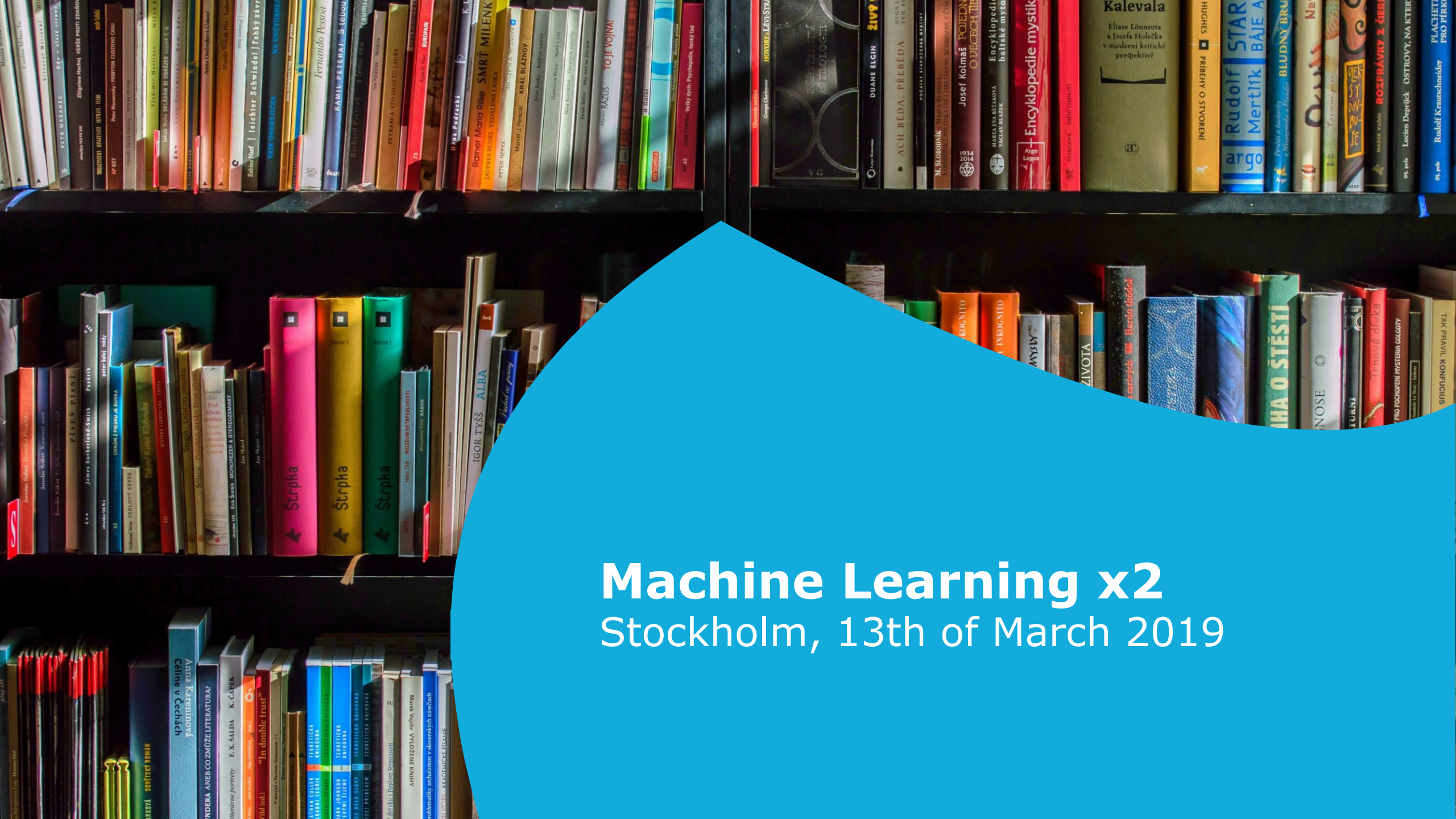
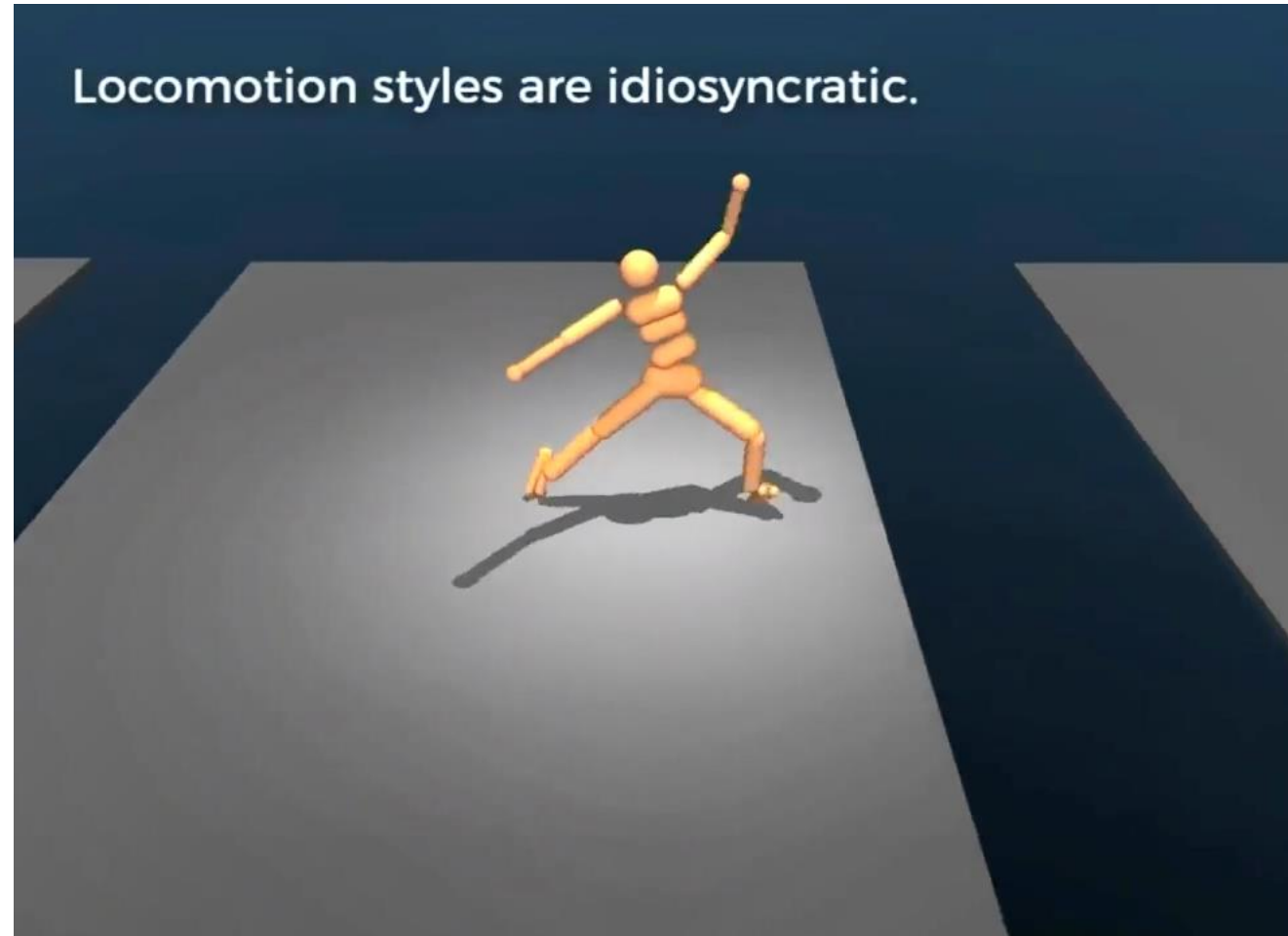




Machine Learning x2

Stockholm, 13th of March 2019

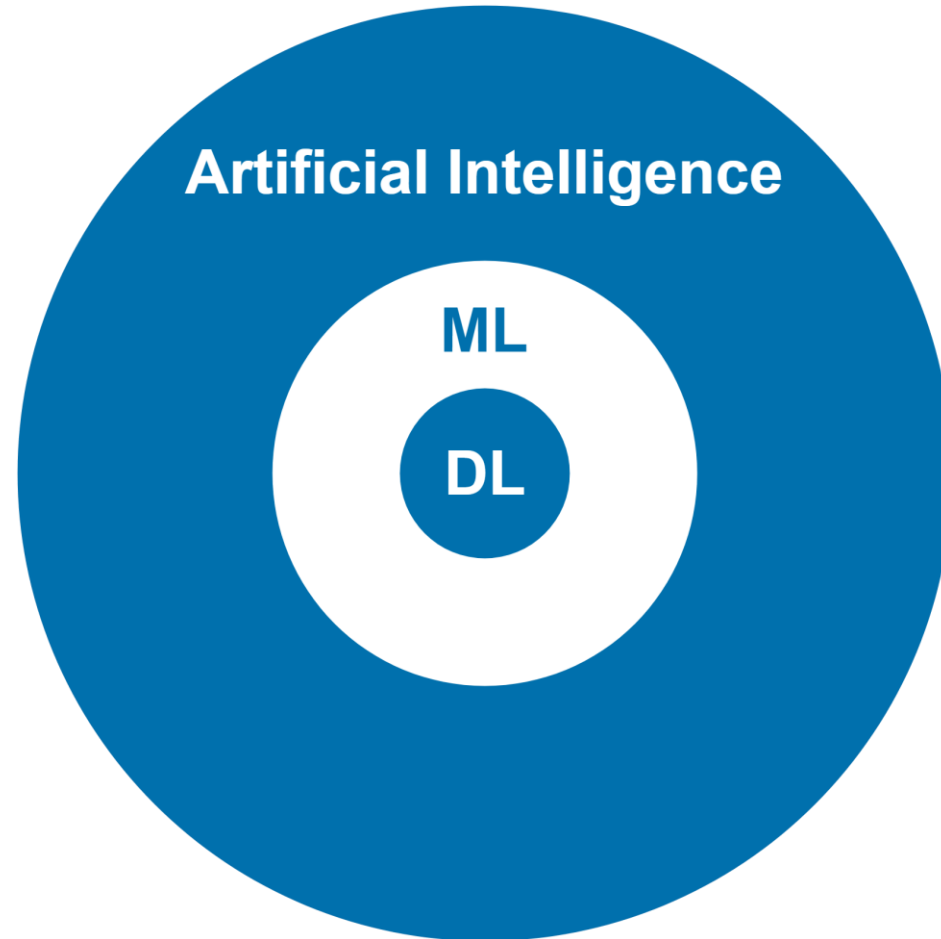
Expectation vs Reality

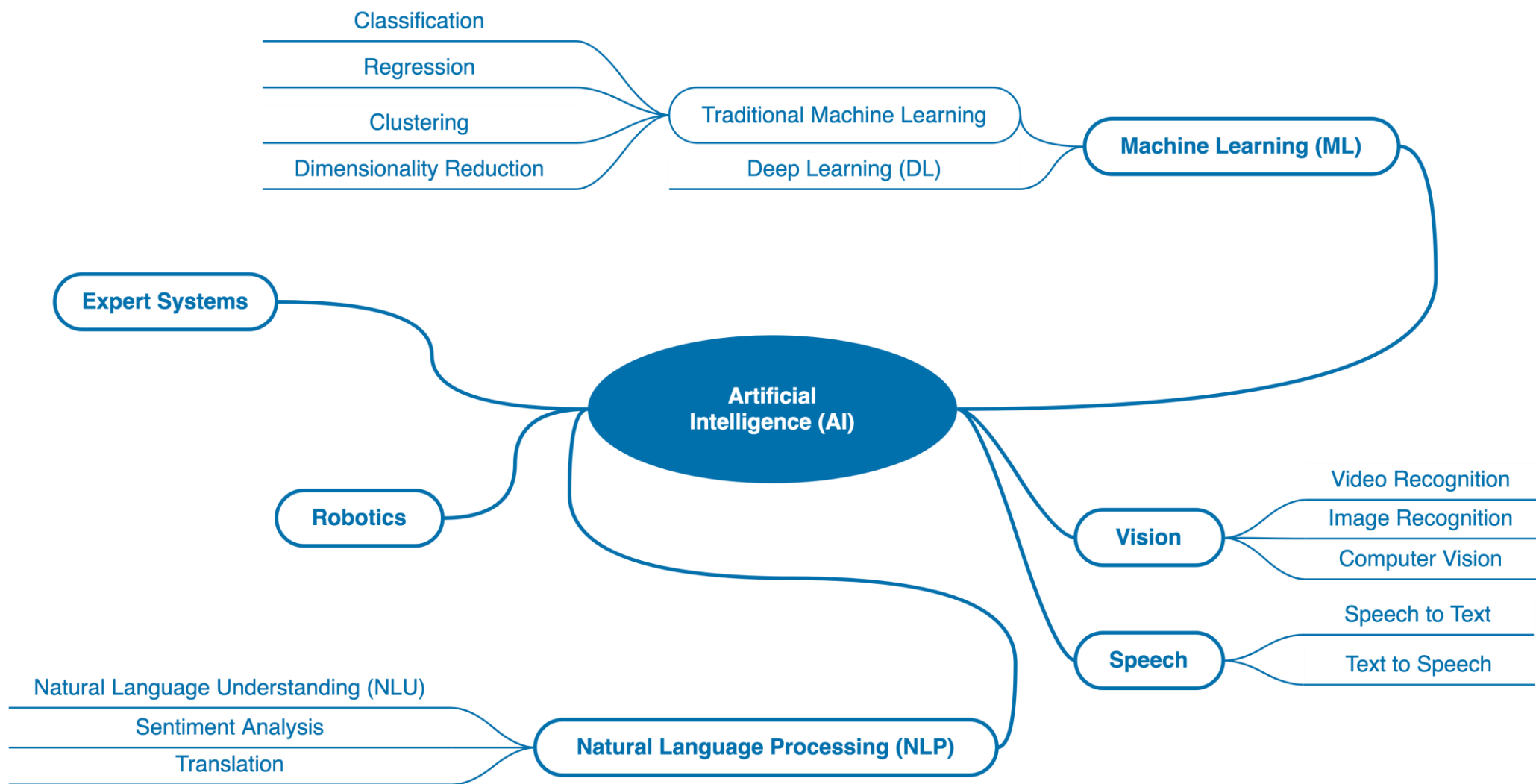


Source: bit.ly/2P0nqTi

Terminology

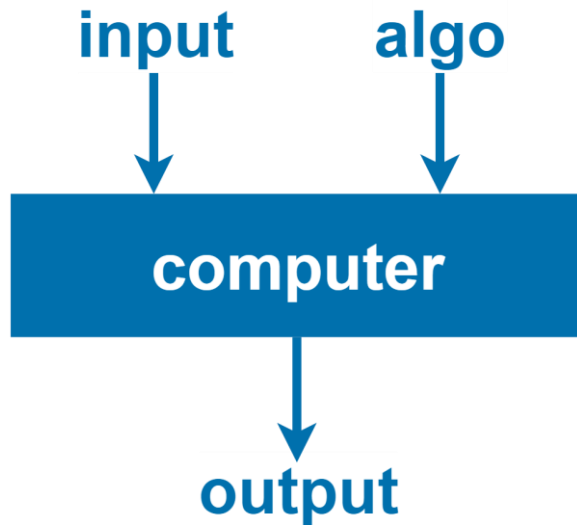
Artificial Intelligence vs Machine Learning



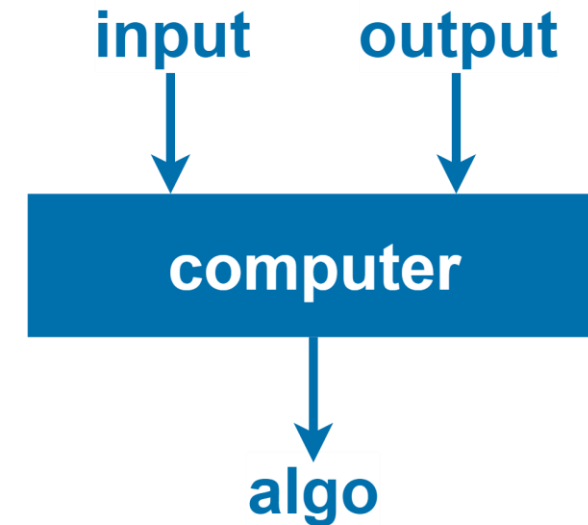


Traditional Programming vs Machine Learning

Traditional Programming



Machine Learning



Traditional Programming vs Machine Learning

Traditional Programming

$$1 + 1 = \square$$

$$1 + 2 = \square$$

Machine Learning

$$1 \square 1 = 2$$

$$1 \square 2 = 3$$

Traditional Programming vs Machine Learning

Traditional Programming

$$1 + 1 = \square$$

$$1 + 2 = \square$$

$$1 + 3 = 4$$

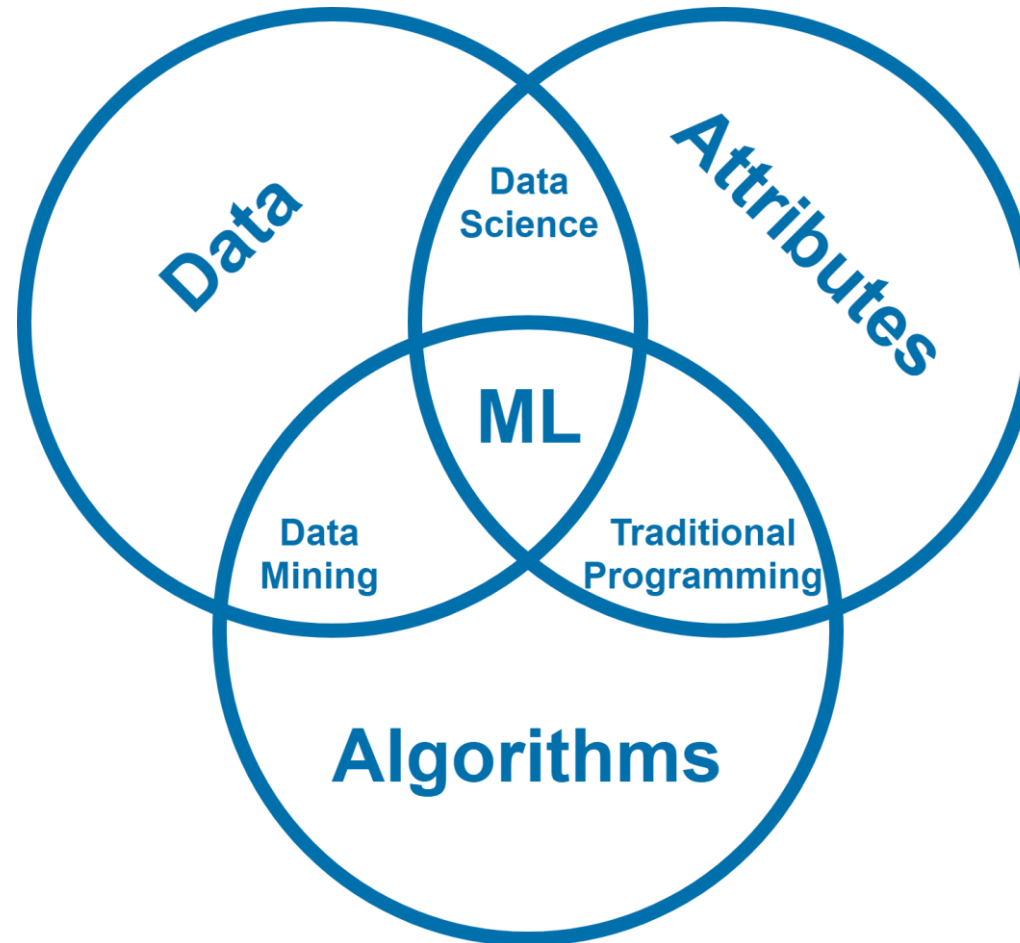
Machine Learning

$$1 \square 1 = 2$$

$$1 \square 2 = 3$$

$$1*1 + 3*1 = 4$$

Data Science vs Data Mining



Data science job roles

Part 1

Data Scientist

- Python
- R
- Pandas, Apache Spark
- Plotly
- Scikit-learn
- TensorFlow, Keras
- Jupyter



Artificial Intelligence Engineer

- Python
- Cloud computing
- Scikit-learn
- TensorFlow, Keras
- Jupyter
- PowerPoint
- Draw.io



Data Architect

- Cloud computing
- PowerPoint
- Draw.io
- Excel



Data science job roles

Part 2

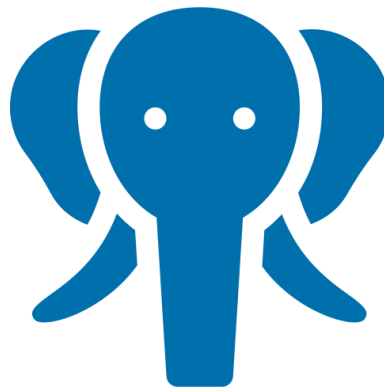
Data Engineer

- Python
- Pandas, Apache Spark
- Cloud computing
- Linux
- Hadoop
- Git



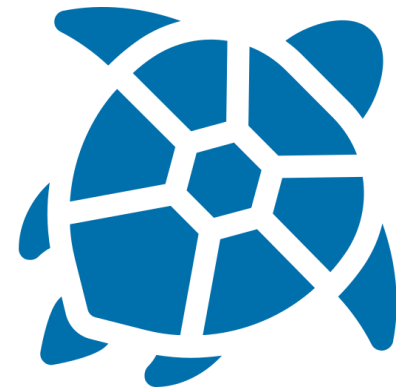
Big Data Specialist

- Java
- Hadoop
- Linux
- Git



Data Analyst

- BI tools
- SQL
- Excel
- PowerPoint





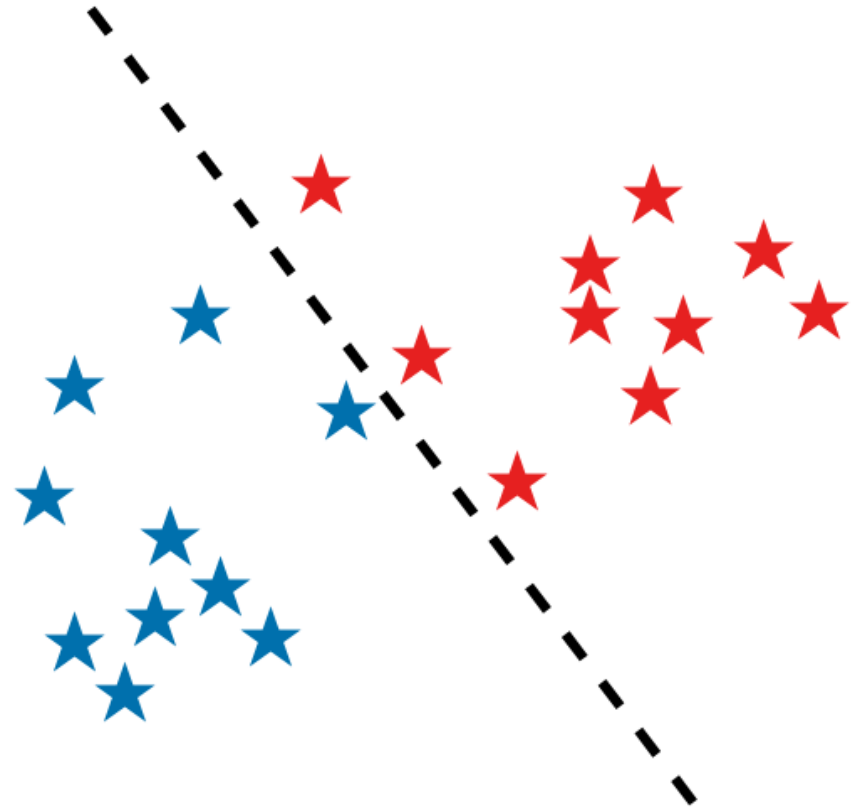
ML tasks

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.



Credit scoring

Classes: *1* – grant credit, *0* – not grant credit.

Attributes:

1. age
2. civil status
3. credit limit
4. education
5. gender
6. income
7. job title
8. work experience

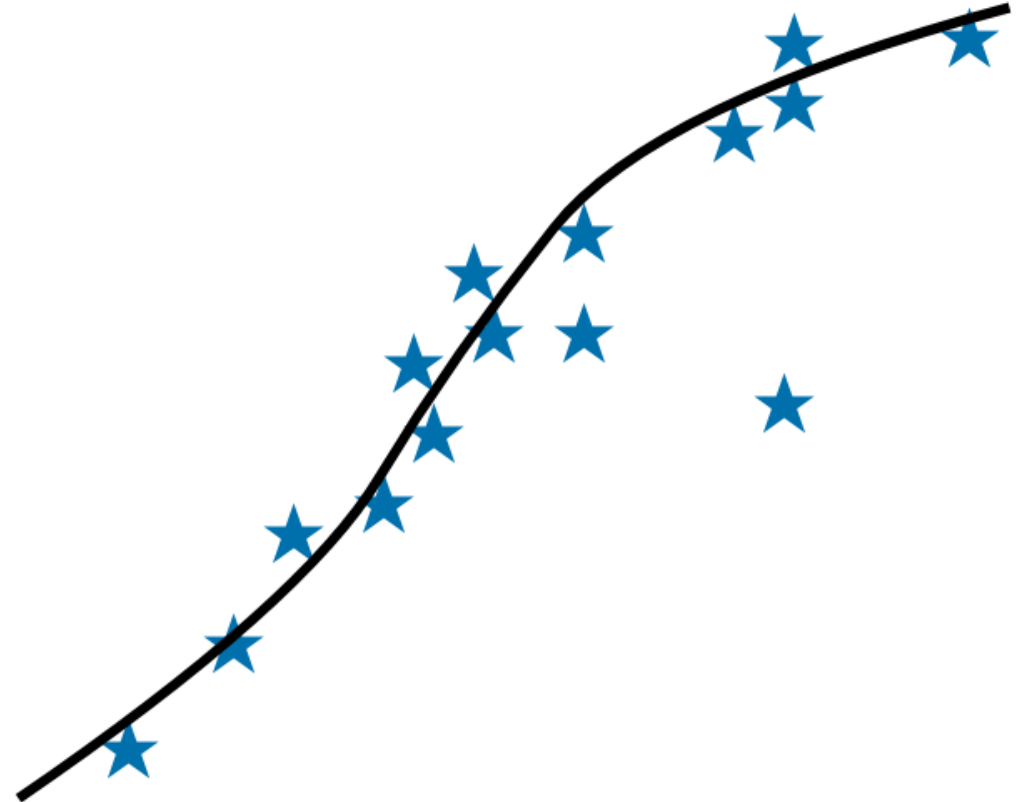


Regression

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.



Stock market prediction

Output: price of share.

Attributes:

1. close price
2. day
3. high price
4. low price
5. next close price
6. next open price
7. open price
8. previous volume
9. quarter
10. volume

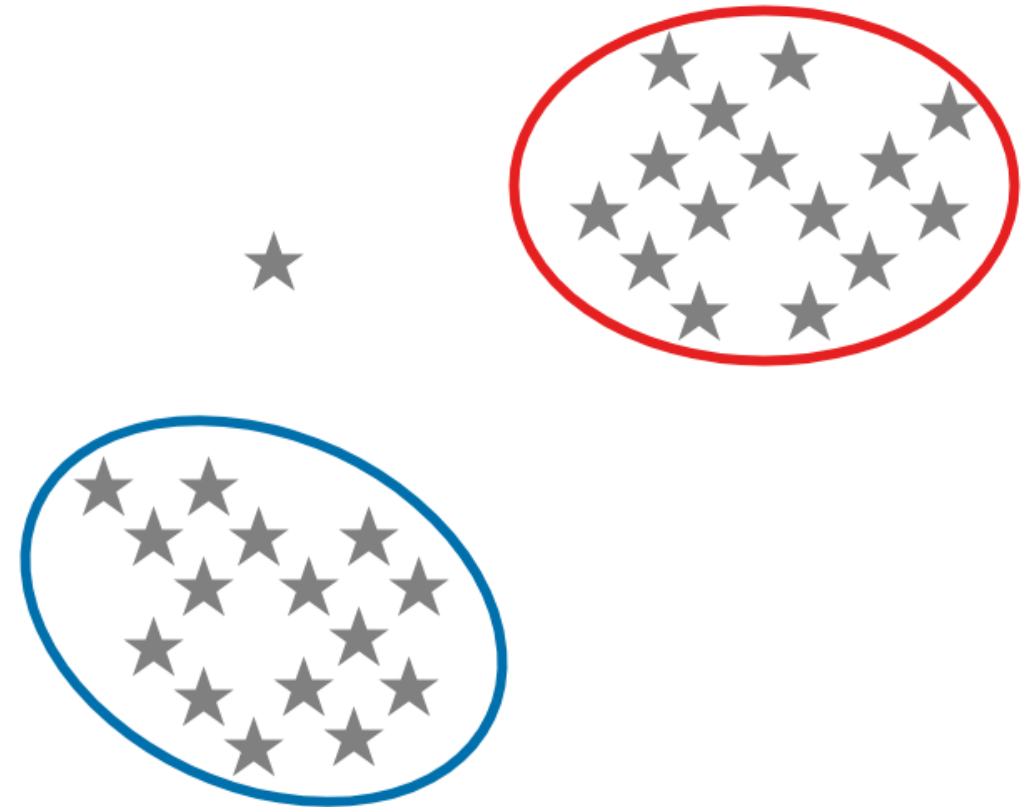


Clustering

Grouping a set of points to clusters.

Examples:

- Fraudulent bank transaction identification.
- Human activity recognition using smartphones.
- Image compression.
- Recommender system.
- One-class classification.

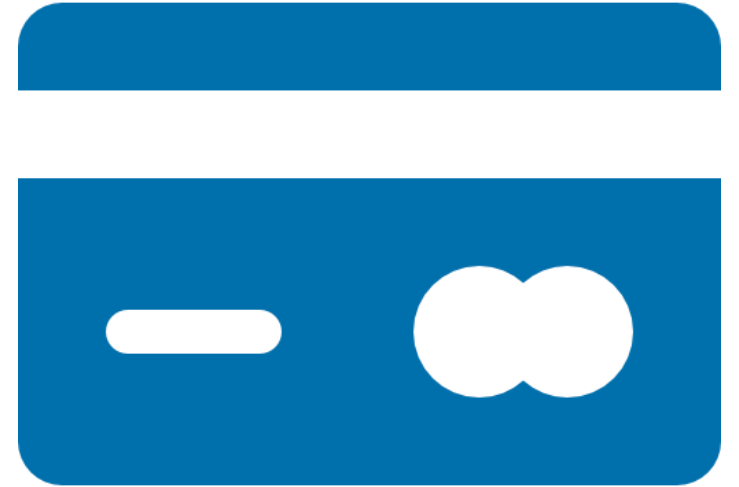


Fraudulent bank transaction identification

Output: money transfer approve or reject.

Attributes:

1. amount
2. currency in
3. currency out
4. quarter
5. receiver
6. receiver country
7. sender
8. sender country
9. weekday

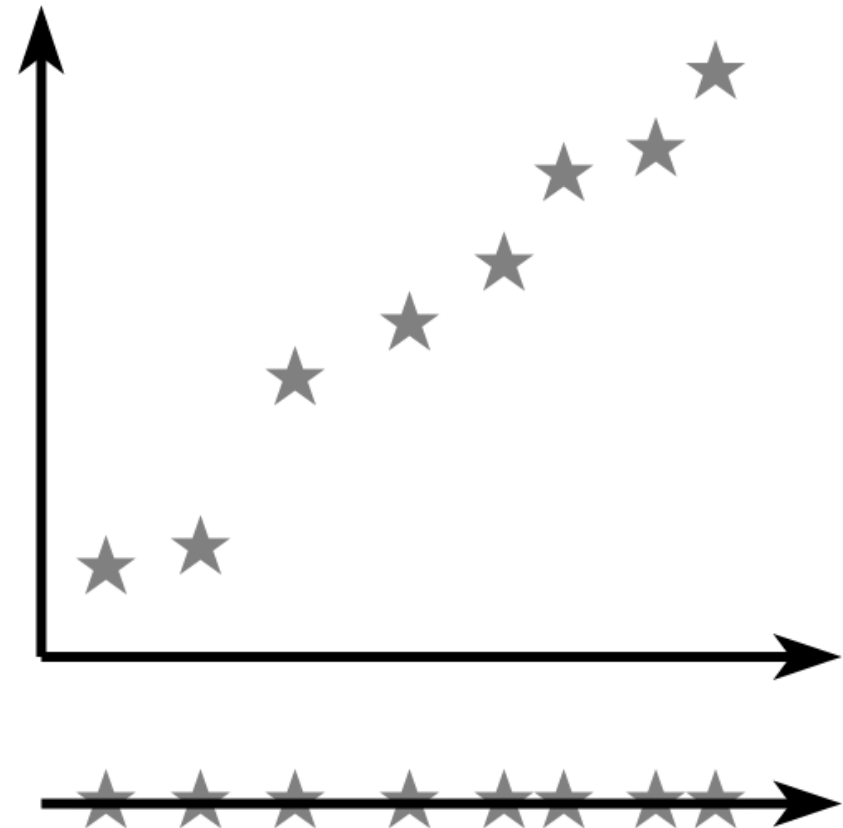


Dimensionality reduction

Reducing the number of attributes.

Examples:

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



Recommender system

	movie1	movie2	...	movieN
name1				
name2				
...				
nameN				

name1				
name2				
...				
nameN				

Recommender system

	movie1	movie2	...	movieN
name1				
name2				
...				
nameN				

	genre1	genre2	...	genreN
name1				
name2				
...				
nameN				

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.
- Image clustering.

Test

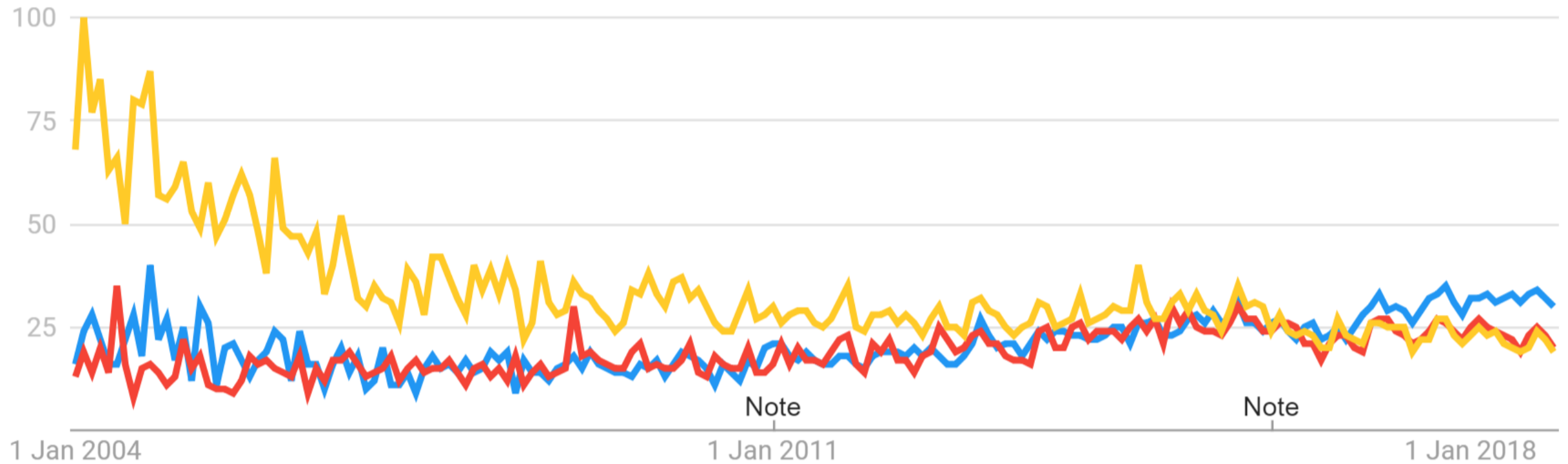
Part 2

- 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.

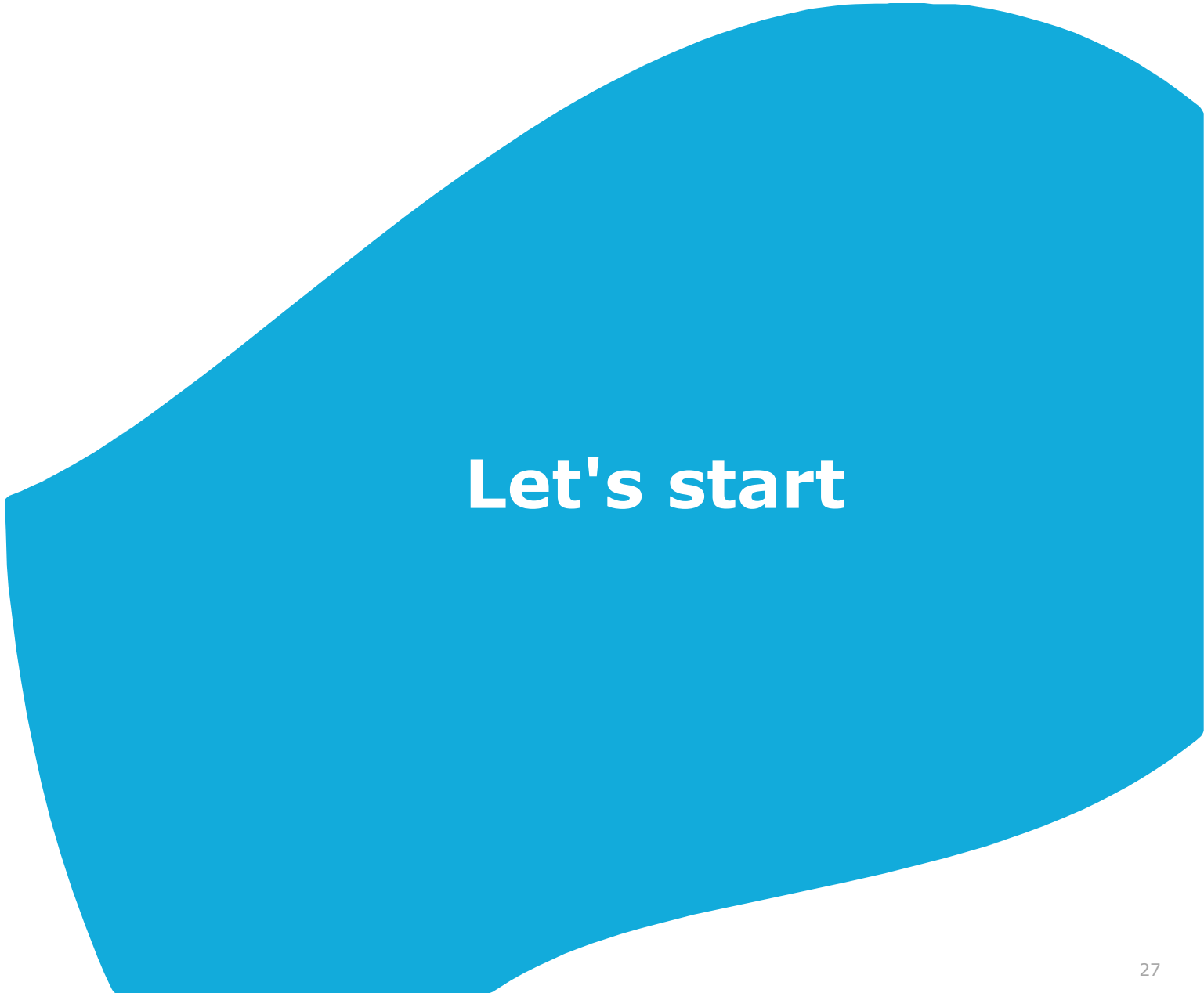
BREAK

Python vs R

Python vs R vs Java



Source: bit.ly/2M6QI7I



Let's start

Classification in Python. Predict Diabetes

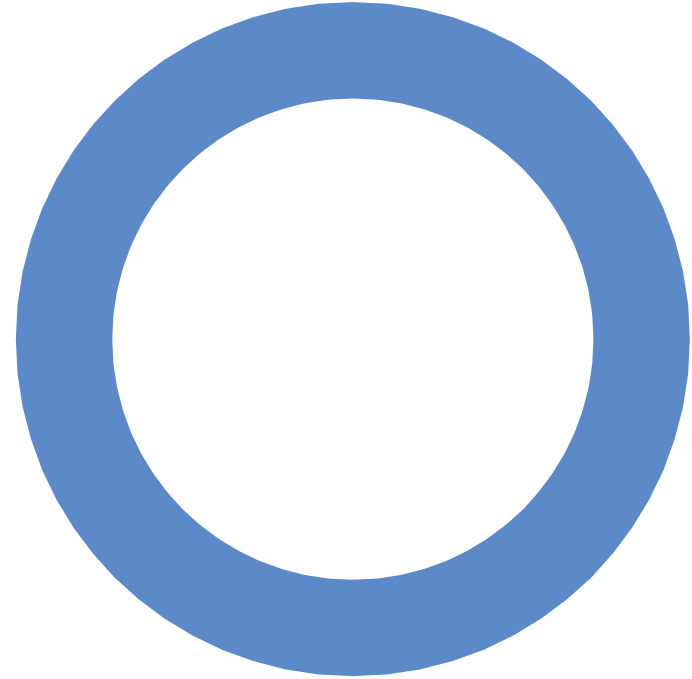
Classes: *TRUE* -- diabetic, *FALSE* -- non-diabetic.

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

Data and code source: bit.ly/2Mbo53w

Image source: bit.ly/2MN4Myi



Classification in R. Banknote Auth

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

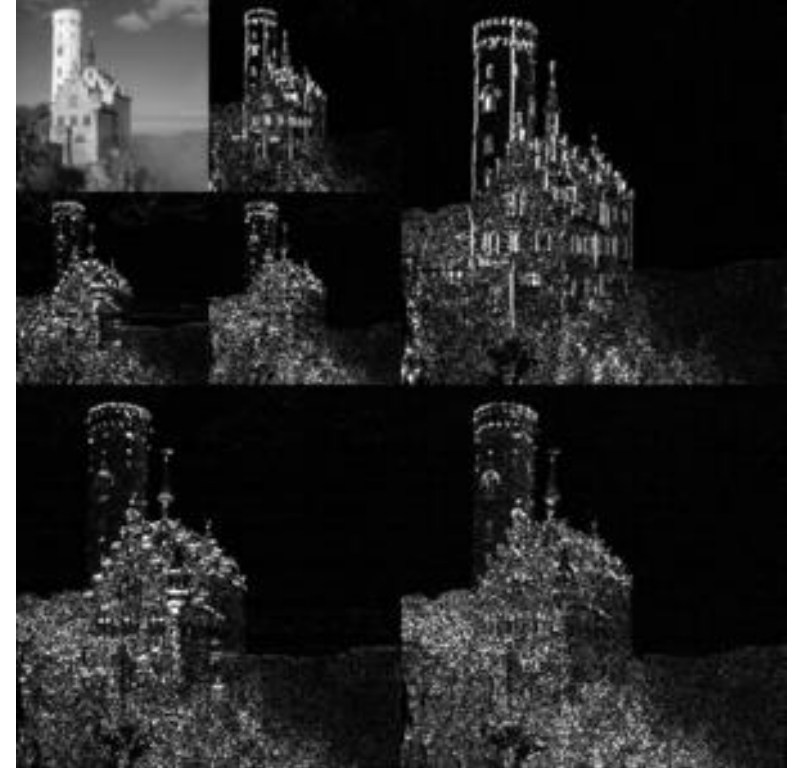
Attributes:

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

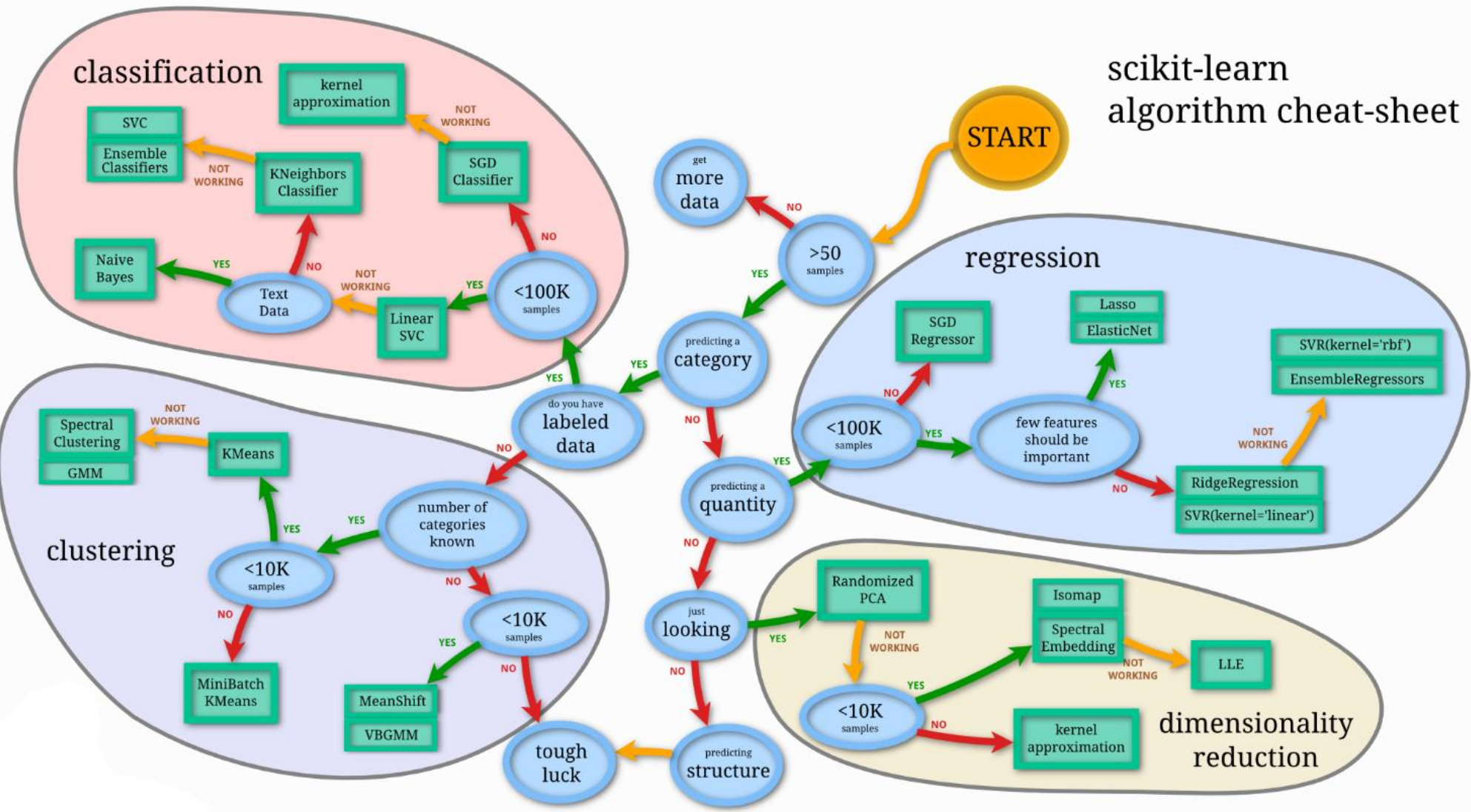
Data source: bit.ly/2JpLsp3

Code source: bit.ly/2K2wcP3

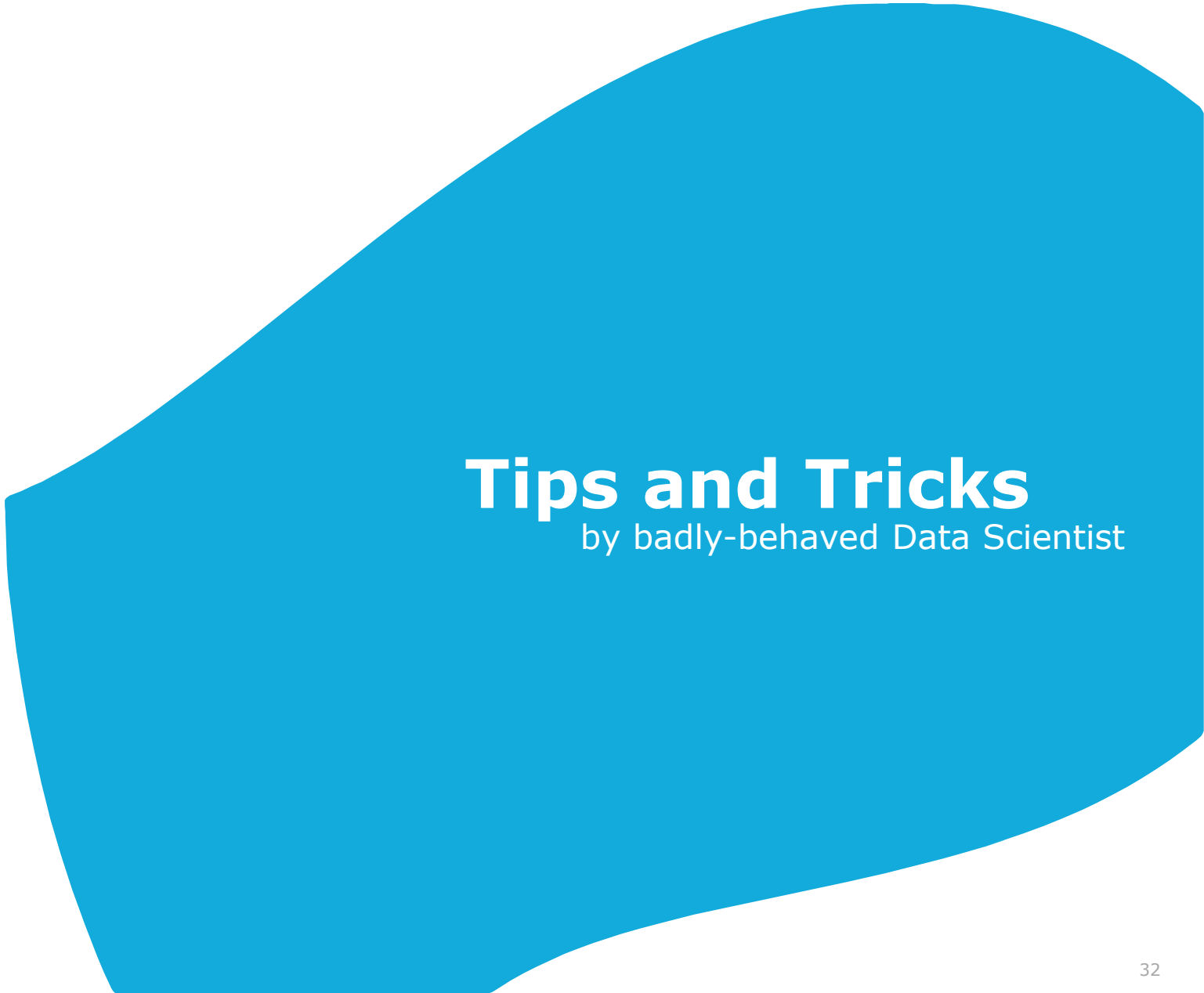
Image source: bit.ly/2kWjwyw



Algorithms



Source: bit.ly/2HuB6II



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

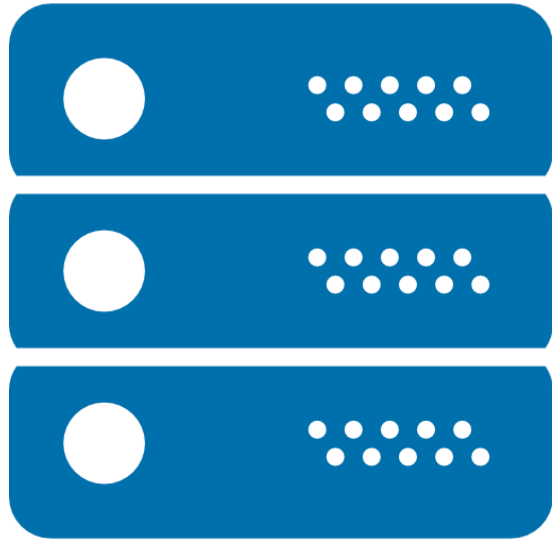


Is Machine Learning Magic?

70% / 30%
train / test
split

3/4 Equipment

Server



Moore's Law

**# of
transistors
x2 every
two years**

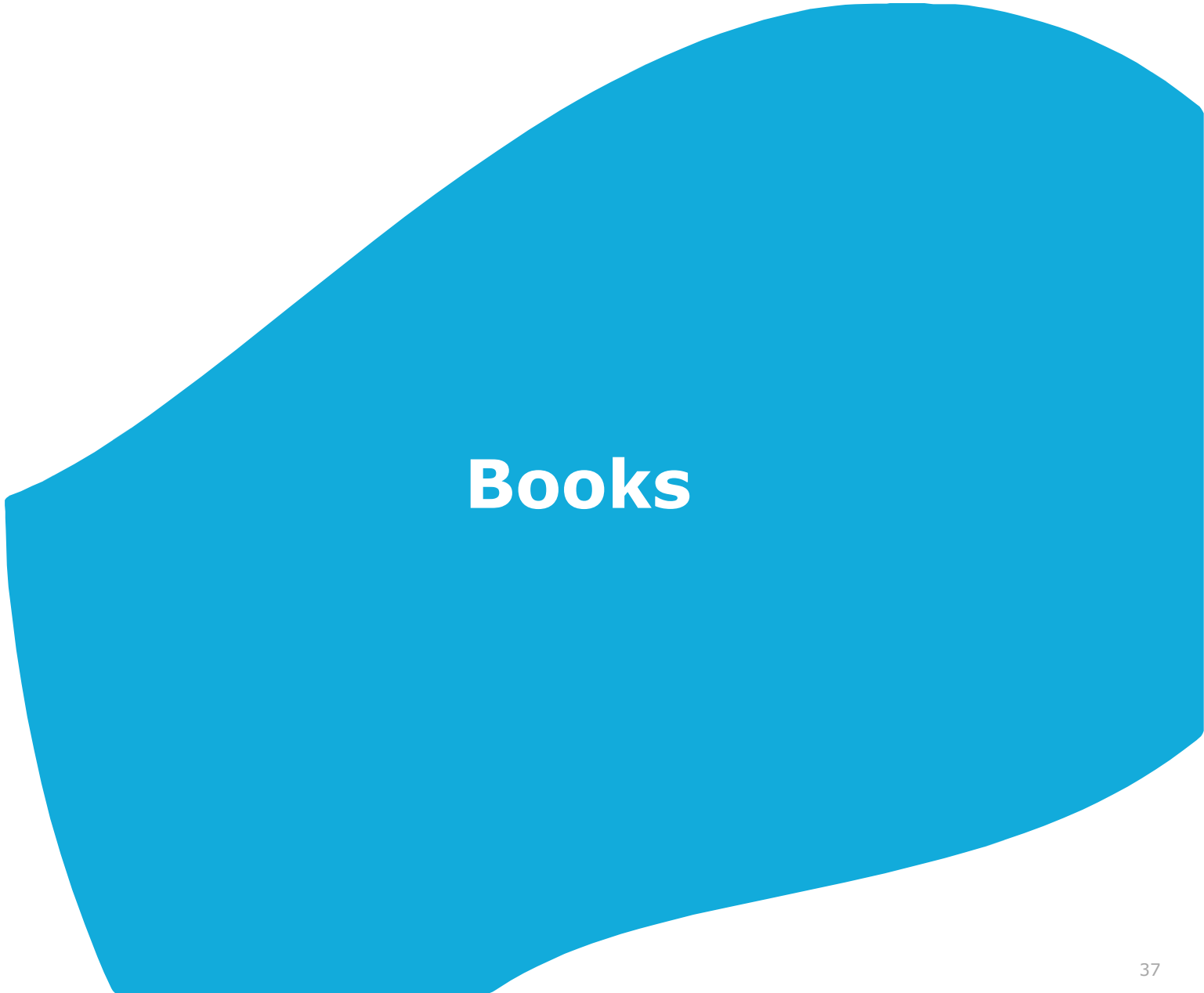
4/4 Domain knowledge

Books



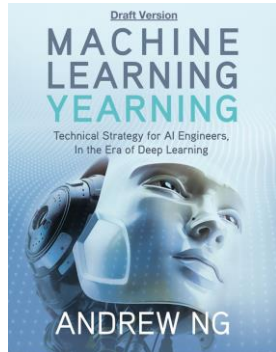
The Role of Domain Knowledge

**Domain
knowledge
is important**



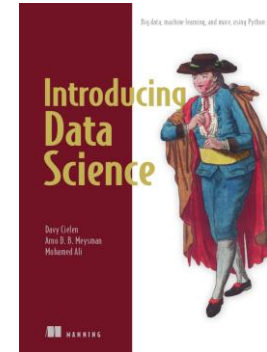
Books

Books



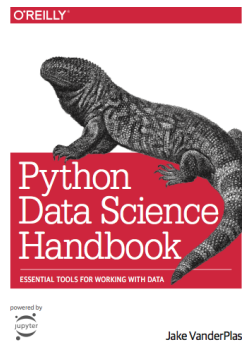
Machine Learning Yearning by Andrew Ng

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



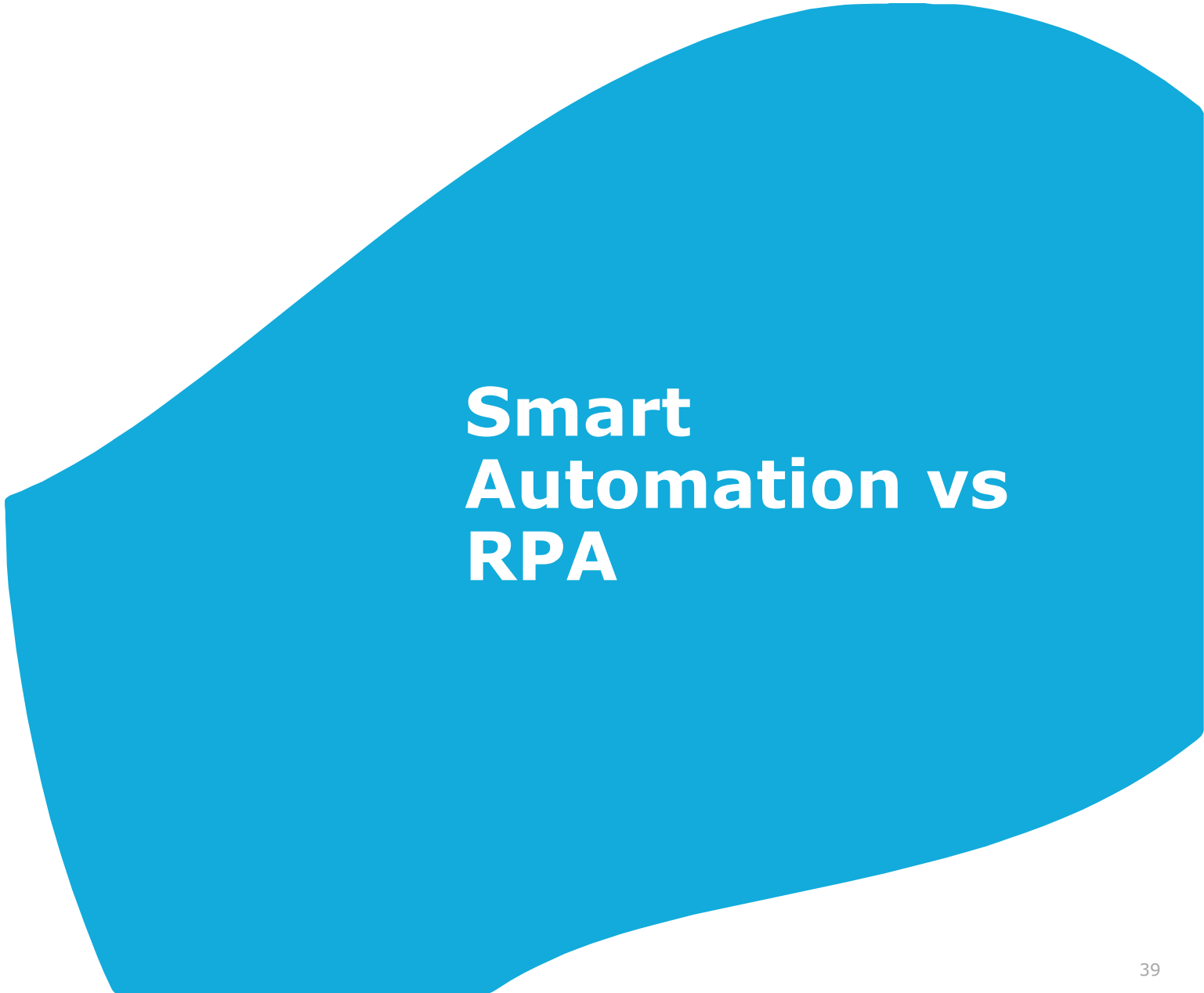
Introducing Data Science by Davy Cielen, Arno Meysman, and Mohamed Ali

- in English
- in Russian



Python Data Science Handbook by Jake VanderPlas

- [in English](#)
- in Russian



Smart Automation vs RPA

RPA

RPA (**Robotic Process Automation**) tools perform "if, then, else" statements on structured data, typically using a combination of user interface (UI) interactions, or by connecting to APIs to drive client servers, mainframes or HTML code.

An RPA tool operates by mapping a process in the RPA tool language for the software "robot" to follow, with runtime allocated to execute the script by a control dashboard.

Source: [Gartner](#)

RPA tool vendors

Specialist RPA software providers

- AutomationEdge
- **Automation Anywhere**
- **Blue Prism**
- Contextor
- EnableSoft
- Epiance
- Kryon Systems
- OpenConnect
- Softomotive
- **UiPath**

Software providers with multiple software products, including RPA

- AntWorks
- HelpSystems
- Infosys (EdgeVerve Systems)
- Jacada
- Kofax
- Nice
- Pegasystems
- Perpetuuti
- Verint
- Redwood Software
- Servicetrace
- SpiceCSM
- Winshuttle
- **WorkFusion**

IT or BPO service providers that provide proprietary RPA software platforms

- Another Monday
- Cognizant
- Conduent
- Sutherland
- Syntel
- Tech Mahindra

Smart Automation

Smart Automation accomplishes a task repeatedly without human intervention.

Smart Automation includes technologies such as machine learning, natural language processing/understanding, optical character recognition.

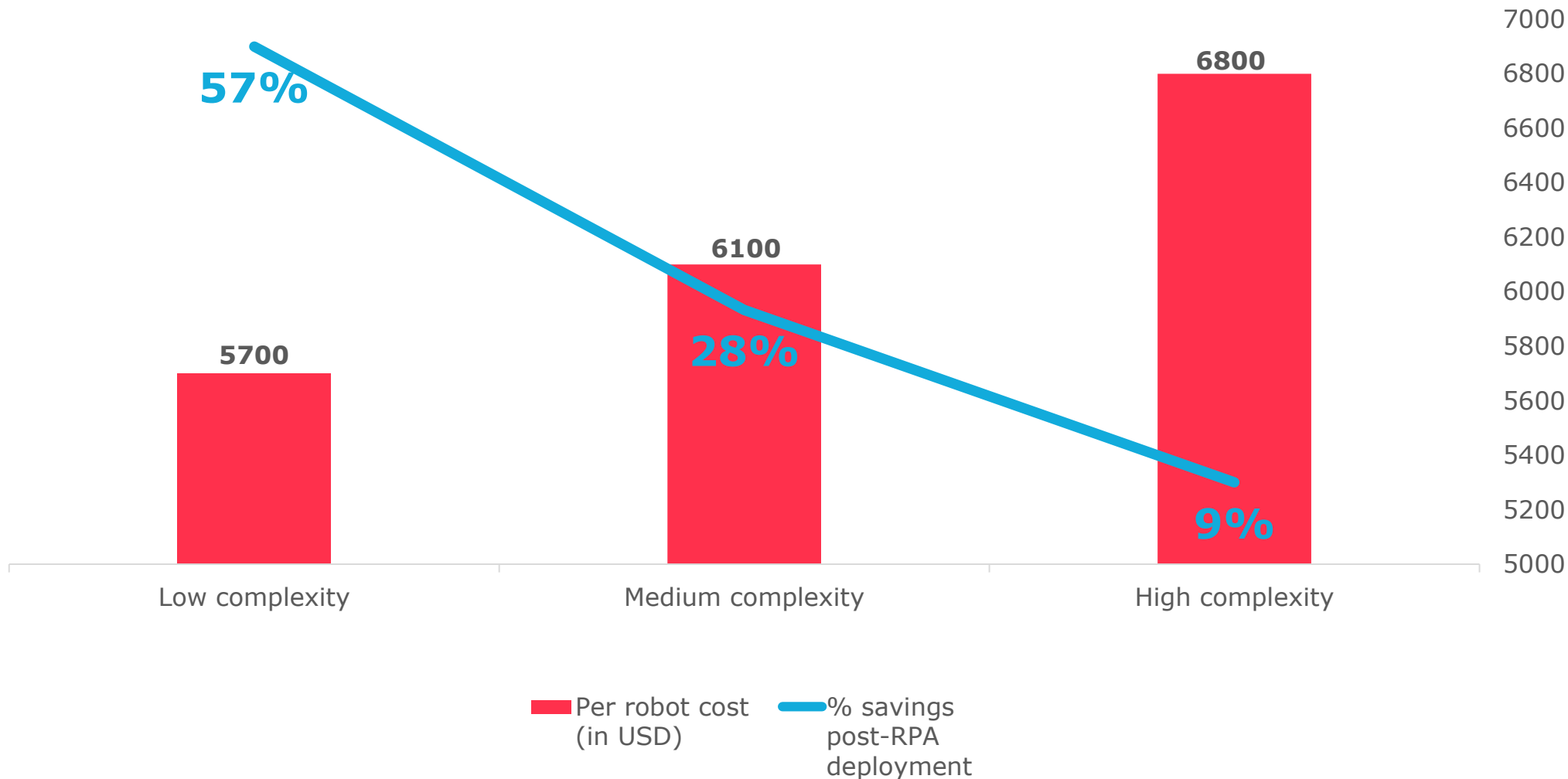


Smart Automation vs RPA

Process complexity and automation opportunities in financial services

Payment processing	Tax calculation & filing	Reconciliations	Invoice processing	Trade processing & settlements	Funds transfer & review	Account closure	Loan account & credit line setup
Customer account opening	Interest payments	Setting credit limits	Periodic accounting books closure	Loan disbursements	KYC/AML process	Loan applications processing	Updating accounts details
Regulatory reporting	Card account interchange & settlement	Trade confirmations & settlements	Transactions authorization	Tax claims management	Document processing	Overdraft/ advances provision management	Credit underwriting
Foreclosure audit review & resolution	Collateral review	Personalized offers & promotions	Risk exposure monitoring	Internal comm monitoring	Fraudulent transaction detection	Customer queries	Personalized financial advice

RPA. Variation of cost and savings



Scenario: Process with 100 FTEs pre-RPA deployment