PERSPEKTYWY WOMEN IN TECH SUMMIT

CODECOOL

DATA MINING AND MACHINE LEARNING FOR BEGINNERS

Presenter: Michał Kruczkowski

Basic techniques for the carbon footprint reduction according to the Industry 4.0 concept

WORKSHOP GOALS

1

GIVE A TRY
DM & ML

2

IMPORTANT
BACKGROUND

3

REAL-WORLD
DATASETS

4

FACE THE
CHALLENGE

BEFORE WE START?!

1. LOG IN TO YOUR GOOGLE ACCOUNT
2. GOOGLE CALAB
3. FIND MY GITHUB
4. OPEN YOUR MIND

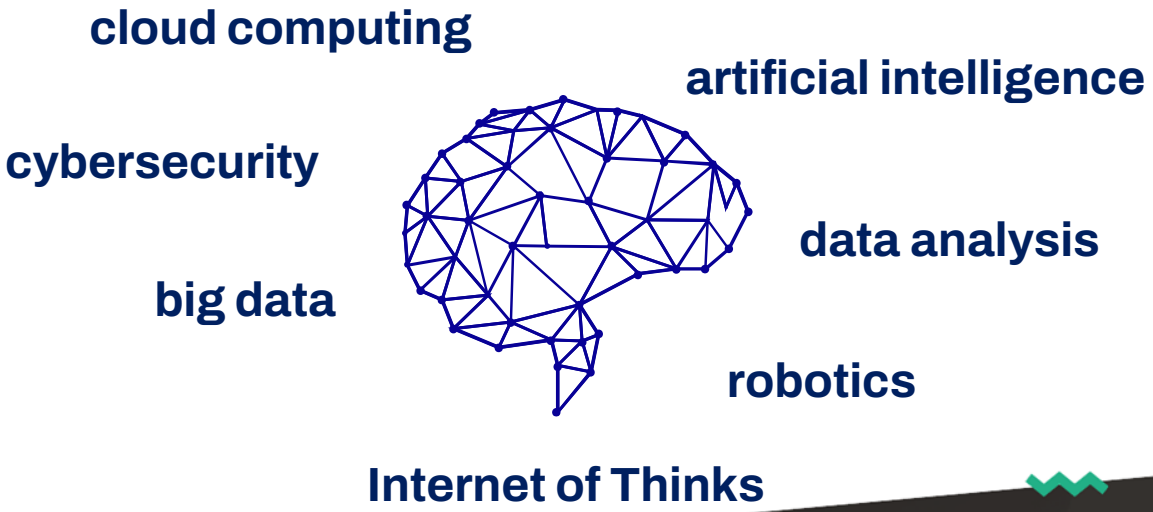
colab.research.google.com

https://github.com/MikiKru

Industry 4.0

CODECOOL

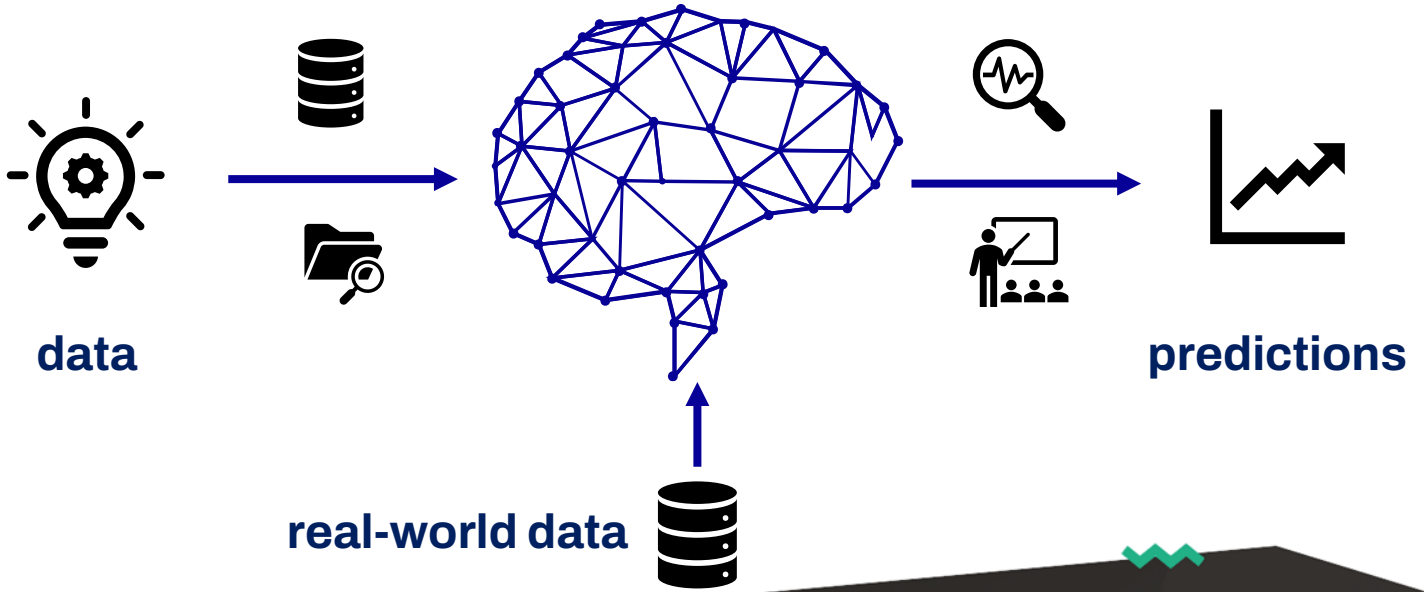
Industry 4.0 is integrating new technologies, including Internet of Things (IoT), cloud computing and analytics, and AI and machine learning into their production facilities and throughout their operations.



Machine Learning & Data Mining

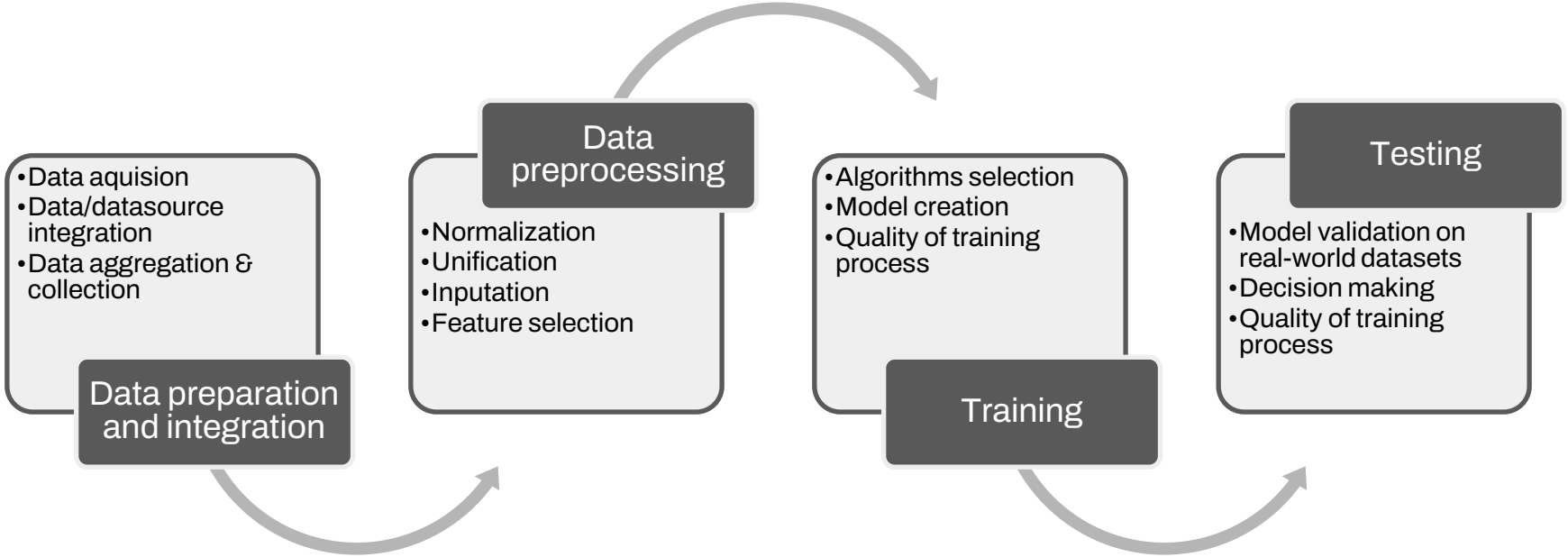
CODECOOL

Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed.



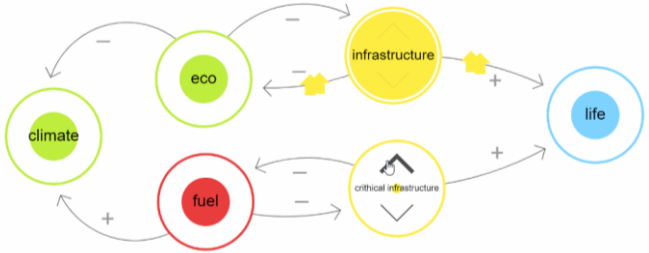
Modelling lifecycle

CODECOOL



Real problem?!

green energy



energy from fossil fuels

crithical infrastructure

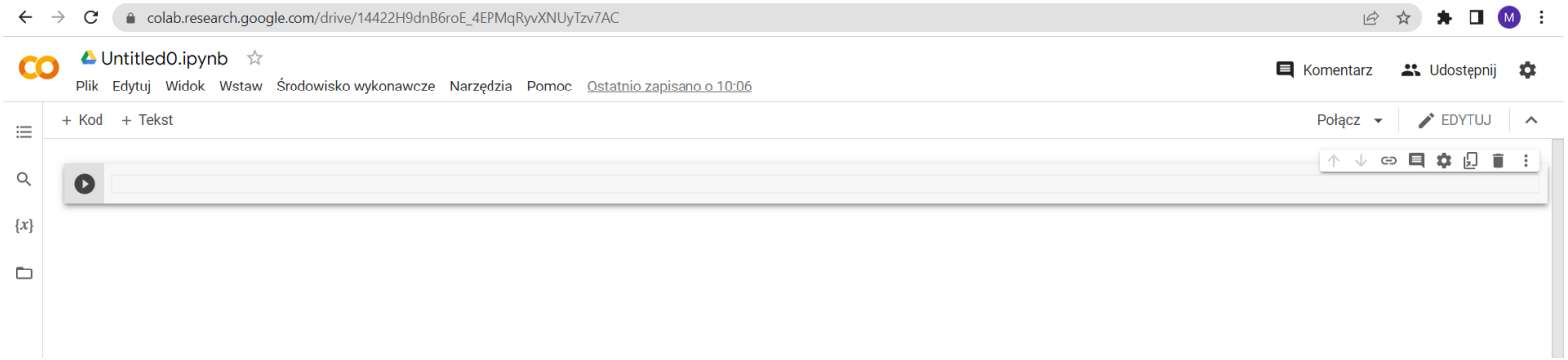
CODECOOL

guaranteed continuity of energy supply

Time to code it!

CODECOOL

https://colab.research.google.com/



Energy profile dataset

CODECOOL

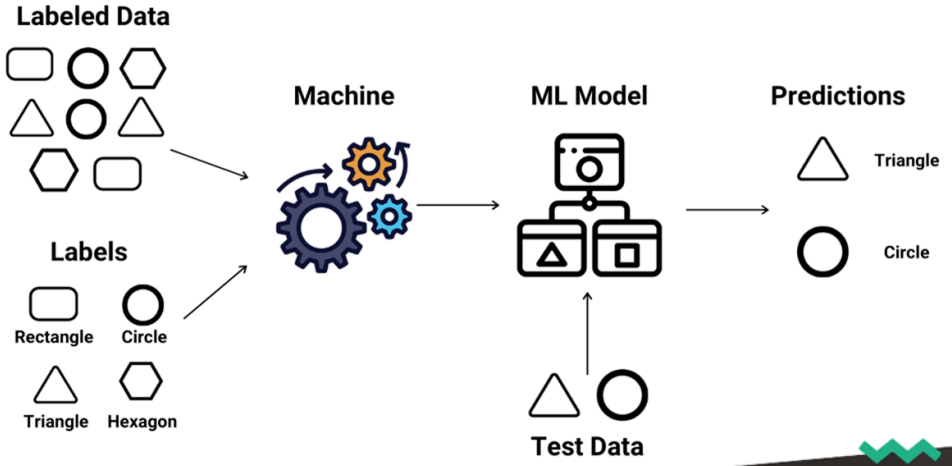
	Time_tick	PV	Demand	Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	AvgTemp_F	Yn	Yn_class
0	1	0.0	3.85	0	1	0	0	0	0	0	0	0	0	0	0	38.8	0.513650	1.0
1	2	0.0	4.08	0	1	0	0	0	0	0	0	0	0	0	0	38.8	0.506669	1.0
2	3	0.0	3.96	0	1	0	0	0	0	0	0	0	0	0	0	38.8	0.499512	0.0
3	4	0.0	3.96	0	1	0	0	0	0	0	0	0	0	0	0	38.8	0.492128	0.0
4	5	0.0	3.90	0	1	0	0	0	0	0	0	0	0	0	0	38.8	0.484515	0.0

Time_tick	*numer of the quarter of an hour in the day*
PV	*average energy production of a photovoltaics*
Demand	*average demand on energy*
Day	*numer of the day*
Jan-Nov	*months*
AvgTemp_F	*temperature in Farenheit scale*
Yn	*energy demand for the next 8 hours*
Yn_class	*target variable (0.0 – low, 1.0 - high)*

Supervised learning

CODECOOL

An approach to creating artificial intelligence (AI), where a computer algorithm is trained on input data that has been labeled for a particular output. The model is trained until it can detect the underlying patterns and relationships between the input data and the output labels, enabling it to yield accurate labeling results when presented with never-before-seen data.



K nearest neighbours

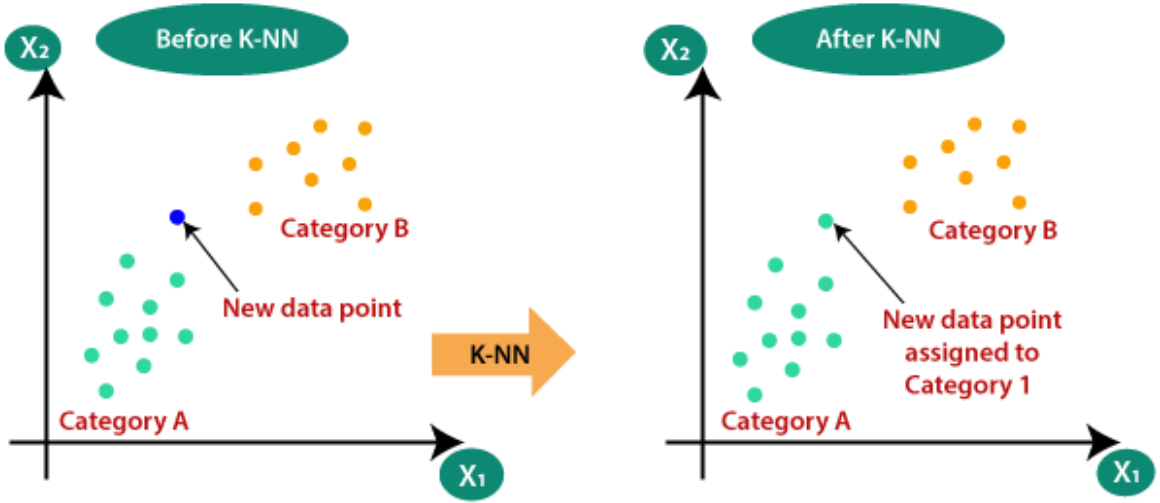
CODECOOL

QUALITY METRICS

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

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

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



		Predicted	
		Negative (N) -	Positive (P) +
Actual	Negative -	True Negatives (TN)	False Positives (FP) Type I error
	Positive +	False Negatives (FN) Type II error	True Positives (TP)

PERSPEKTYWY WOMEN IN TECH SUMMIT

CODECOOL

TIME TO CHALLENGE!

we invite you to our stand

Codecool – hall -1

deadline for sending solutions: 15.06.2022 (23:59)

PERSPEKTYWY WOMEN IN TECH SUMMIT

Thank you for your atention!

CODECOOL

MICHAŁ KRUCZKOWSKI

michal.kruczkowski@codecool.com