



Machine Learning

Katowice, 21st of June 2018

Expectation vs Reality

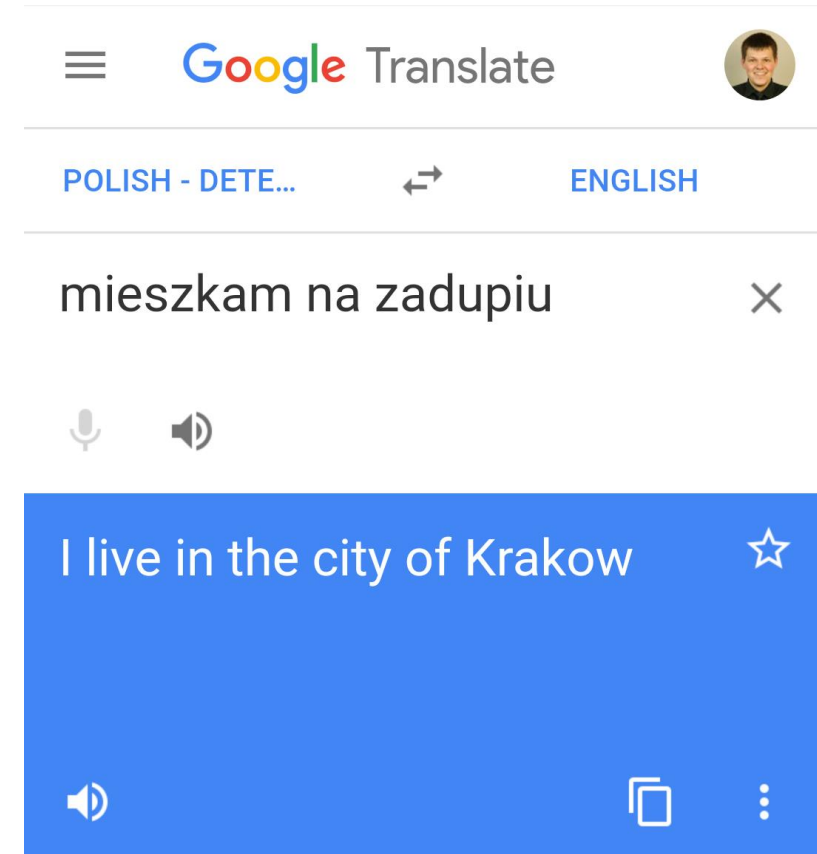


Expectation

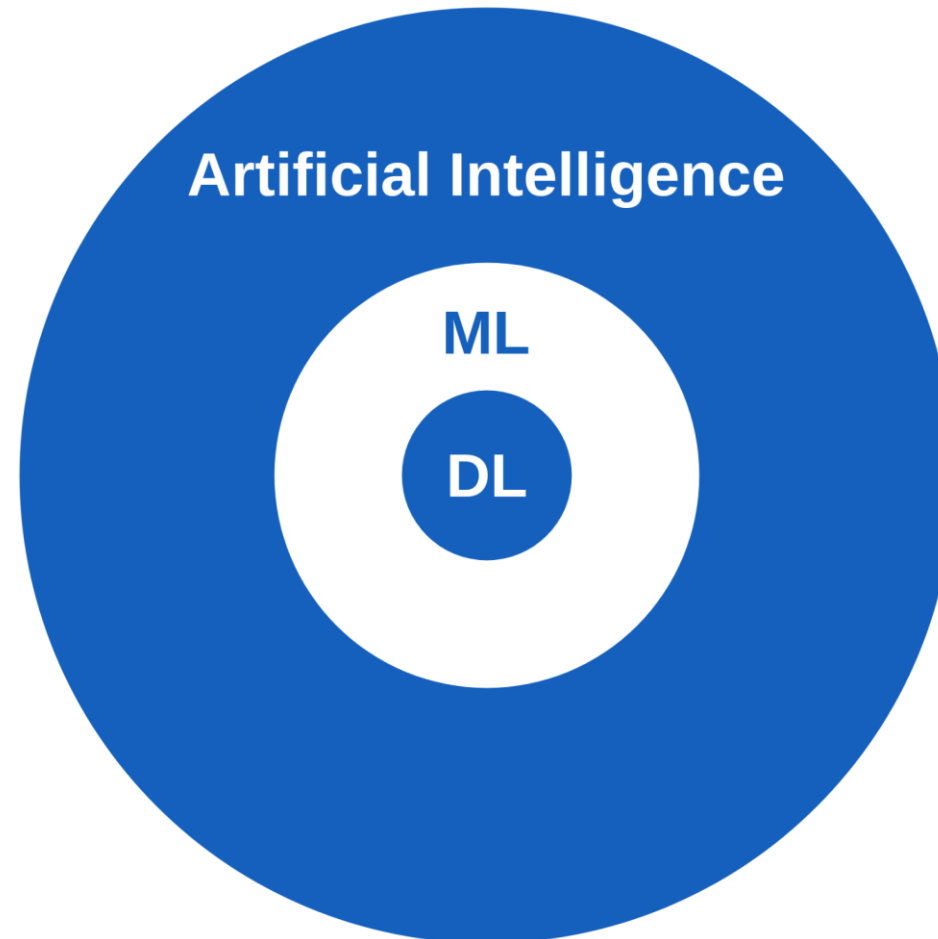


bit.ly/2I6sSAN

Reality



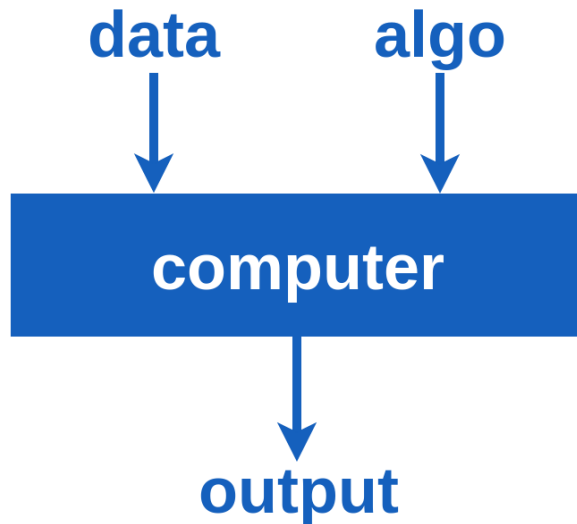
AI vs ML



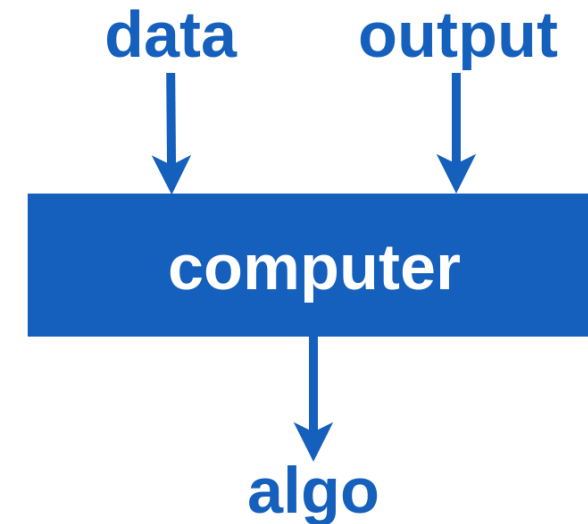
TP vs ML



Traditional Programming



Machine Learning





Problems/tasks

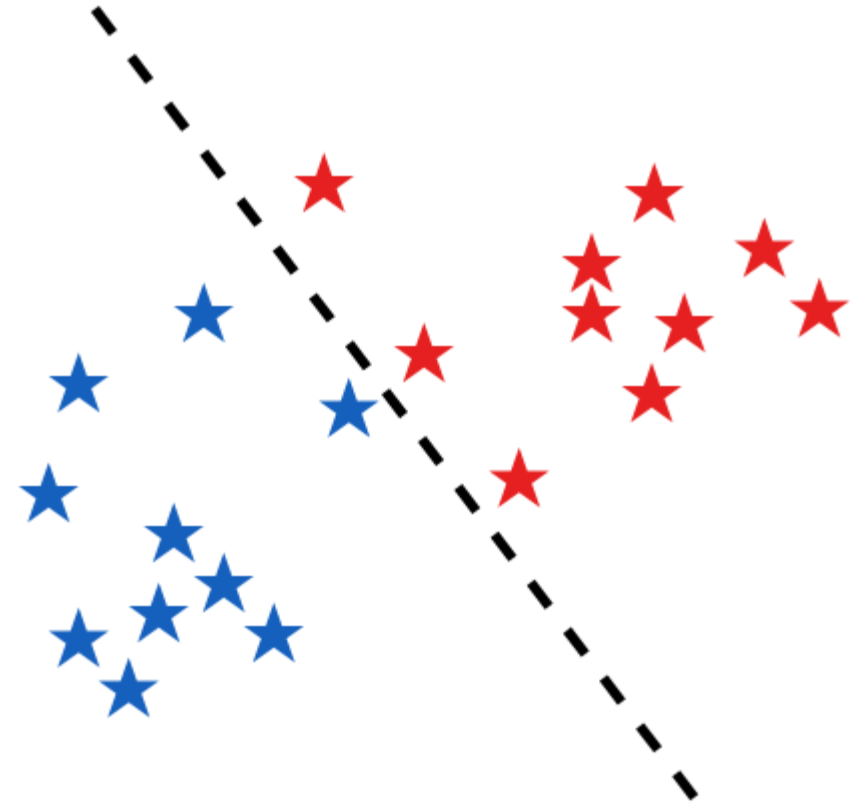
Classification



Prediction value as category like yes/no, positive/negative, etc.

Example:

- Finding named entities in text: person, organization, location, etc.
- Predicting which team will win the Champions League in soccer.



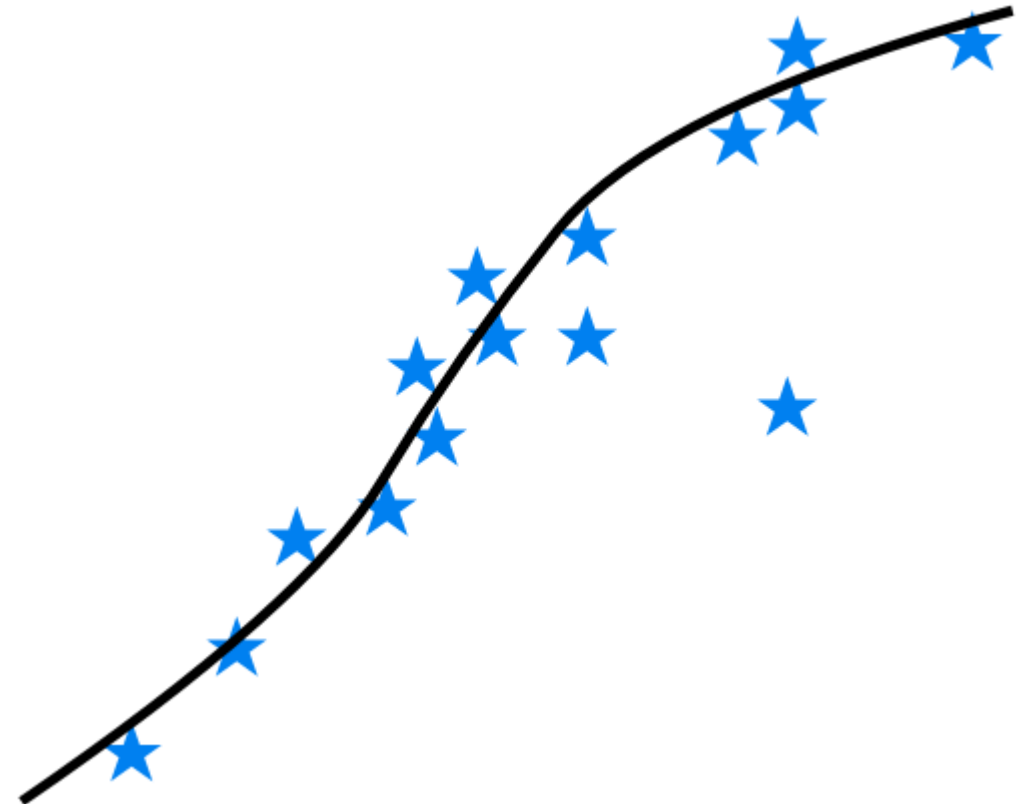
Regression



Prediction value as continuous value.

Example:

- Predicting house price.
- Predicting company's yearly revenue.



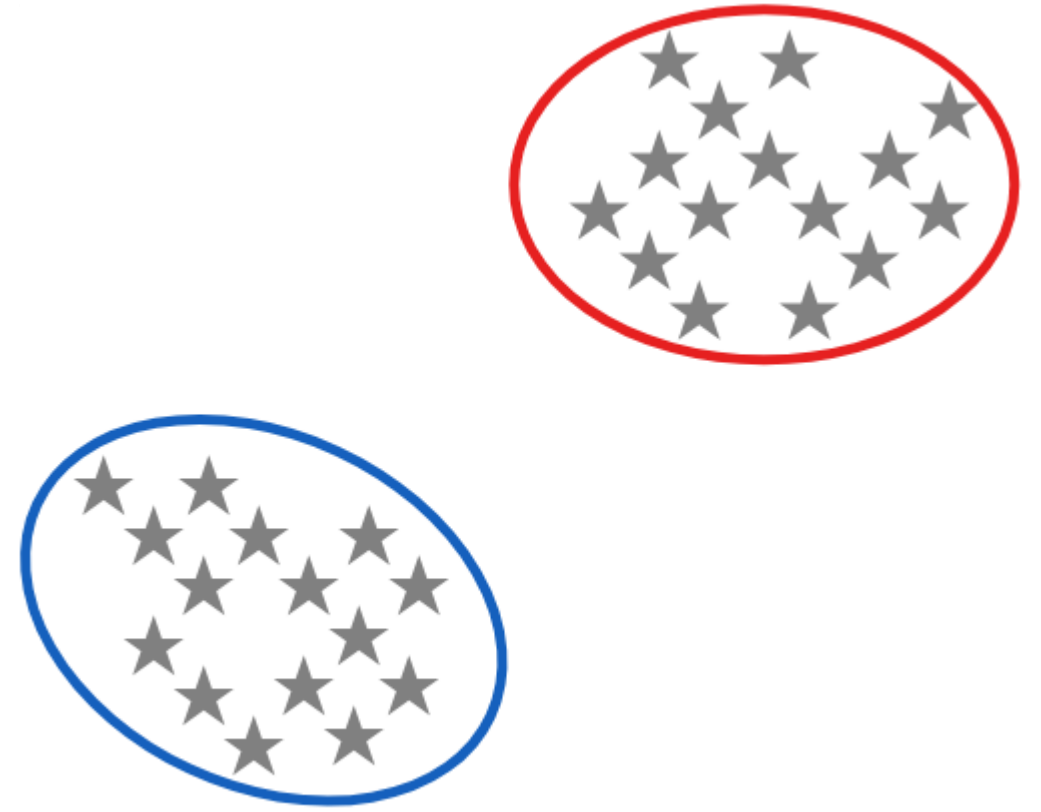
Clustering



Grouping a set of points to given number of clusters.

Example:

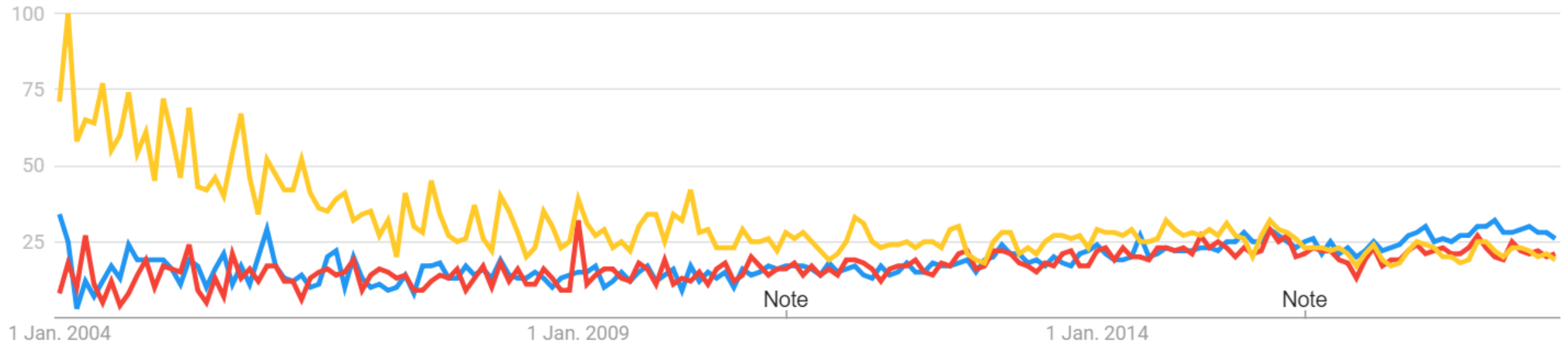
- Image clustering.
- Human activity recognition using smartphones.





Python vs R

Python vs R vs Java

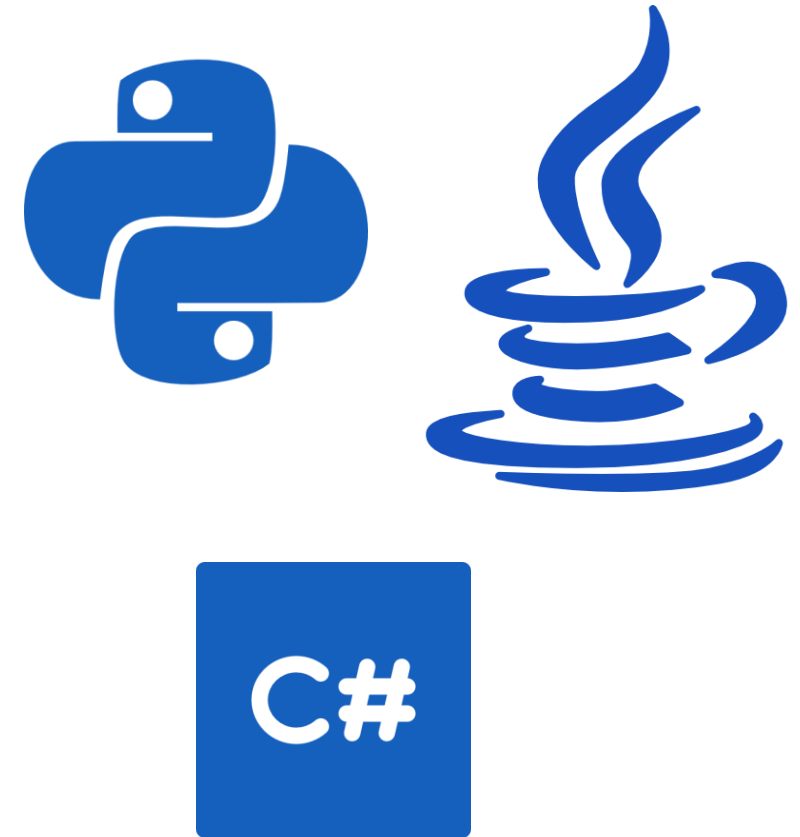


Source: bit.ly/2M6QI7I

Python. IT Salaries



	Warsaw	Katowice	Lodz	Poznan
Python	100%	95%	75%	70%
Java	100%	65%	85%	75%
.NET C#	100%	70%	75%	65%



Source: bit.ly/2tgRxhj



Let's start

Classification in Python. Predict Diabetes



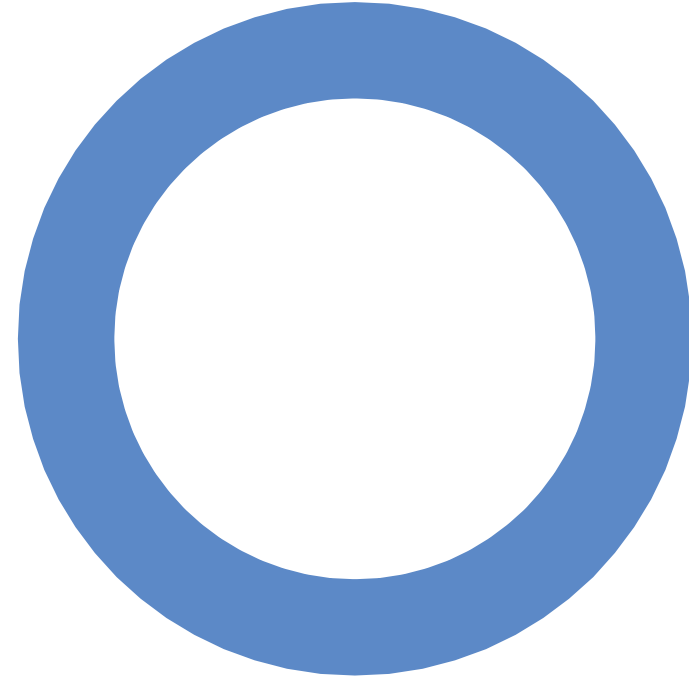
Classes: *TRUE* -- diabetic, *FALSE* -- non-diabetic.

Attributes:

1. number of pregnancies,
2. plasma glucose concentration,
3. blood pressure,
4. skin fold thickness,
5. serum insulin,
6. body mass index,
7. diabetes pedigree function,
8. age,
9. class.

Data and code source: bit.ly/2Mbo53w

Image source: bit.ly/2MN4Myi



Classification in R. Banknote Auth



Classes: *1* -- genuine, *0* -- counterfeit (fake).

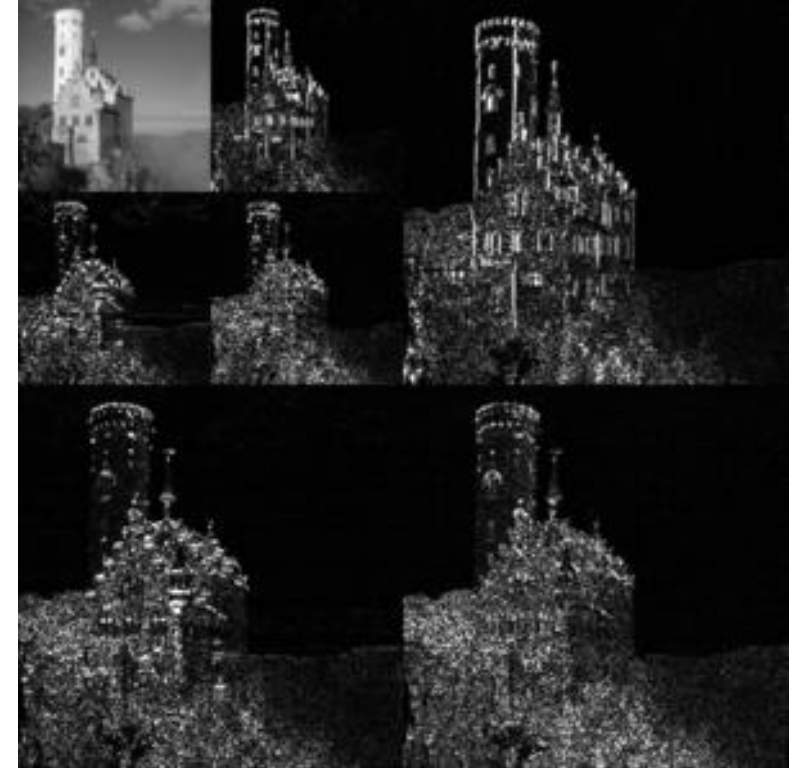
Attributes:

1. [variance](#) of Wavelet Transformed image,
2. [skewness](#) of Wavelet Transformed image,
3. [kurtosis](#) of Wavelet Transformed image,
4. entropy of image,
5. class.

Data source: bit.ly/2JpLsp3

Code source: bit.ly/2K2wcP3

Image source: bit.ly/2kWjwyw





Algorithms



Source: bit.ly/2HuB6lI



Tips and Tricks

by badly-behaved Data Scientist

1/3 Data



Iceberg



The 90/10 Rule

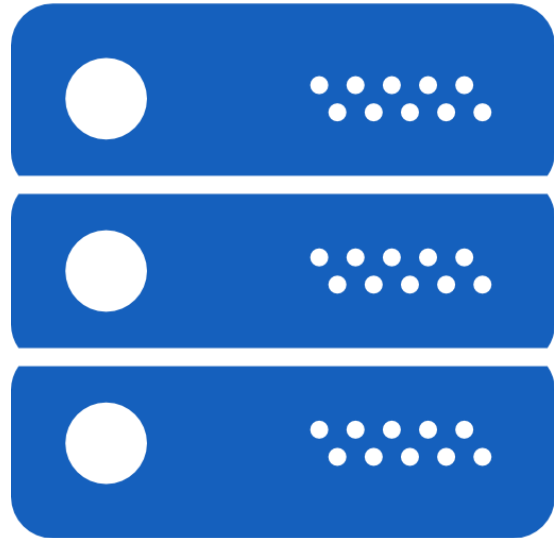
10%
analysis

90% data

2/3 Equipment



Server



Moore's Law

**# of
transistors
x2 every
two years**

3/3 Polish language



Poland



Language Difficulty Ranking

**Polish is the
hardest
language!**



bit.ly/AIKATO



About Capgemini

A global leader in consulting, technology services and digital transformation, Capgemini is at the forefront of innovation to address the entire breadth of clients' opportunities in the evolving world of cloud, digital and platforms. Building on its strong 50-year heritage and deep industry-specific expertise, Capgemini enables organizations to realize their business ambitions through an array of services from strategy to operations. Capgemini is driven by the conviction that the business value of technology comes from and through people. It is a multicultural company of 200,000 team members in over 40 countries. The Group reported 2017 global revenues of EUR 12.8 billion.

Learn more about us at

www.capgemini.com



People matter, results count.

This presentation contains information that may be privileged or confidential and is the property of the Capgemini Group.
Copyright © 2018 Capgemini. All rights reserved.

Ruslan Korniiichuk

Senior ITS Consultant

Capgemini

ruslan.korniichuk@capgemini.com

+48 791 065 877

Nikolay Pawlenko

Data Scientist

Bank Zachodni WBK

www.pawlenko.com

+48 787 398 477