



Machine Learning

Krakow, 1st of September 2018

Expectation vs Reality

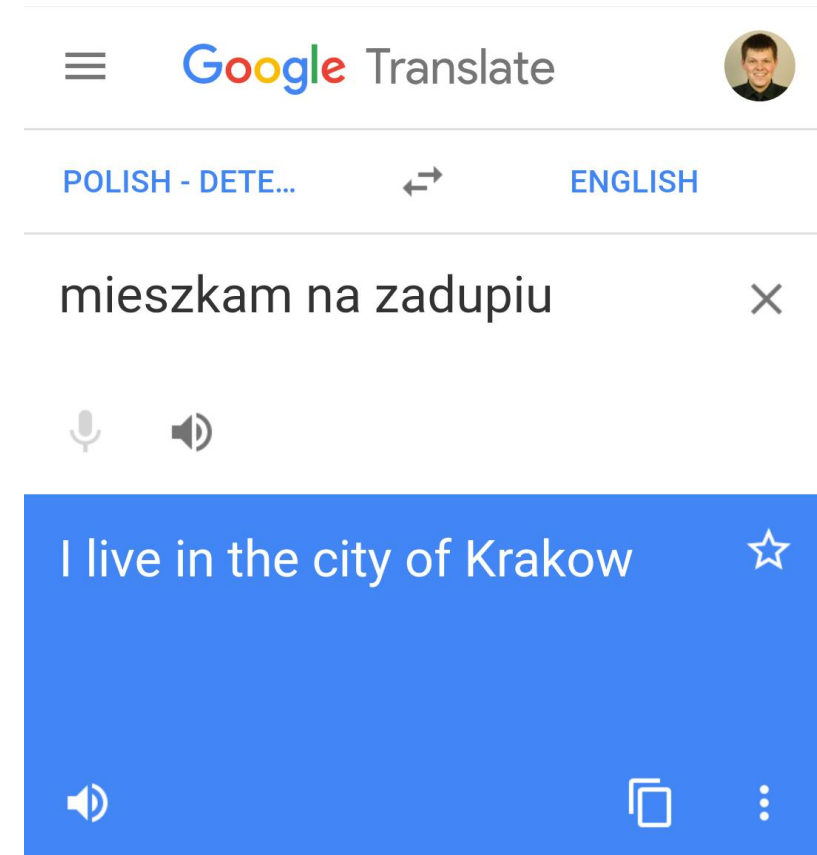


Expectation

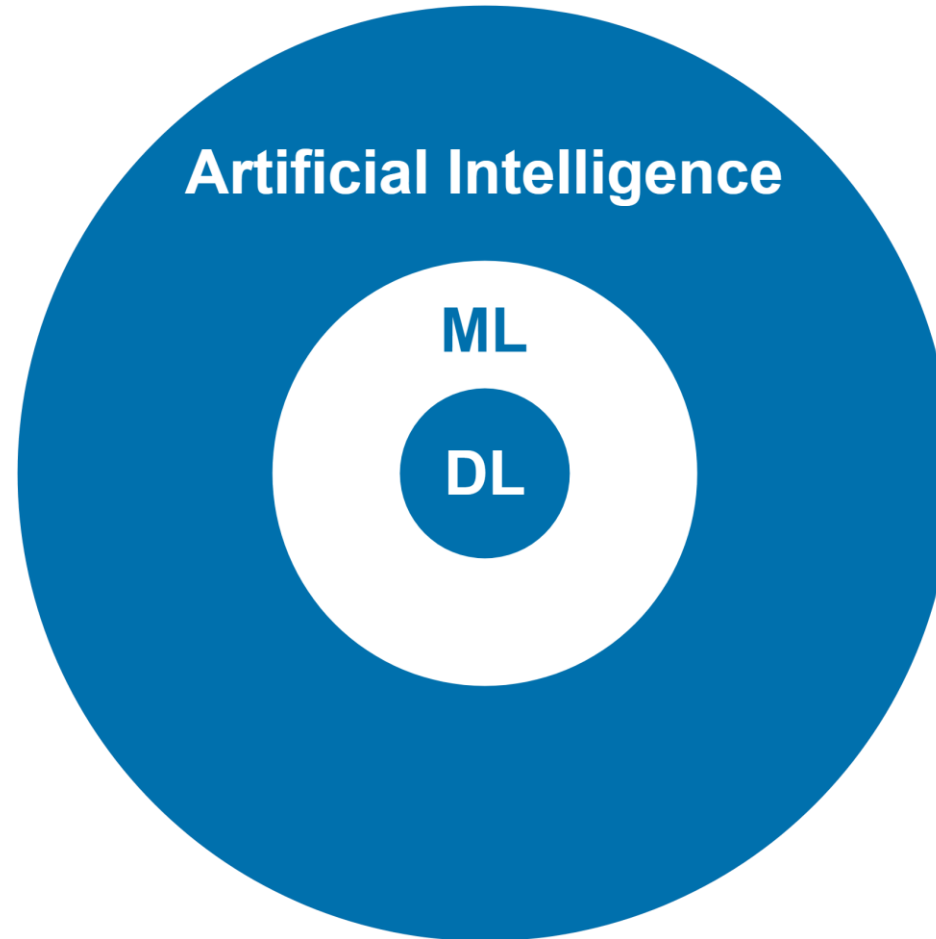


bit.ly/2I6sSAN

Reality



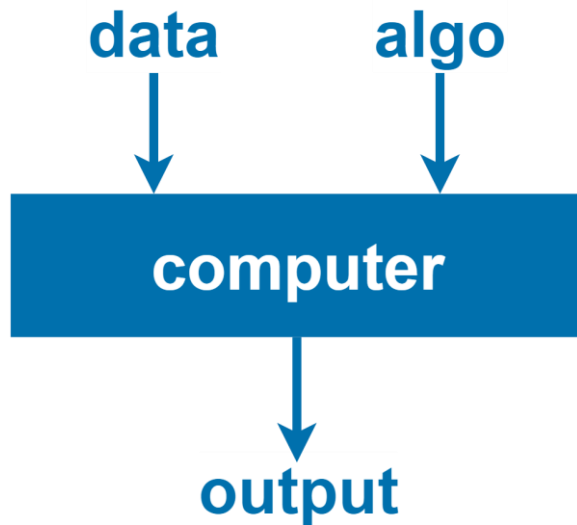
Artificial intelligence vs Machine Learning



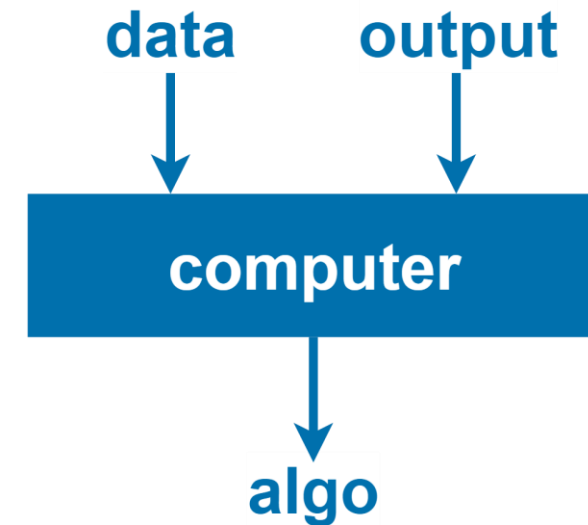
Traditional Programming vs Machine Learning



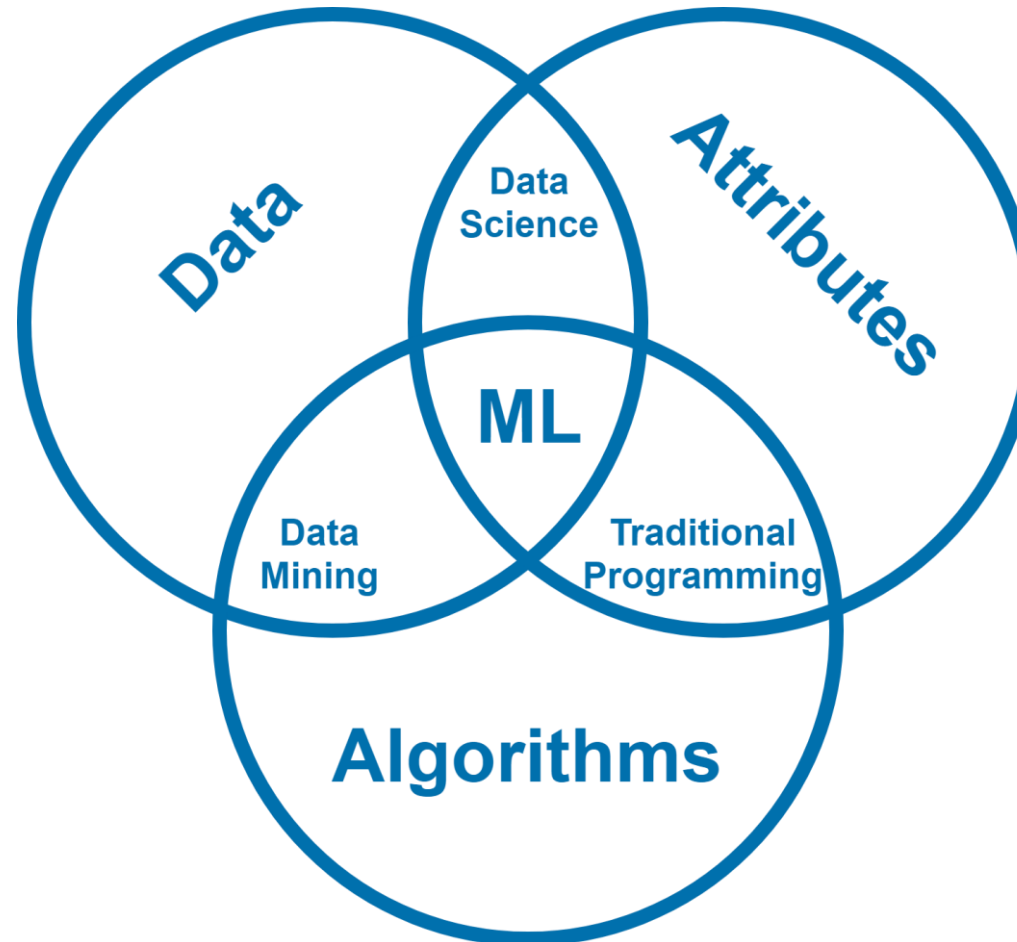
Traditional Programming



Machine Learning



Data Science vs Data Mining



ML tasks

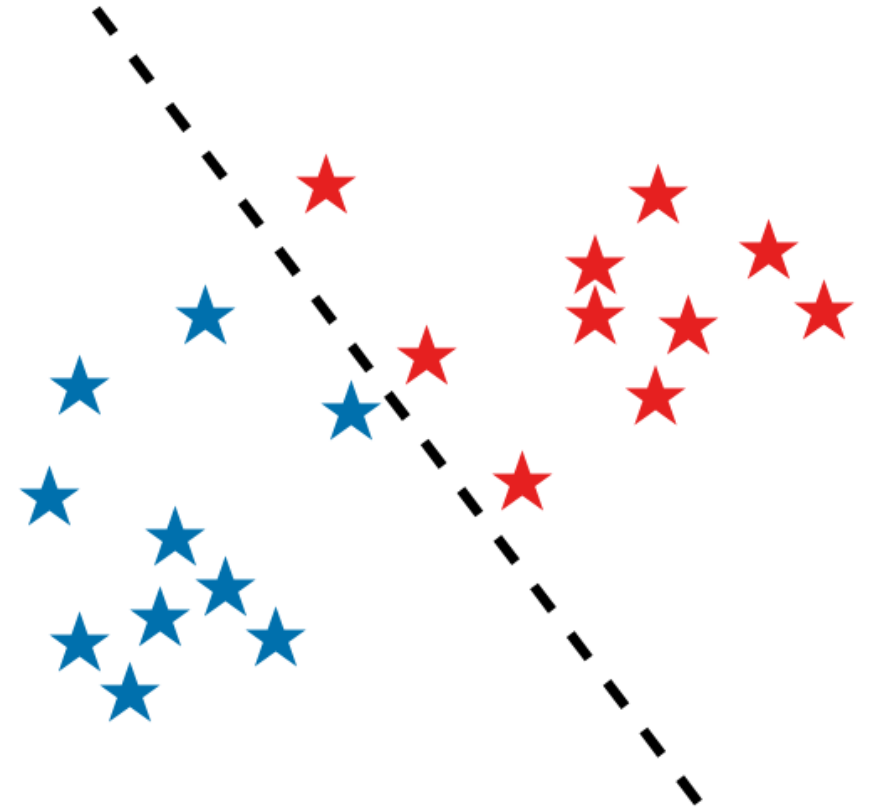
Classification



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

Examples:

- Email filtering.
- Credit scoring.
- Finding named entities in text: person, organization, location, etc.
- Language identification.
- Prediction of diabetes.



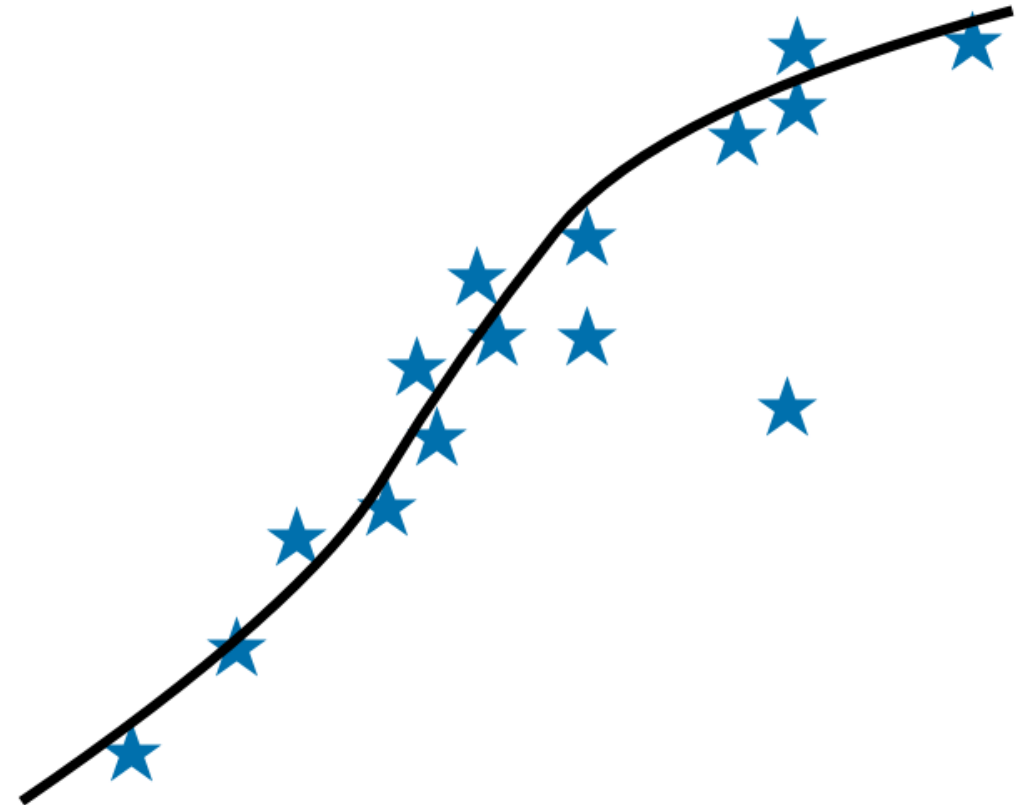
Regression



Prediction value as continuous value.

Examples:

- Company's yearly revenue prediction.
- Foreign exchange market prediction.
- Job salary prediction.
- Predicting house and car price.
- Stock market prediction.



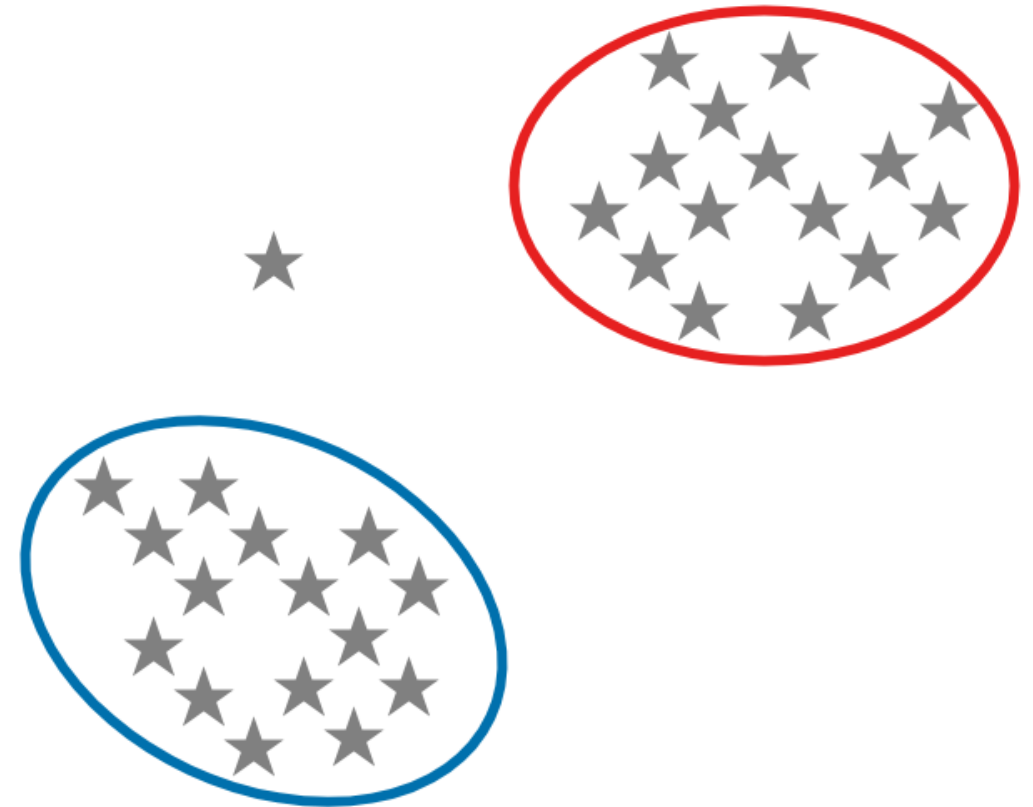
Clustering



Grouping a set of points to clusters.

Examples:

- Human activity recognition using smartphones.
- Image clustering.
- Image compression.
- Recommender system.
- One-class classification.



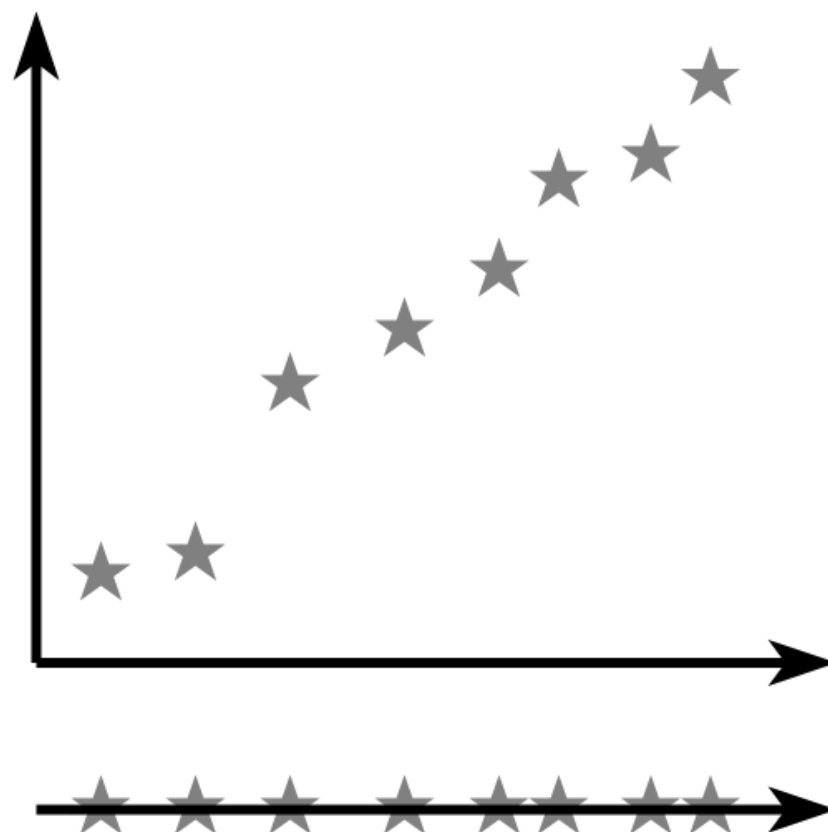


Dimensionality reduction

Reducing the number of attributes.

Examples:

- Data visualization.
- Fake image detection.
- Recommender system.
- Risk management.
- Topic Modelling.



Test Part 1



- Anomaly detection.
- Cancer prediction.
- Champions League winner prediction.
- Character recognition.
- E-Commerce business forecasting.
- Employee attrition prediction.
- Employee's unplanned absence prediction.
- Fraudulent bank transaction identification.

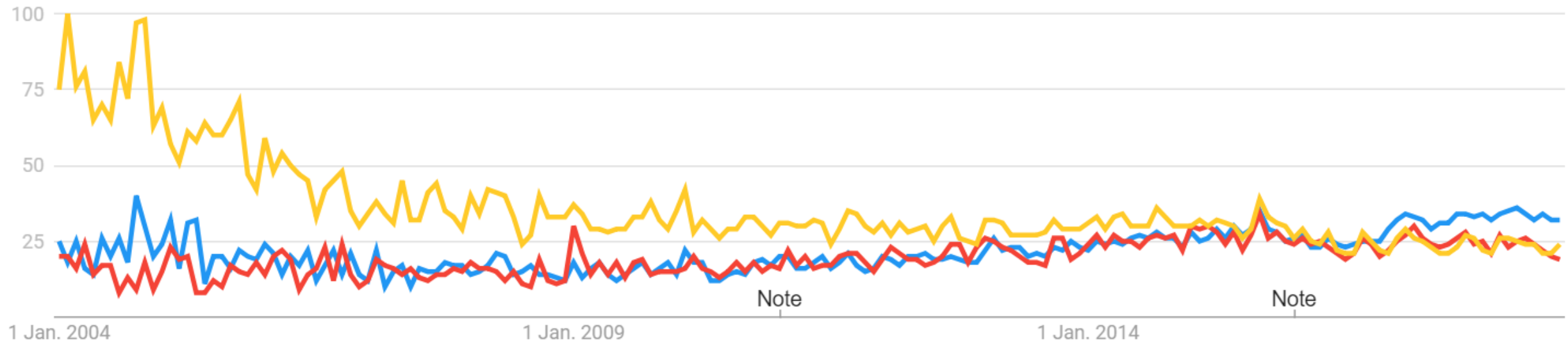
Test Part 2



- IT system availability prediction.
- Joiner prediction in recruitment.
- Map points combining.
- Market basket analysis.
- Market segmentation (types of customers, loyalty).
- Project overrun prediction.
- Sentiment analysis.
- Telecom industry churners prediction.

Python vs R

Python vs R vs Java

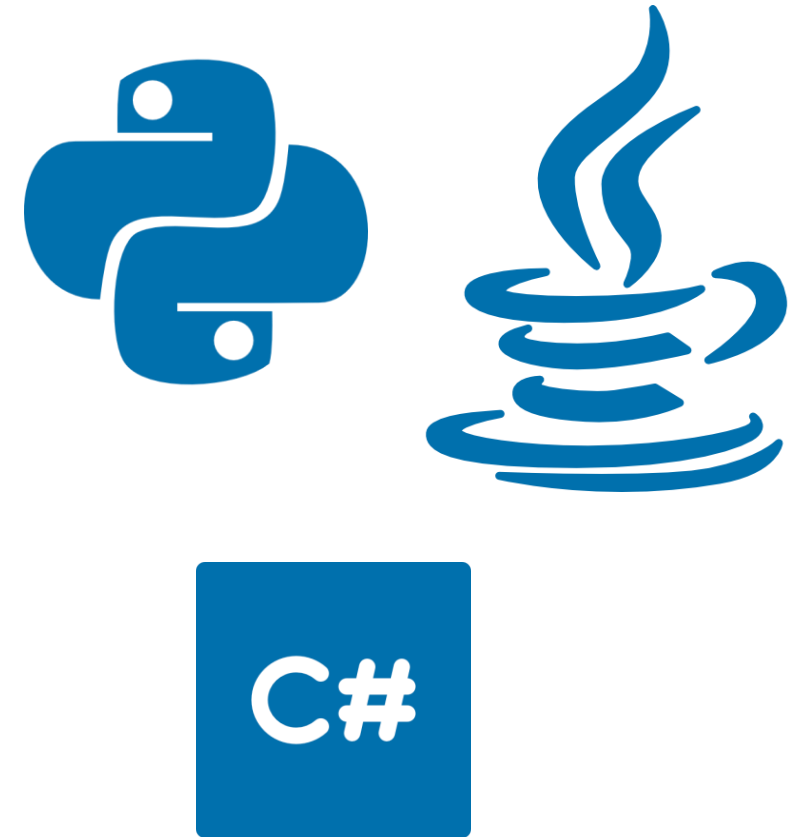


Source: bit.ly/2M6QI7I

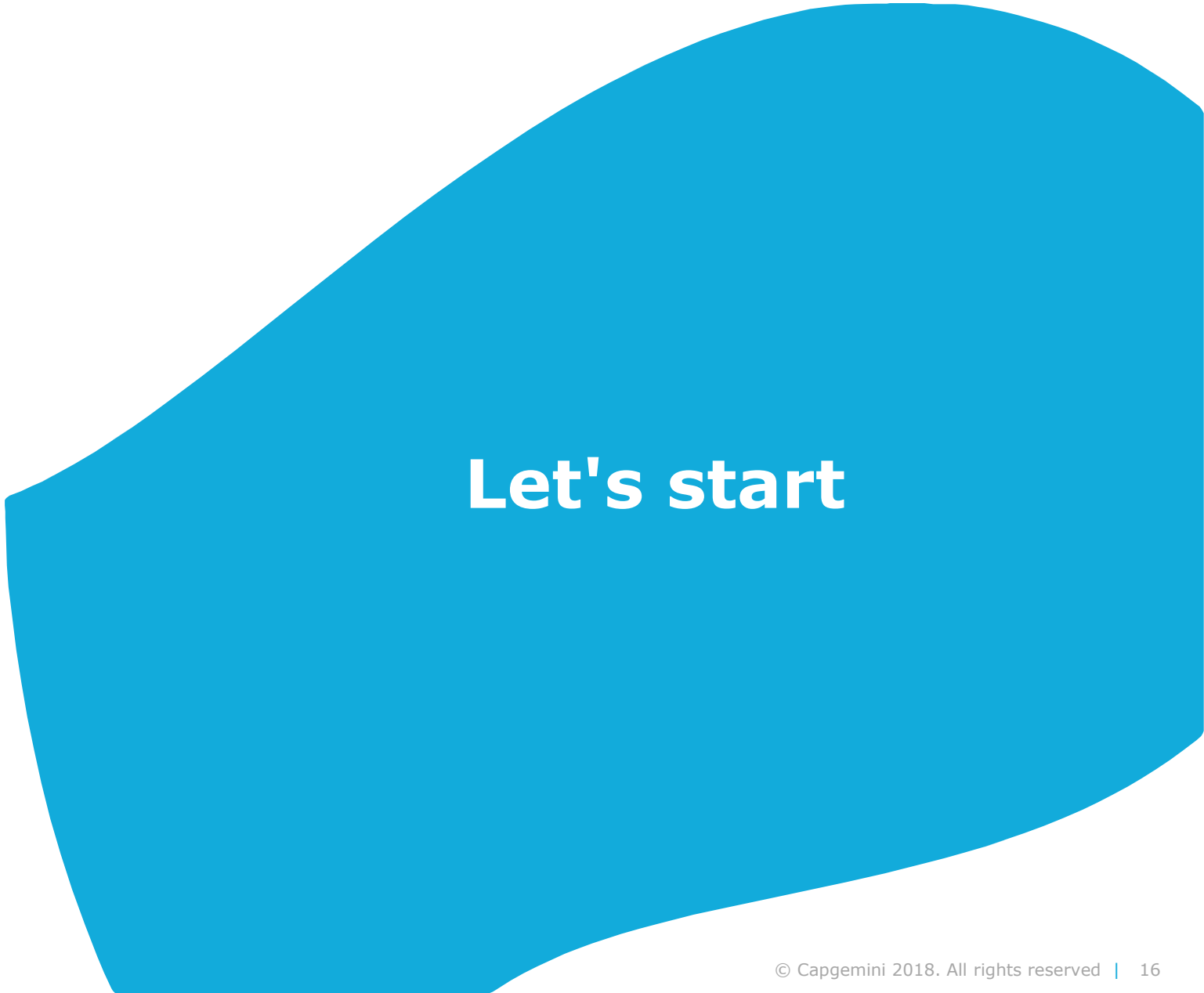
Python. IT Salaries



	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 R. Banknote Auth

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

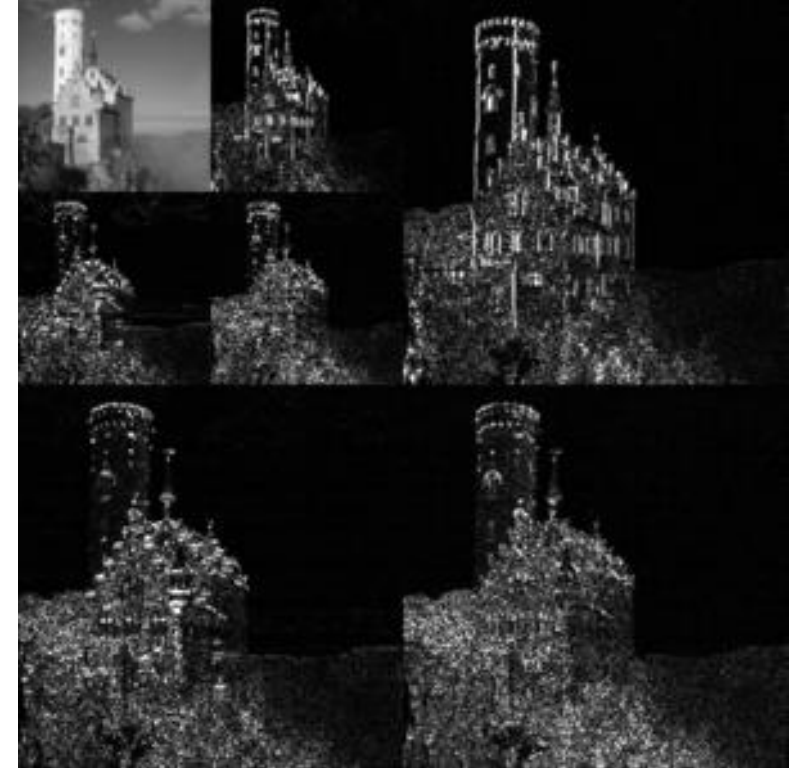
Attributes:

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





Classification in Python. Predict Diabetes

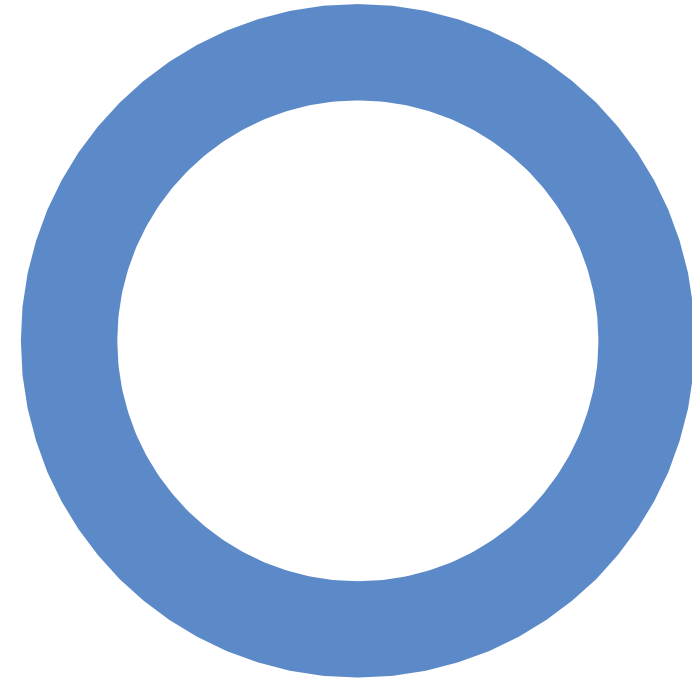
Classes: *TRUE* -- diabetic, *FALSE* -- non-diabetic.

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. skin,
10. class.

Data and code source: bit.ly/2Mbo53w

Image source: bit.ly/2MN4Myi



Algorithms



Source: bit.ly/2HuB6lI



Tips and Tricks

by badly-behaved Data Scientist

1/4 Data



Iceberg



The 90/10 Rule

10%
analysis

90% data

2/4 Magic numbers



Magic



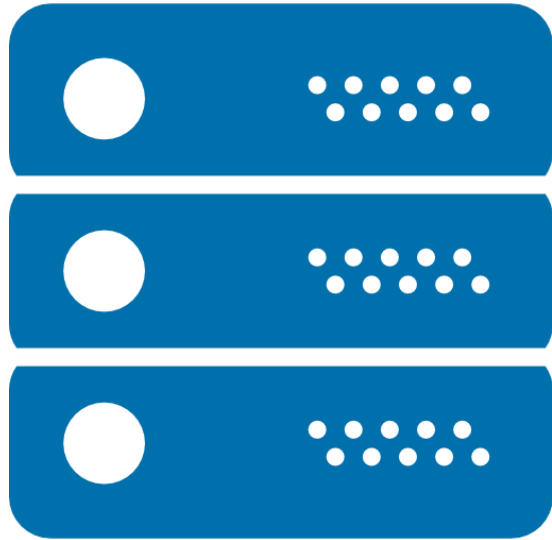
Is Machine Learning Magic?

70% / 30%
train / test
split

3/4 Equipment



Server



Moore's Law

**# of
transistors
x2 every
two years**

4/4 Polish language



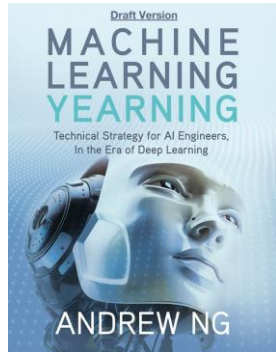
Poland



Language Difficulty Ranking

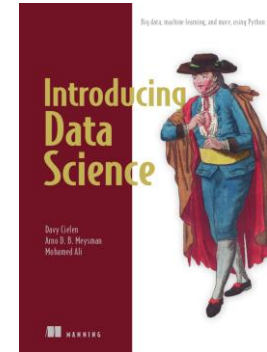
**Polish is the
hardest
language!**

Books



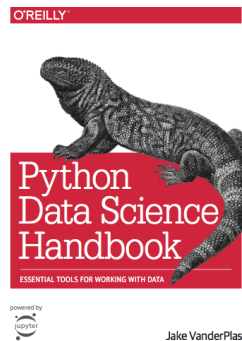
Machine Learning Yearning by Andrew Ng

- [in English](#)
- [in Russian](#)



Introducing Data Science by Davy Cielen, Arno Meysman, and Mohamed Ali

- in English
- in Russian



Python Data Science Handbook by Jake VanderPlas

- [in English](#)
- in Russian



bit.ly/CPKRRK

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

Santander

www.pawlenko.com

+48 787 398 477