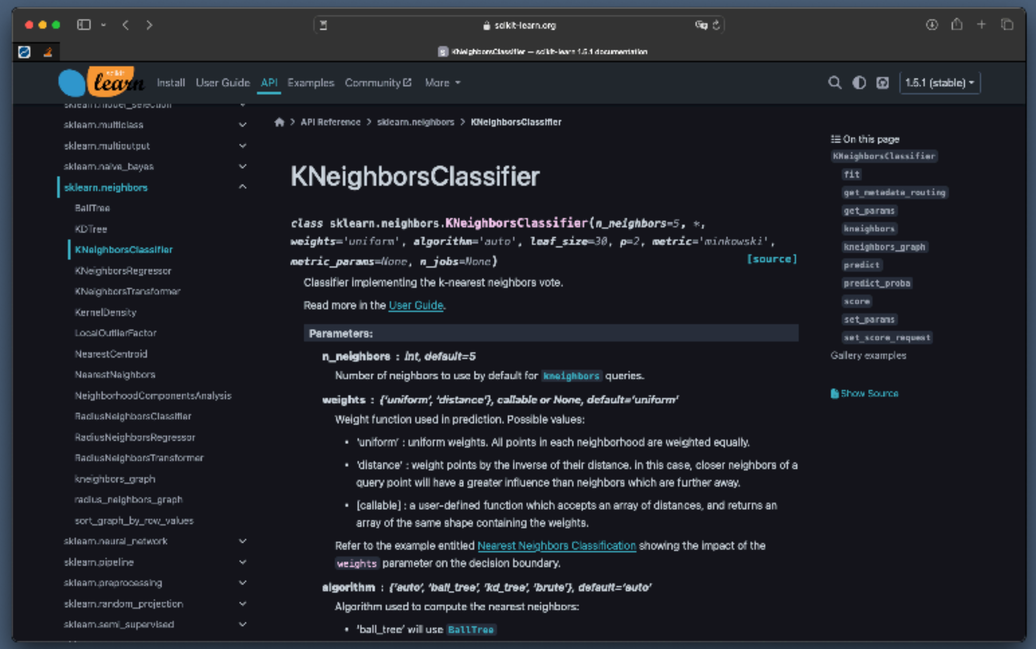
Powerful

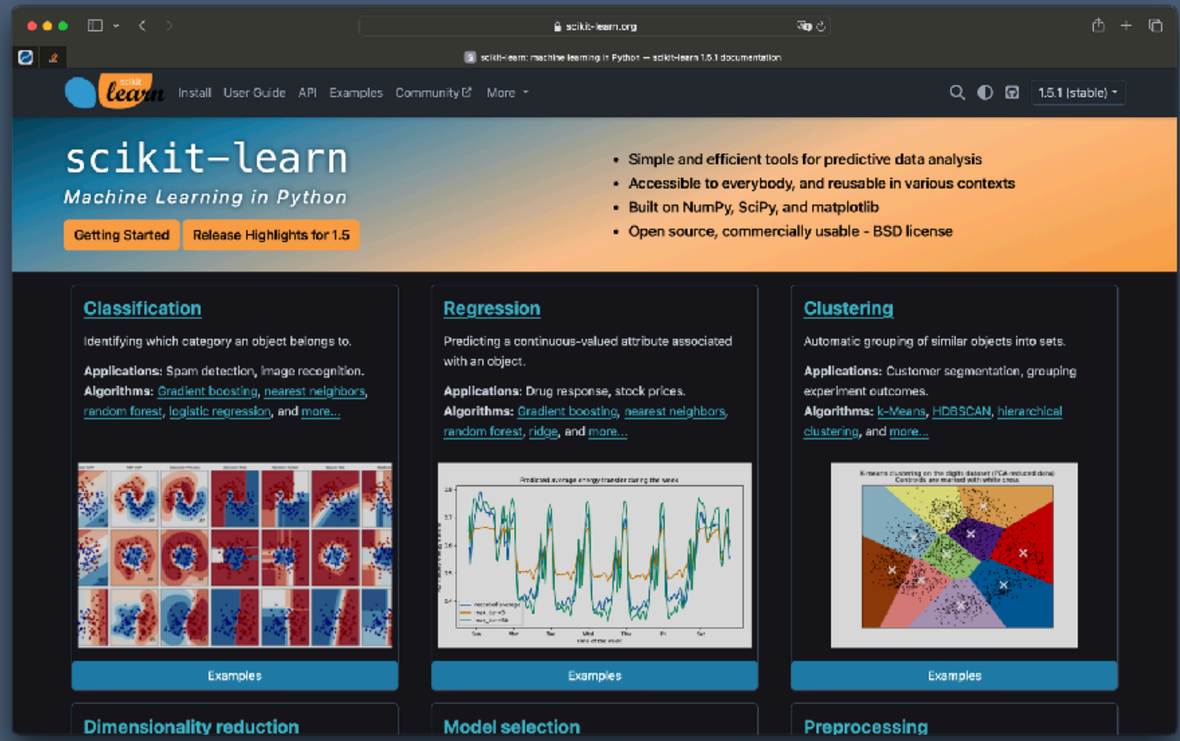
Free

python SOFTWARE FOUNDATION

Scikit-learn

Free and open-source machine learning library for Python. It’s build on NumPy, SciPy and matplotlib.



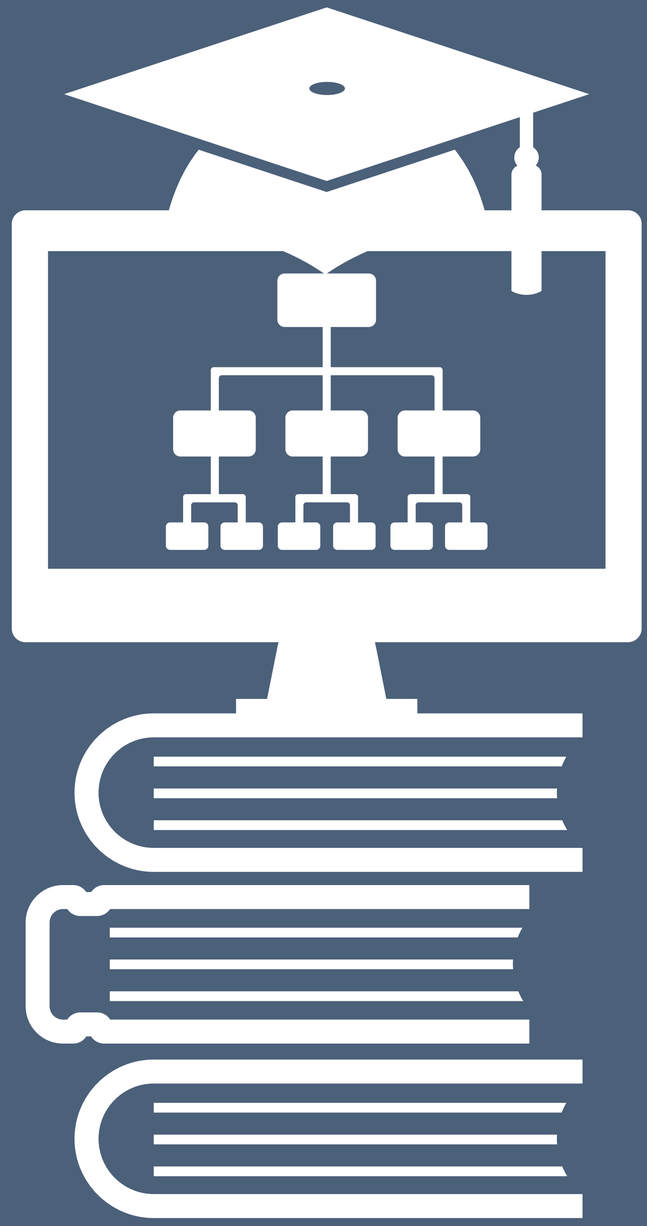


Machine learning

What is it and why is it so popular?

„Machine learning (ML) is a branch of artificial intelligence (AI) and computer science that focuses on the using data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy.” (IMB, 2024)

What usually follows is prediction. It is widely use in business, examples: Google, Netflix, Government institutions, banks, etc.



Types of machine learning

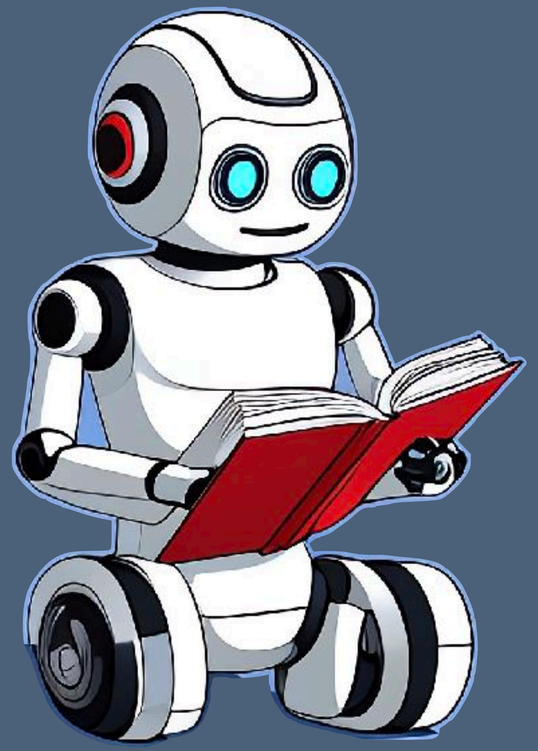
- Supervised
- Unsupervised
- Self-supervised
- Reinforcement
- Semi-supervised

Supervised machine learning

It’s a type of machine learning where model is trained on labeled dataset

Supervised machine learning algorithms:

- Regression algorithms
- Classification algorithms
- Neural networks
- Random forest algorithms

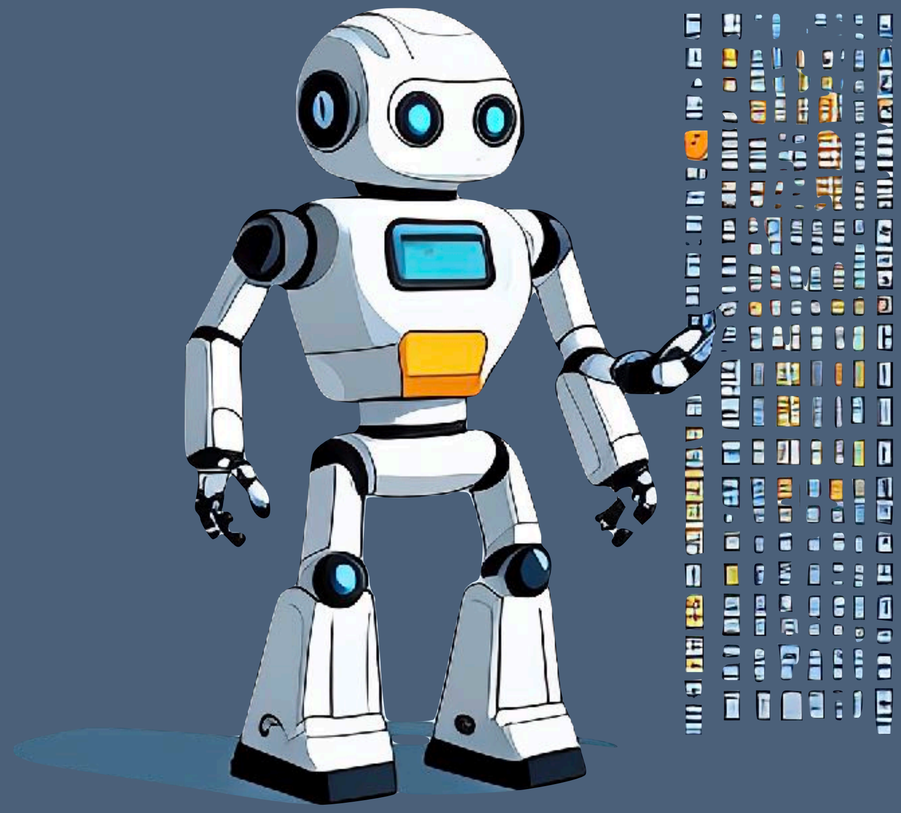


Unsupervised machine learning

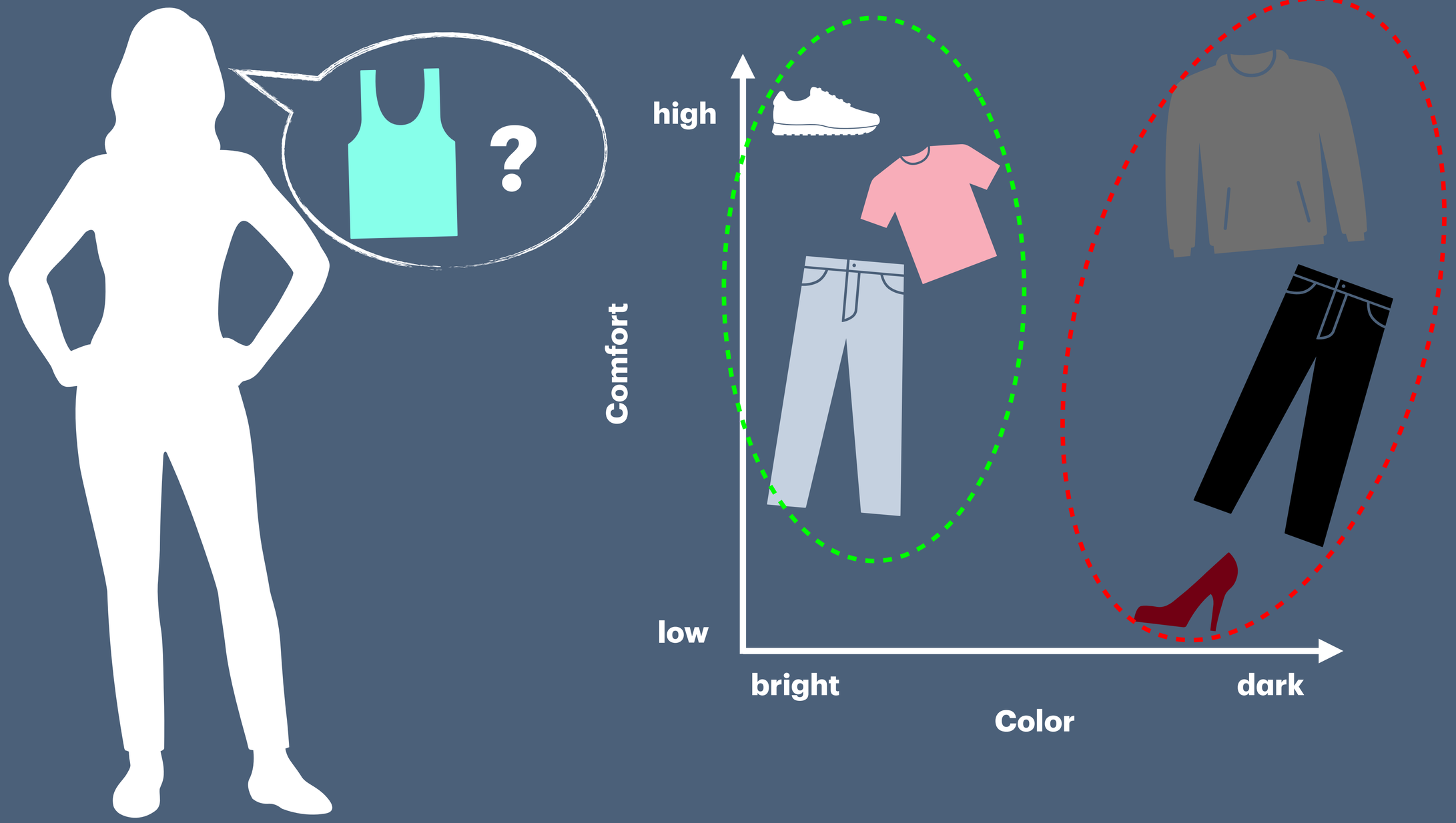
It’s a type of machine learning where model is trained on unlabeled dataset - trying to find patterns without explicit help.

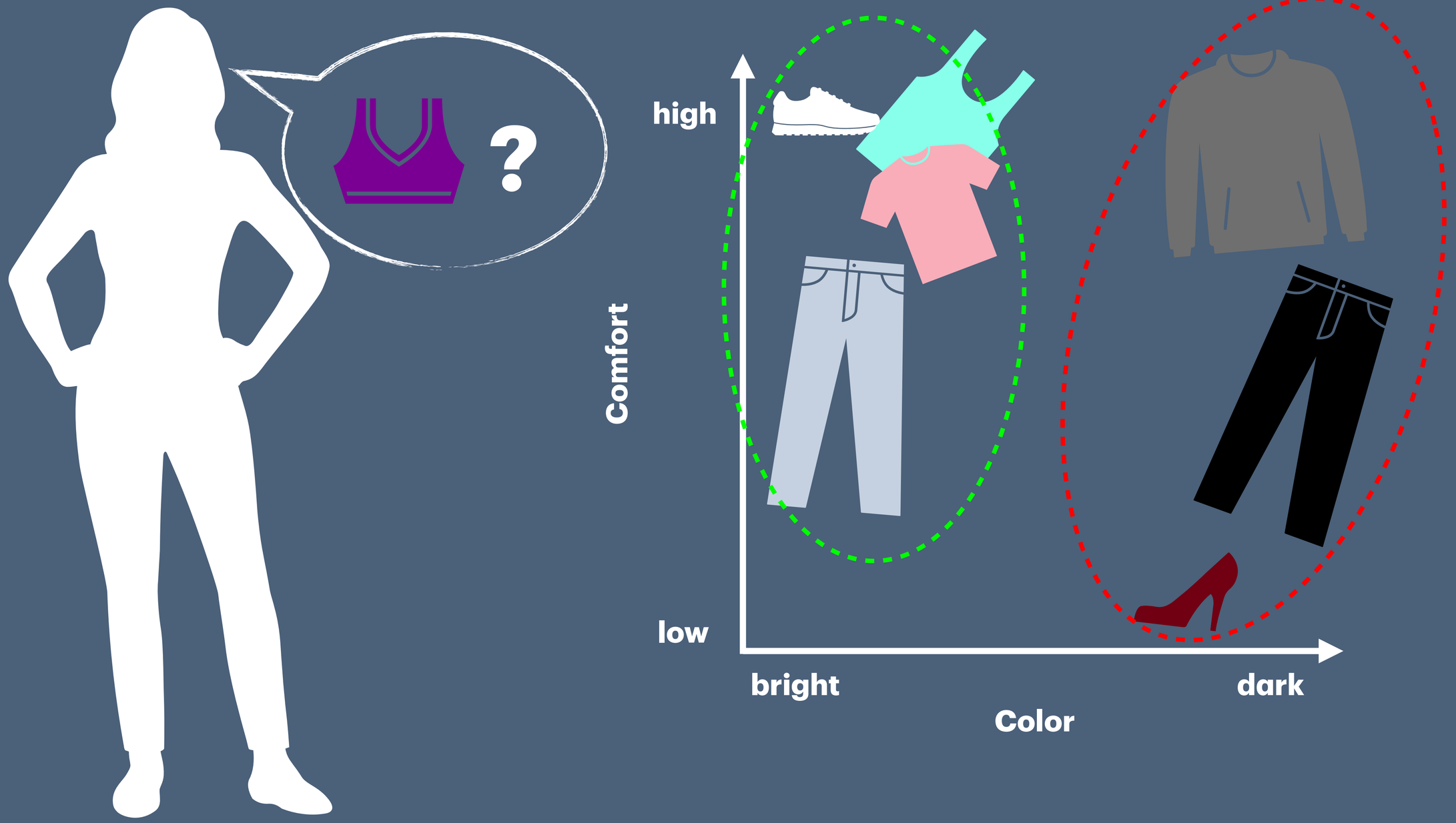
Unsupervised machine learning algorithms:

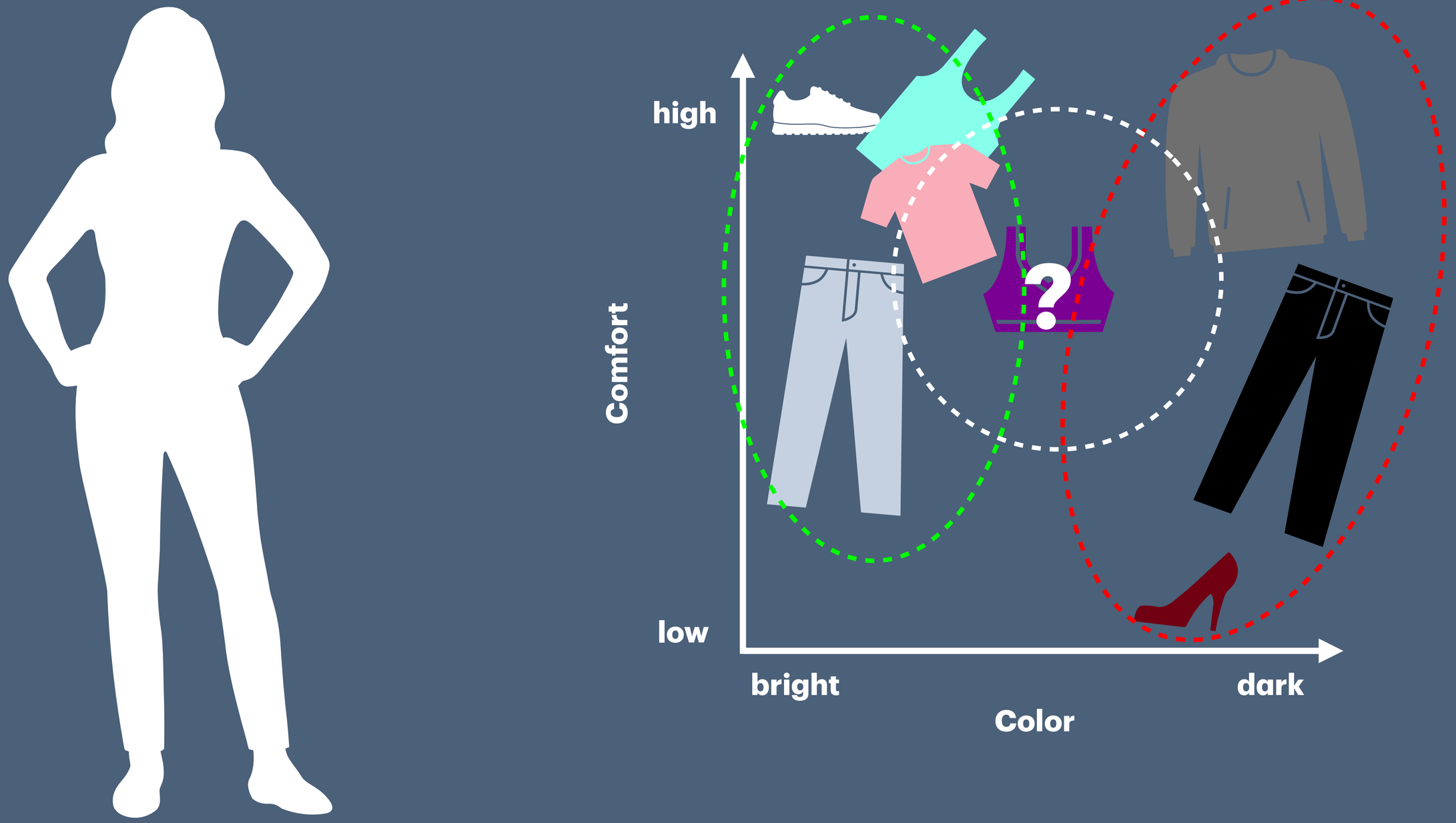
- K-means clustering
- Spectral clustering
- Hierarchical clustering

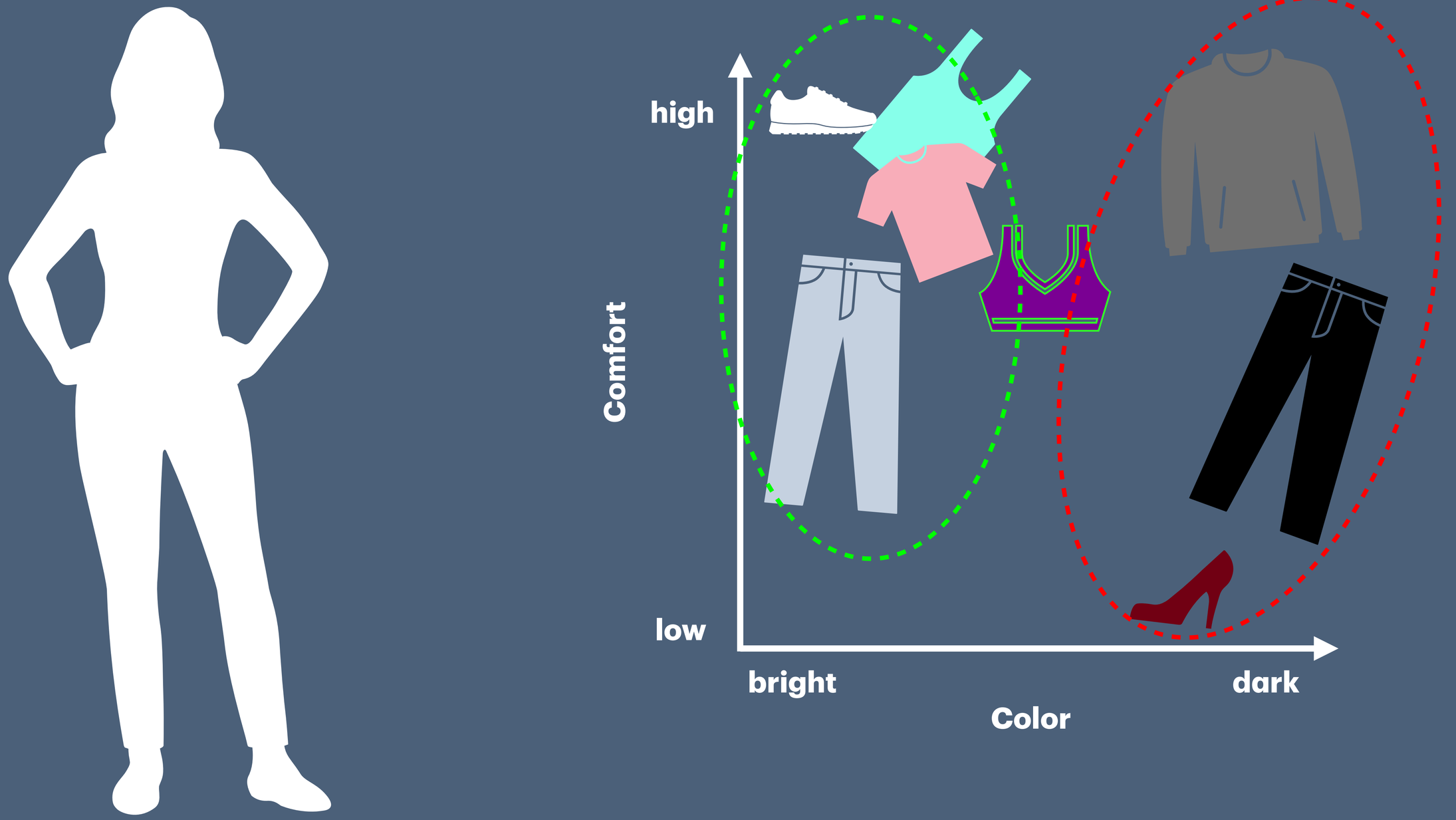


Which clothes will she buy?


?
high
low
Comfort
bright
dark
Color


?
high
low
Comfort
bright
dark
Color


high
low
Comfort
?
bright
dark
Color


K-Nearest Neighbors
high
Comfort
low
bright
dark
Color

Supervised machine learning

It’s a type of machine learning where model is trained on labeled dataset

Supervised machine learning algorithms:

- Regression algorithms
- Classification algorithms
- Neural networks
- Random forest algorithms

Supervised machine learning

- Regression algorithms
- Classification algorithms

Supervised machine learning

Regression:

- When what we want to predict is a continuous numerical value

Classification algorithms

- When what we want to predict is a category

Where to get data?

Google

Dataset Search

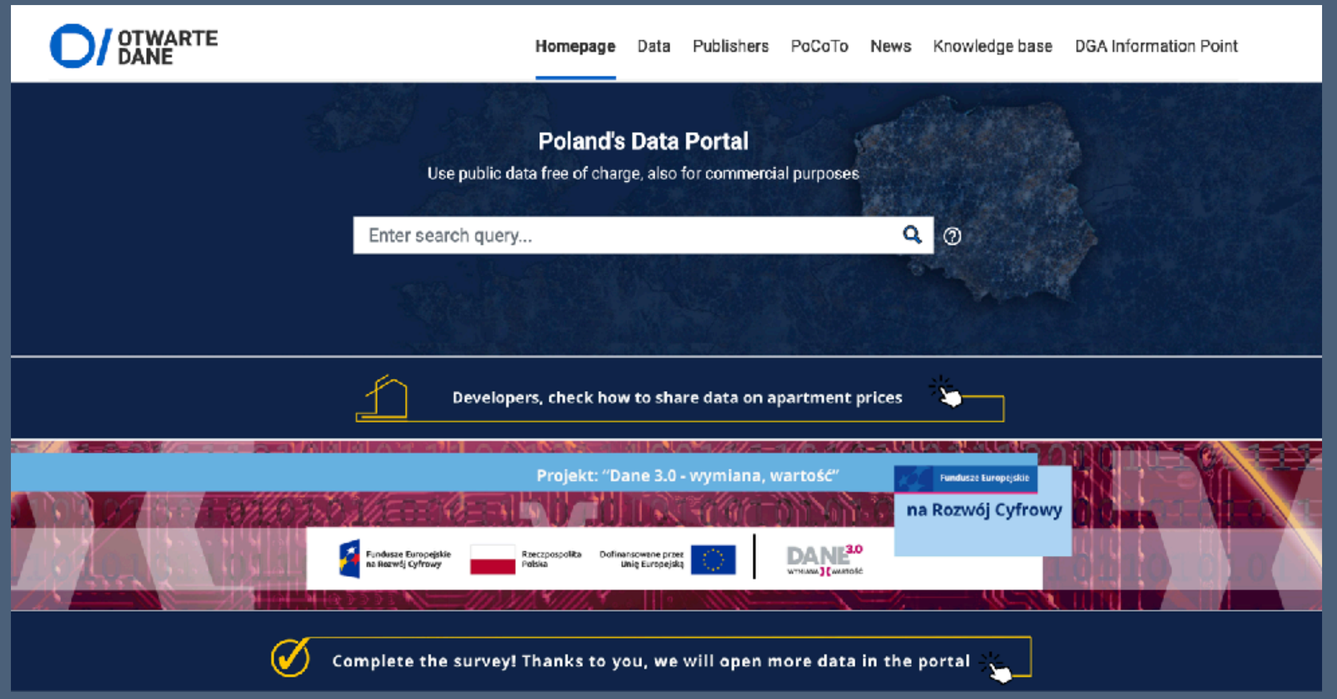
Search for Data Sets

Try boston education data or weather site:noaa.gov

Find out more about including your datasets in Dataset Search.

kaggle™

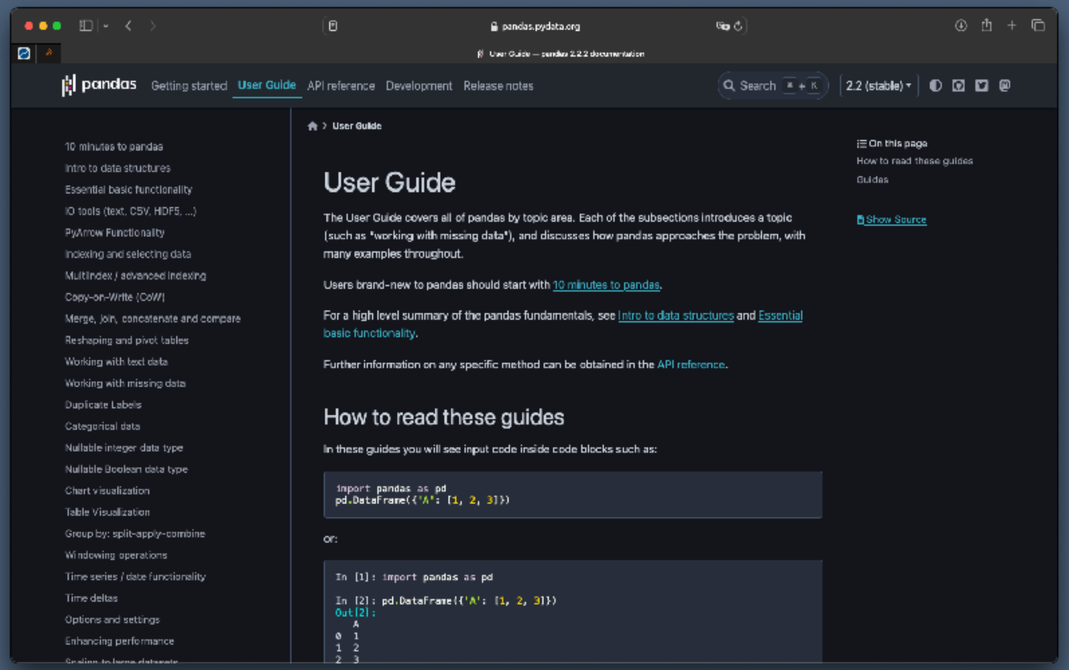
DATA.GOV

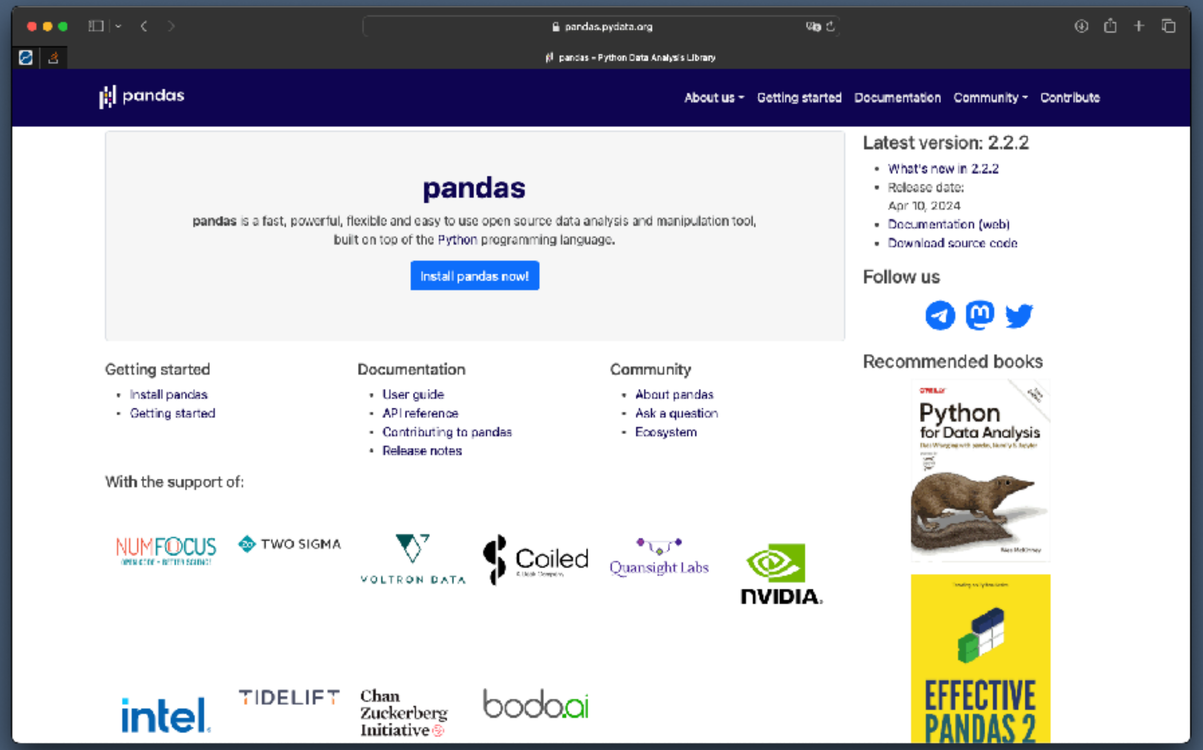


Tutorial

Pandas

Free and open source Python library for data manipulation and analysis.





https://github.com/jszydlowska/EuroSciPy2025.git