



NATIONAL RESEARCH
UNIVERSITY

School of Data Analysis and Artificial
Intelligence Department of Computer Science

DATA SCIENCE FOR BUSINESS

Lecture 1. Introduction to Data Science

Moscow, April 8th, 2022.



COURSE TECHNICALITIES

Lectures: Friday 18.10 - 19.30 , 10 lectures

Seminars: Friday 19.40 - 21.00, 10 seminars

ZOOM: <https://us02web.zoom.us/j/88281792929>

ZOOM: <https://us02web.zoom.us/j/86324644675>

Class website: <http://www.leonidzhukov.net/hse/2022/datascienc>

Seminar github: <https://github.com/kurmukovai/ds-for-business-2022>

Telegram Group: <https://t.me/joinchat/ENzQEhr-hra2WhEjxvgayw>

Modeling software: Python, Orange

TEACHING TEAM

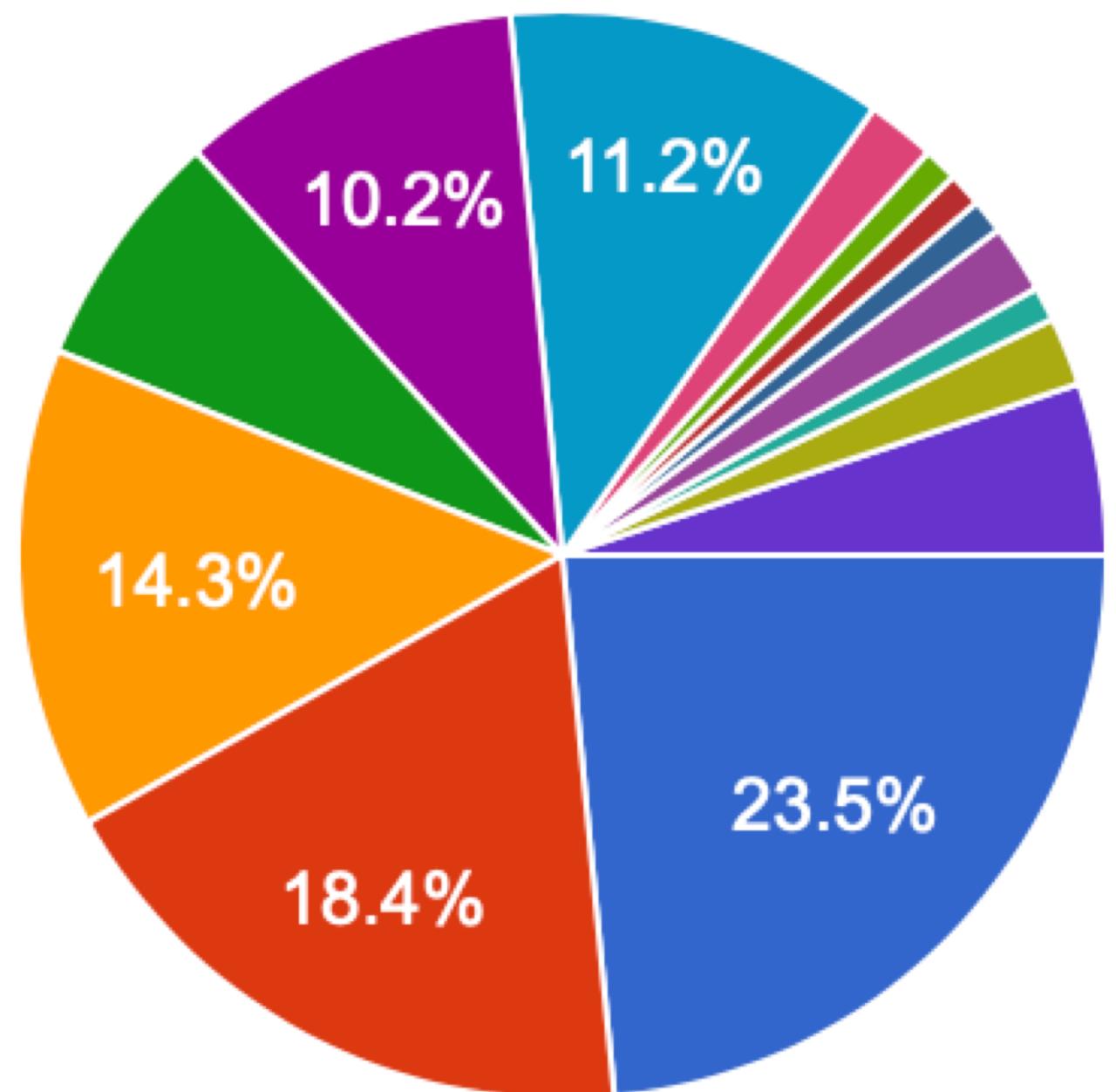


Prof. Leonid Zhukov
lzhukov@hse.ru



Anvar Kurmukov
kurmukovai@gmail.com

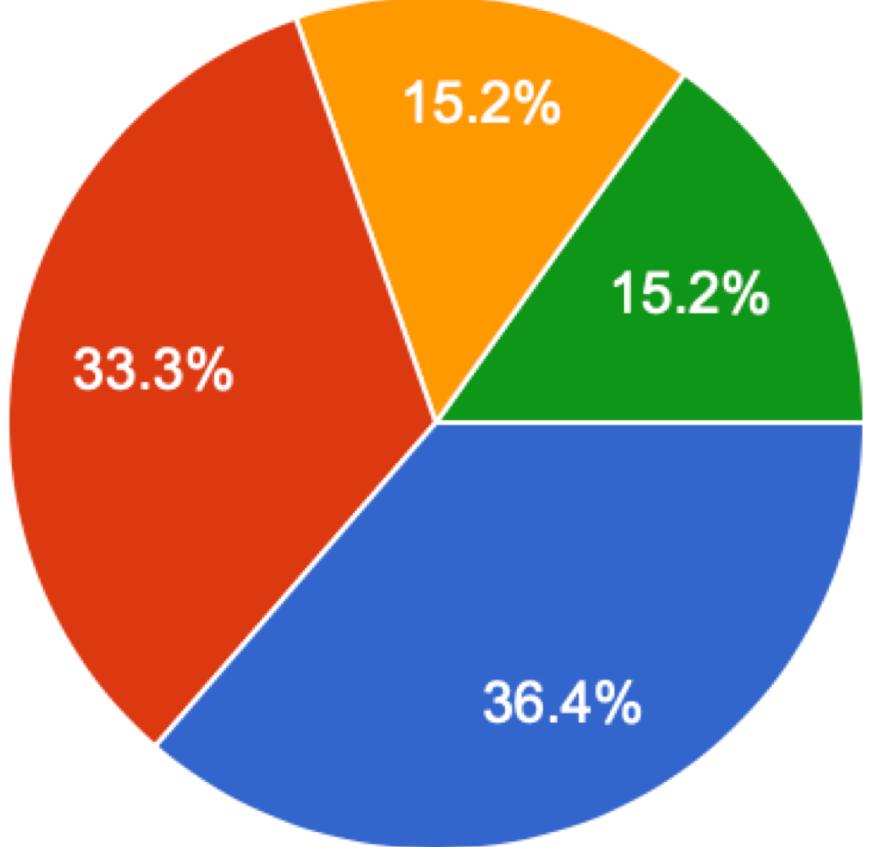
CLASS 2022



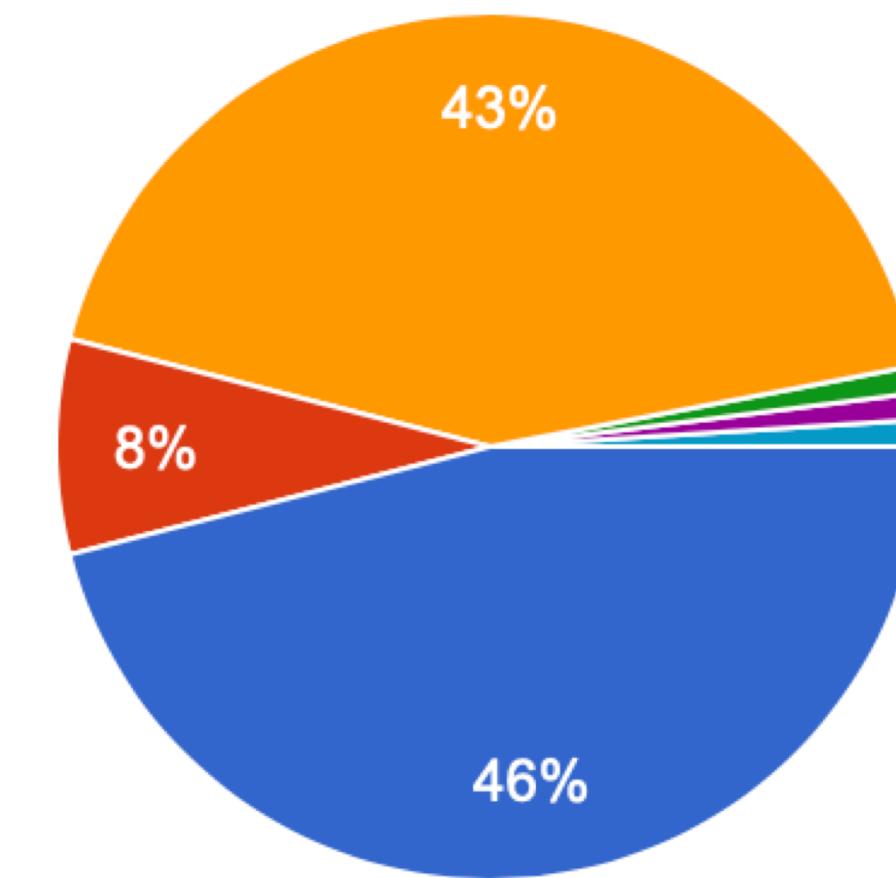
- Высшая школа бизнеса
- Факультет компьютерных наук
- Факультет социальных наук
- Факультет экономических наук
- Московский институт электроники и...
- Факультет Санкт-Петербургская школа...
- Институт статистических исследова...
- Институт когнитивных нейронаук

- Институт торговой политики
- Кафедра менеджмента инноваций
- Факультет гуманитарных наук
- Факультет информатики, математик...
- Факультет математики
- Факультет мировой экономики и мир...
- Факультет Санкт-Петербургская школа...
- Другой факультет

CLASS 2022



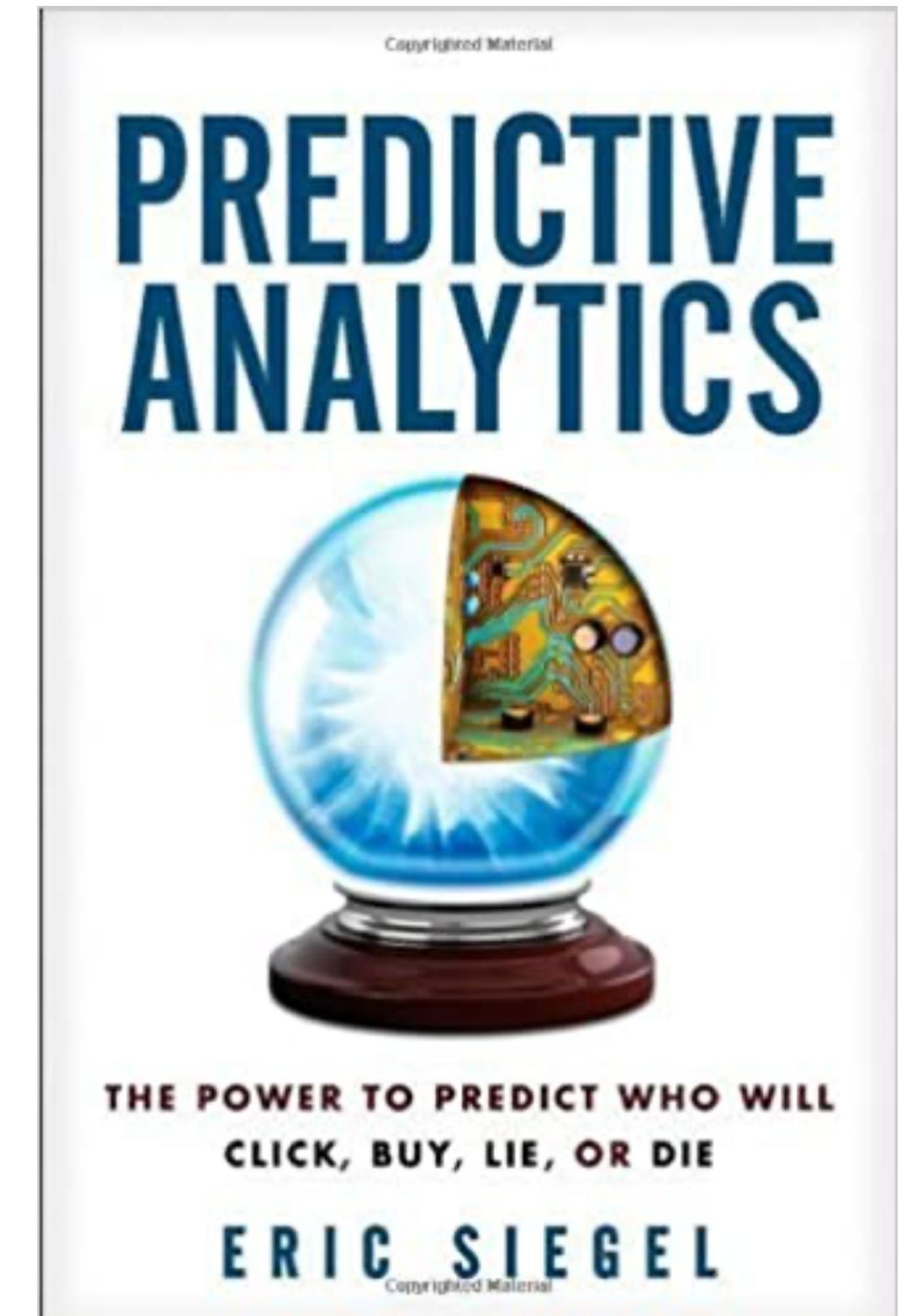
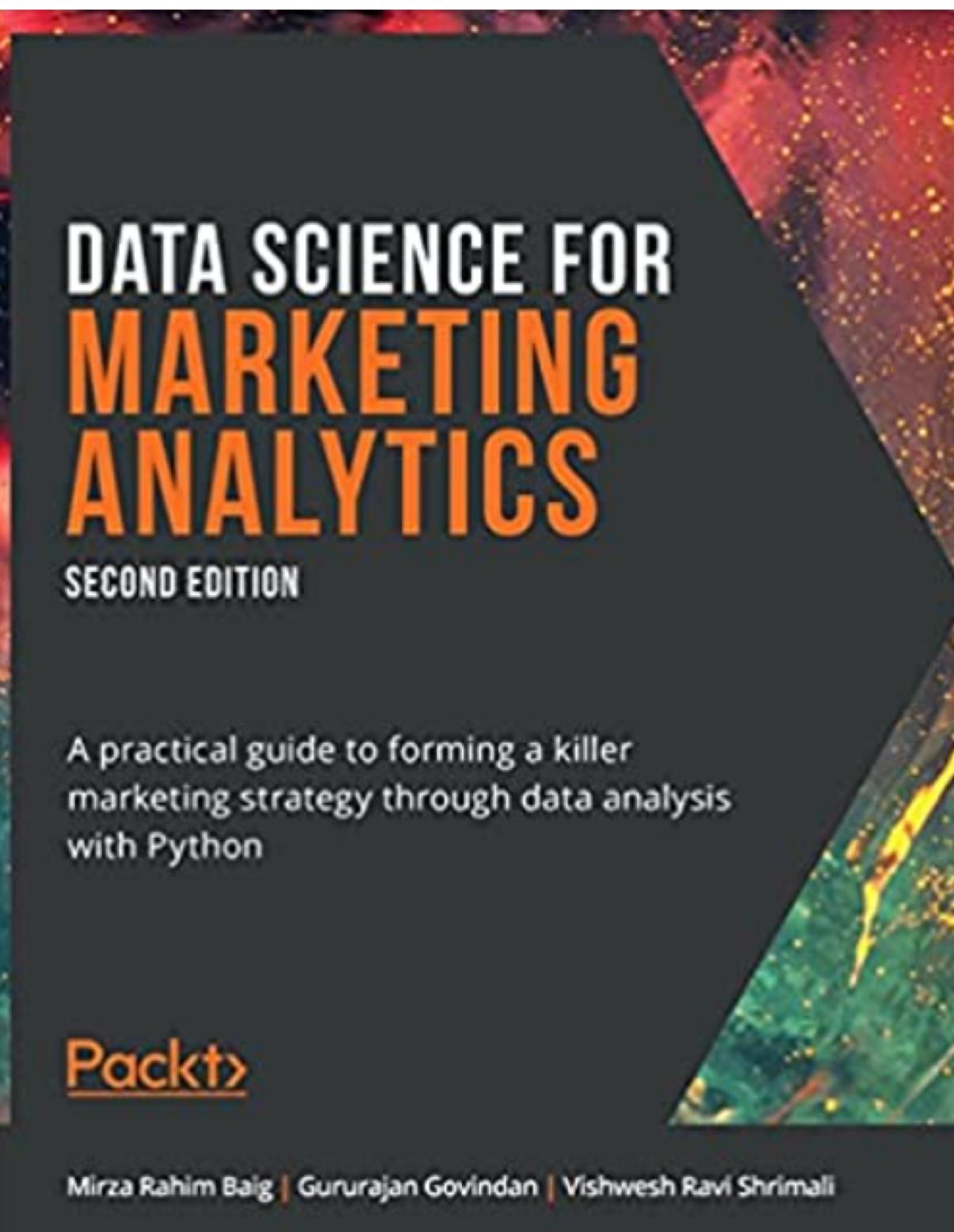
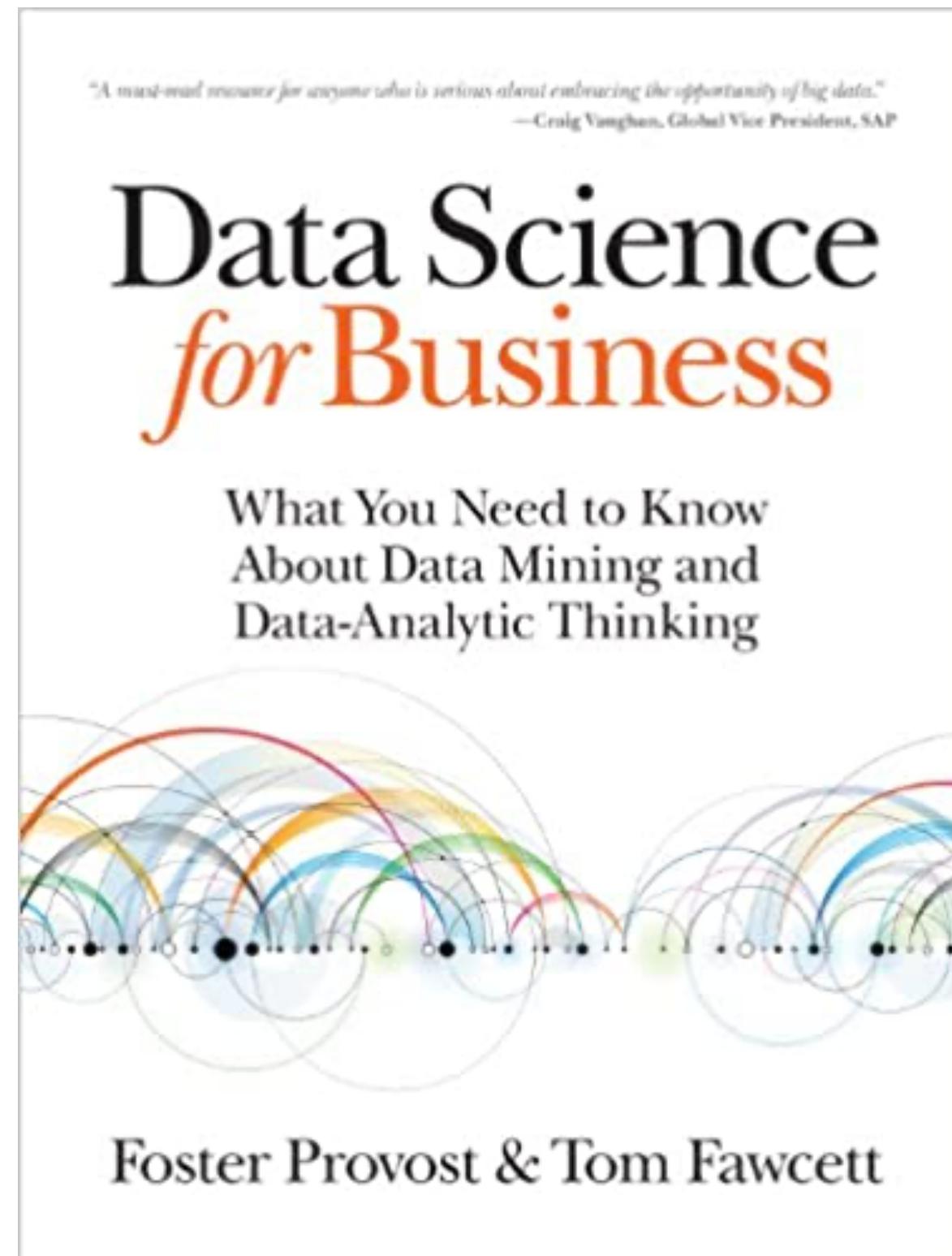
- beginner level (had an introductory Python course)
- good familiarity (use Python in more than one edu courses)
- advance (use Python to earn money)
- not familiar with Python



- DS/ML course in HSE
- DS/ML course on MOOCs platforms
- no DS/ML course
- Yandex School of Data Analysis finished
- no
- on Coursera HSE



TEXTBOOKS FOR THE COURSE





COURSE SCHEDULE

Lecture topics

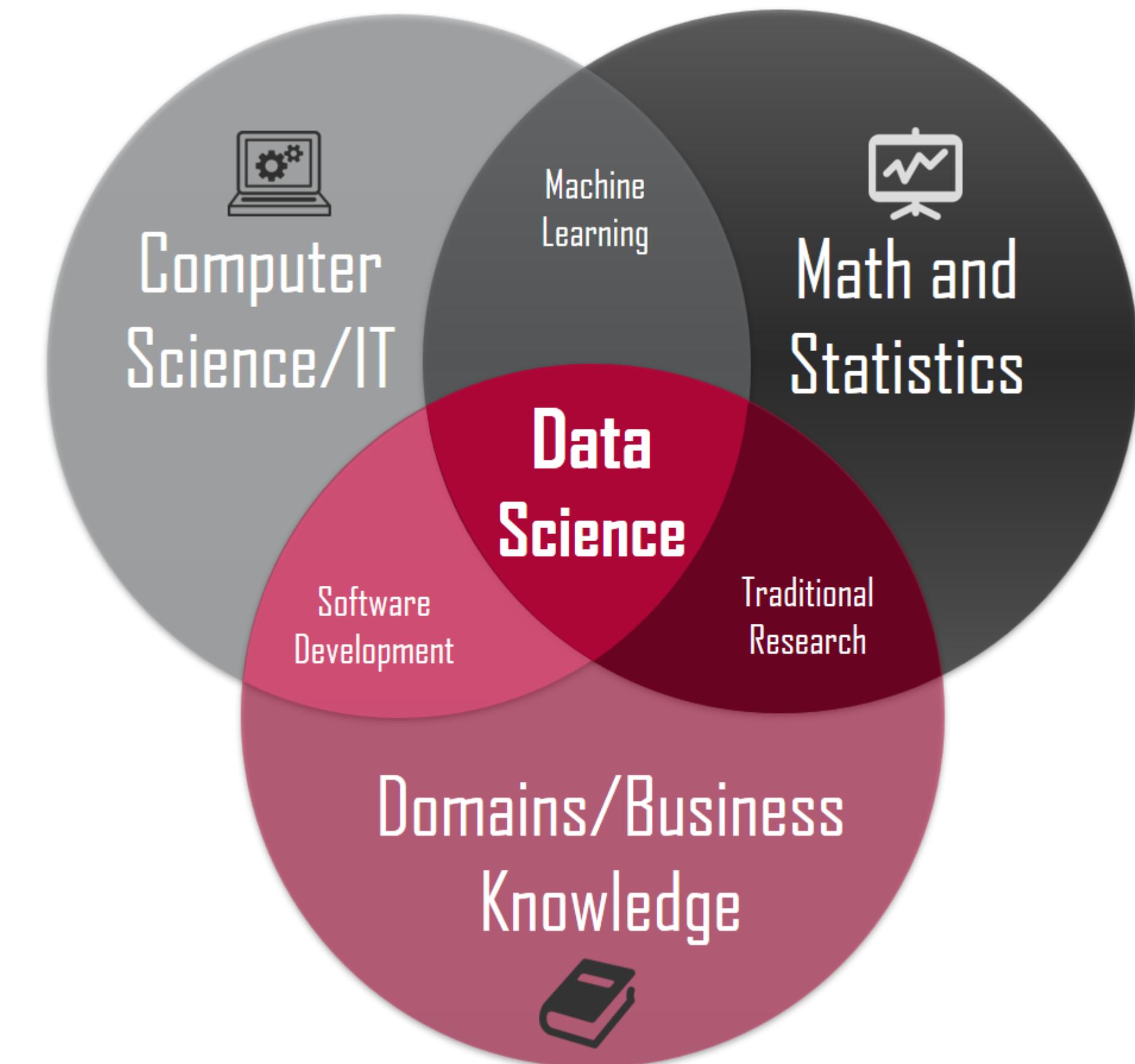
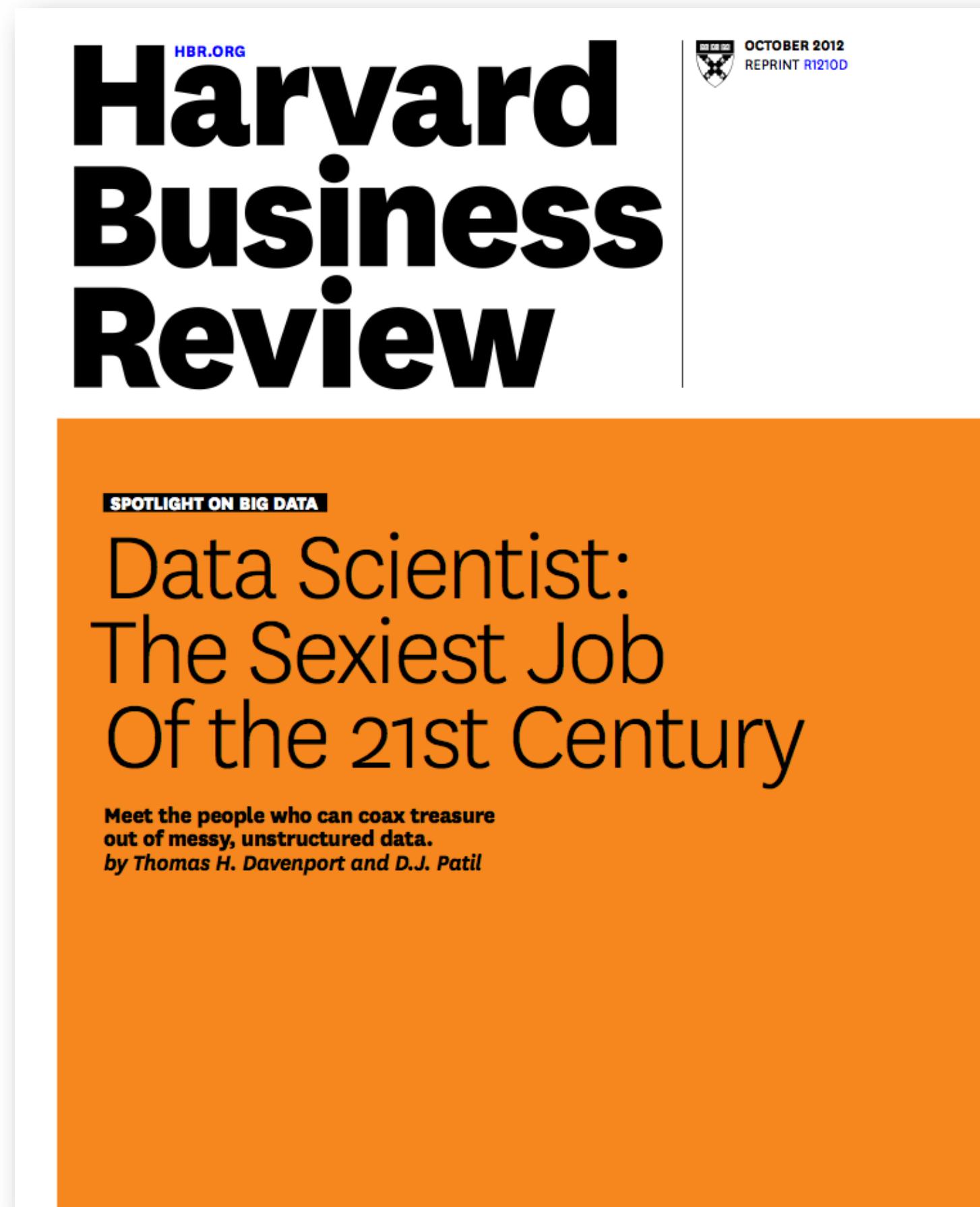
1. Introduction to data science.
2. Data analysis and machine learning
3. Case study 1: Churn modeling
4. Case study 2: Customer segmentation
5. Case study 3: Demand forecast
6. Case study 4: Personalization
7. Case study 5. Pricing
8. Impacting the business

Seminars - exercises

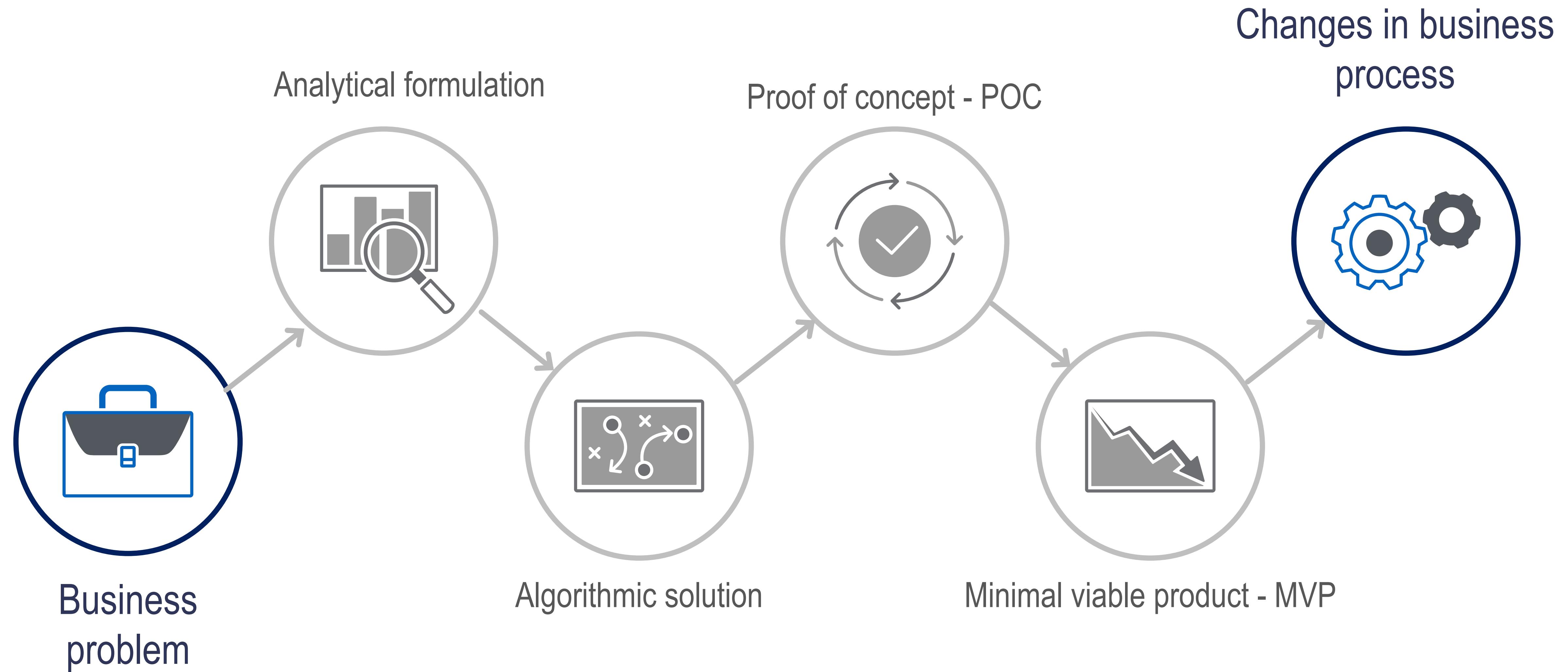
1. Data flow modeling and Orange, ETL process, data exploration
2. ML modeling pipeline
3. Classification
4. Clustering
5. Regression
6. Recommender systems
7. Time series forecasting
8. Problem solving



DATA SCIENCE



DATA SCIENCE BUSINESS PROCESS



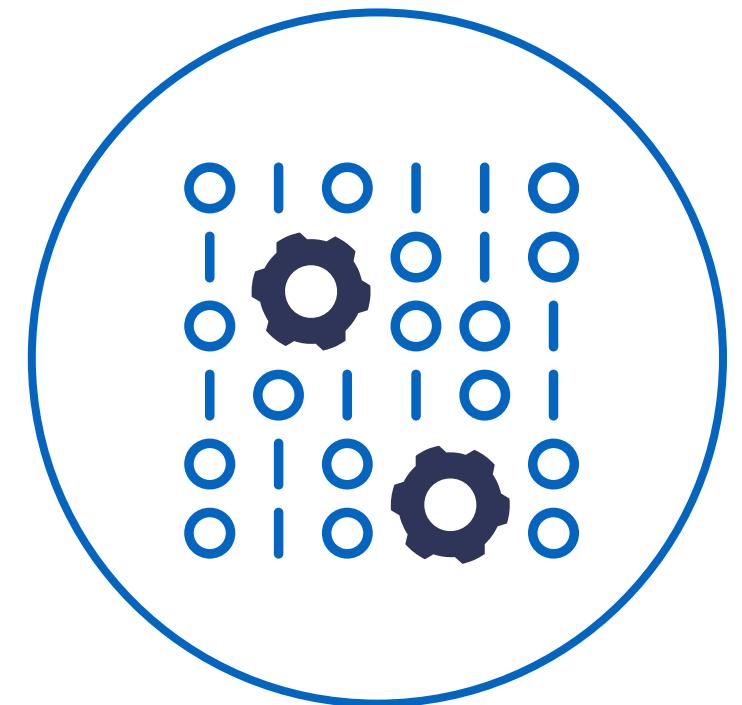
ANALYTICAL METHODS IN DATA SCIENCE



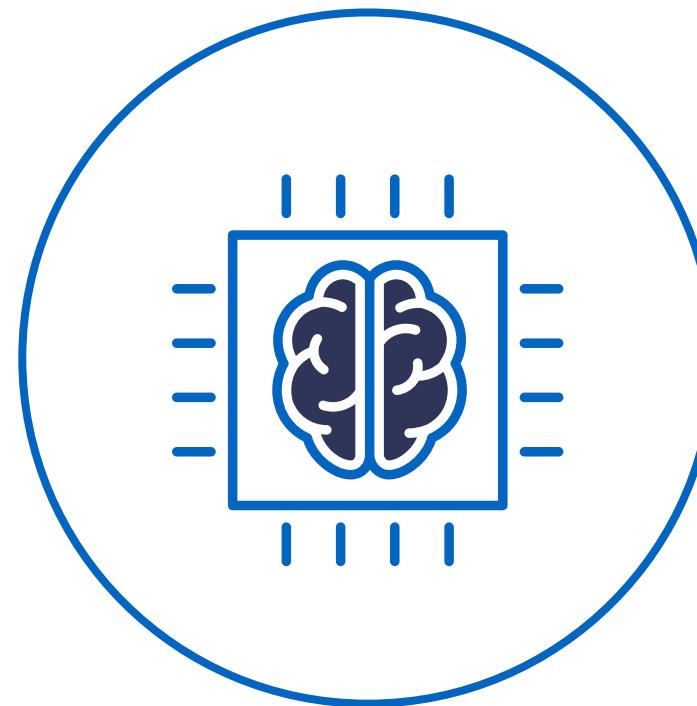
Predictive modeling
Machine learning
Data mining



Operations research
Optimization



Agent based modeling
Simulations

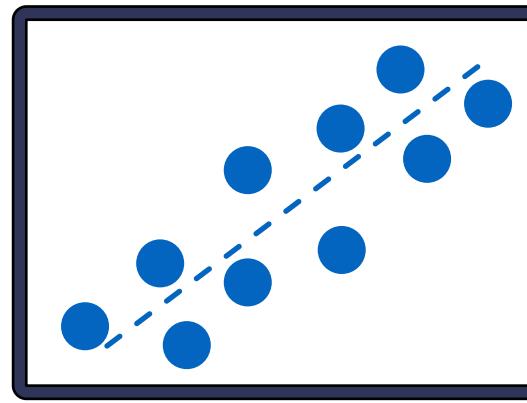


Geo analytics
Text analysis (NLP)
Computer vision

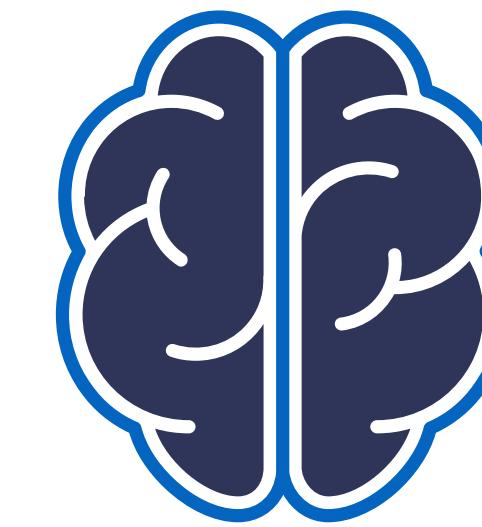
THREE MAIN REASONS TO USE ML IN BUSINESS



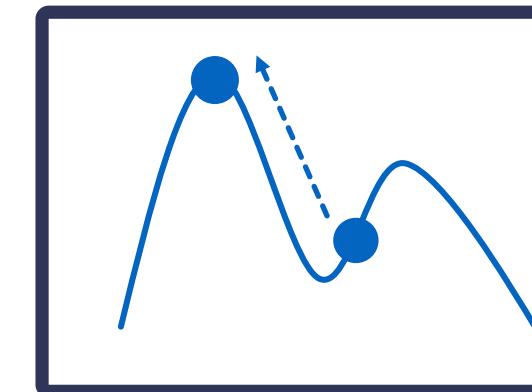
Detect & describe



Predict



Explain



Optimise and improve

SIMPLE EXAMPLE

Statistical Analysis

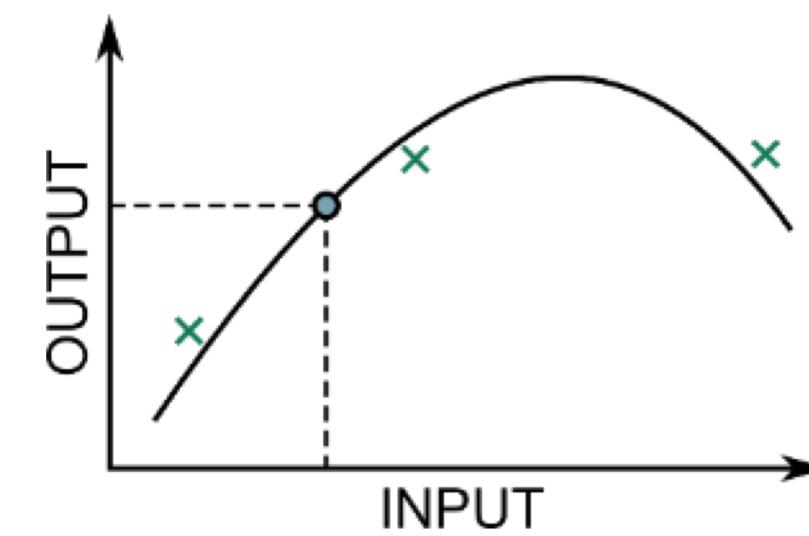
"Measure and understand"



Data exploration
Descriptive statistics

Predictive modeling

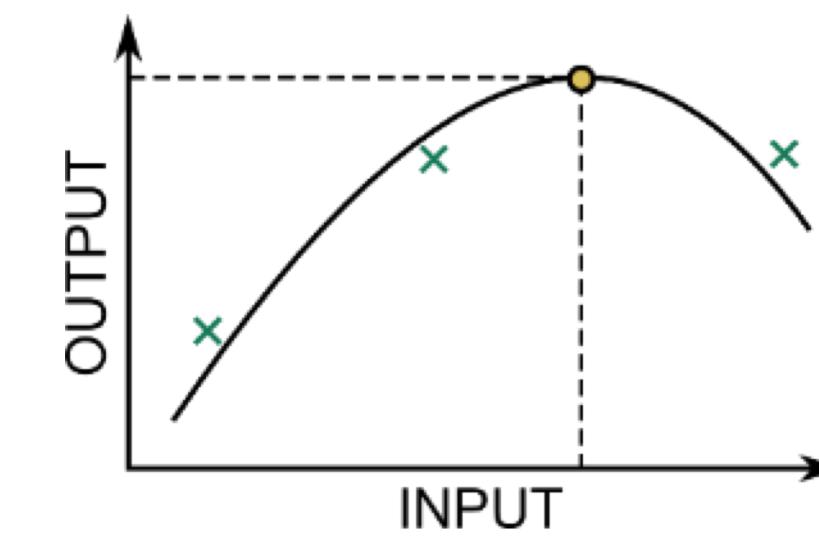
"ML predict outcomes"



Finding patterns
Predicting outcomes

Optimization

"Make optimal decisions"



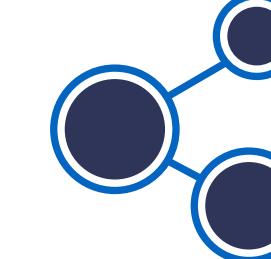
Finding optimal values
Finding optimal parameters

BUSINESS USE CASES



Consumer Goods

Demand forecast
Marketing personalization
Pricing and promo effectiveness
Assortment optimization
Cross sell and upsell



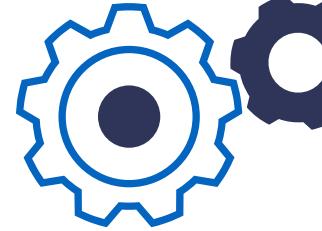
Telecoms

Next best offer
Churn and retention modeling
Network optimization
Infrastructure capacity and utilization



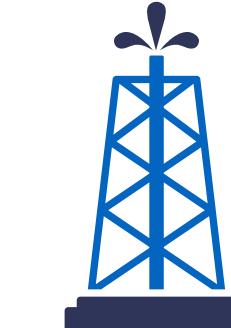
Banking/FI

Credit risk assessment
Fraud detection
Claim management
Churn and retention modeling
Next best offer



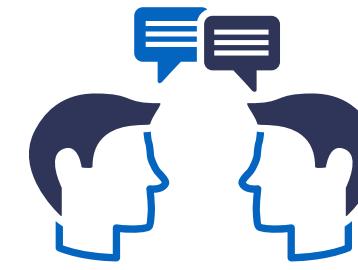
Industrial Goods

Manufacturing process optimization
Predictive maintenance
Demand and supply forecast
Operations planning



Energy

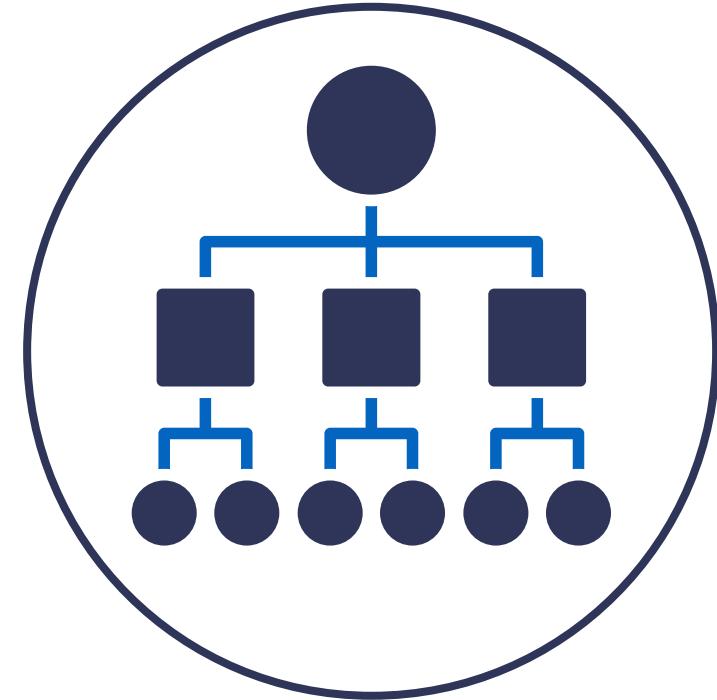
Production optimization
Predictive maintenance
Logistics optimization
Project risk management
Robotics and automation



Enterprise

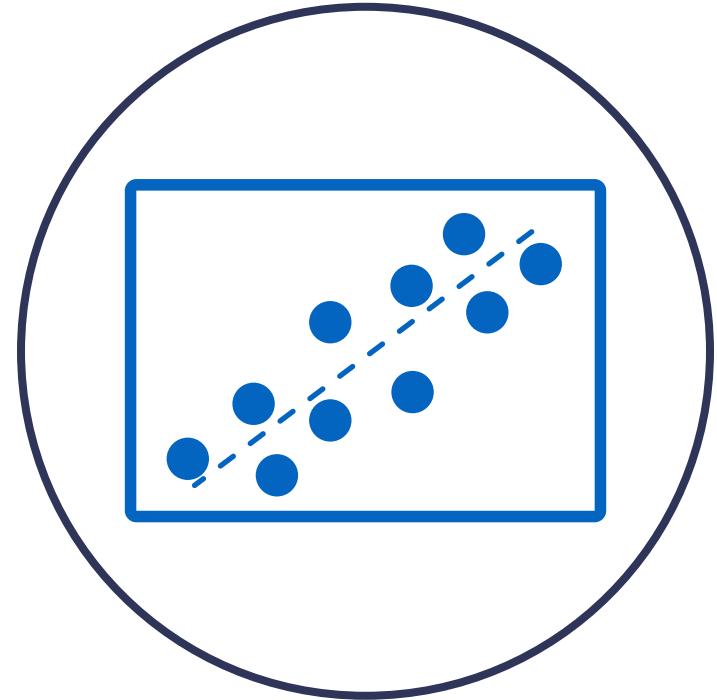
Back office automation RPA
Performance management
Workforce planning
Scenario simulations

THREE TYPES OF MACHINE LEARNING



Unsupervised Learning

Aim to discover structure:
no target variable known



Supervised Learning

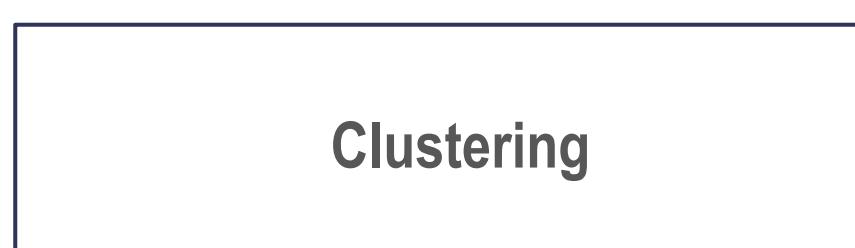
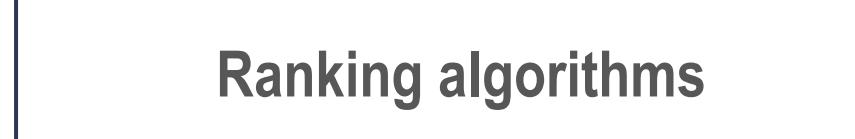
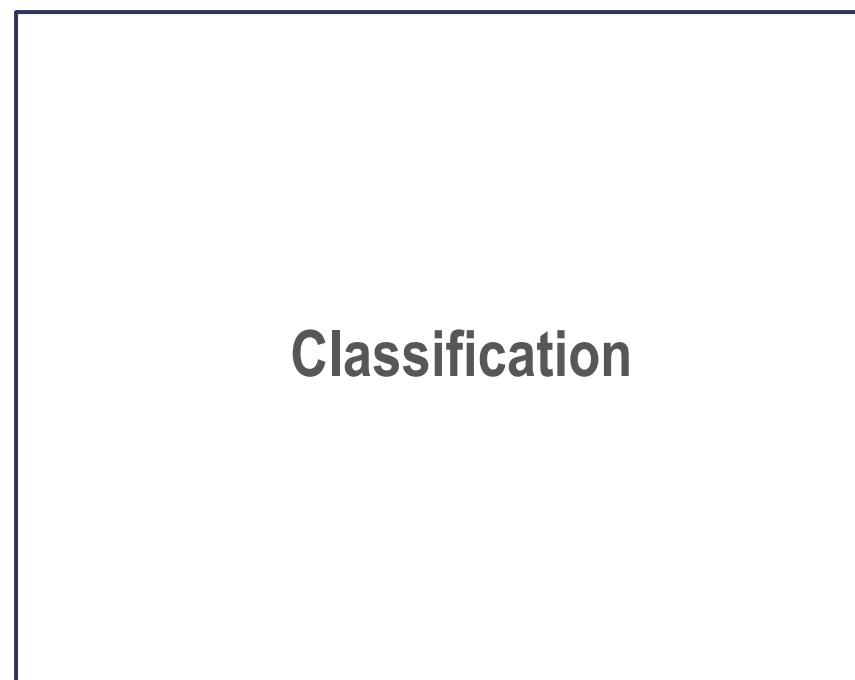
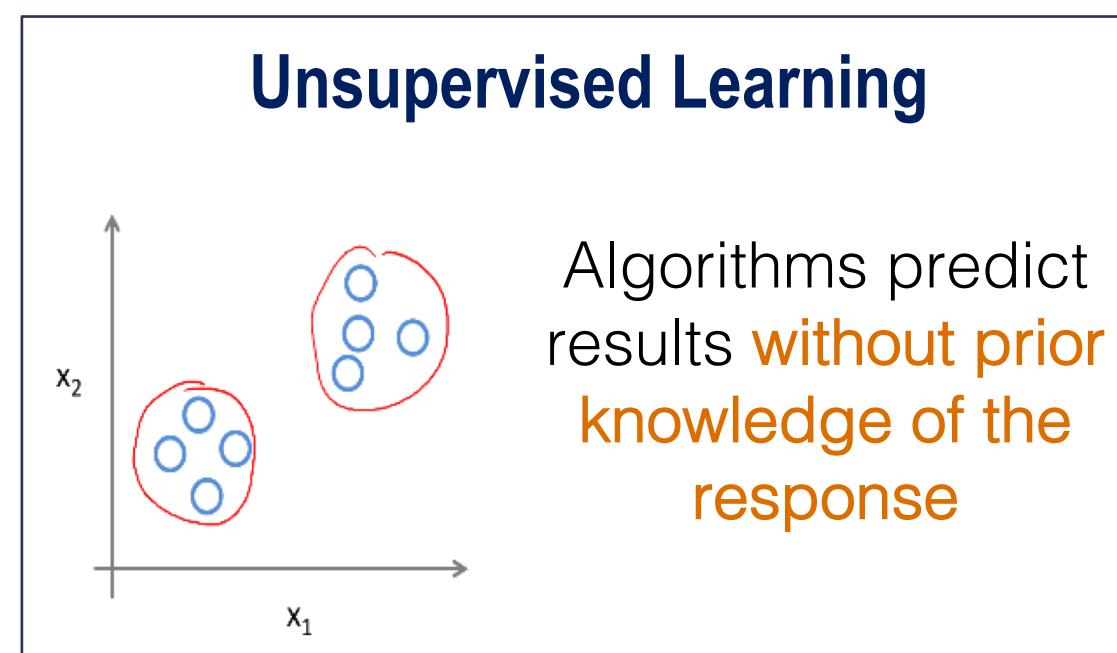
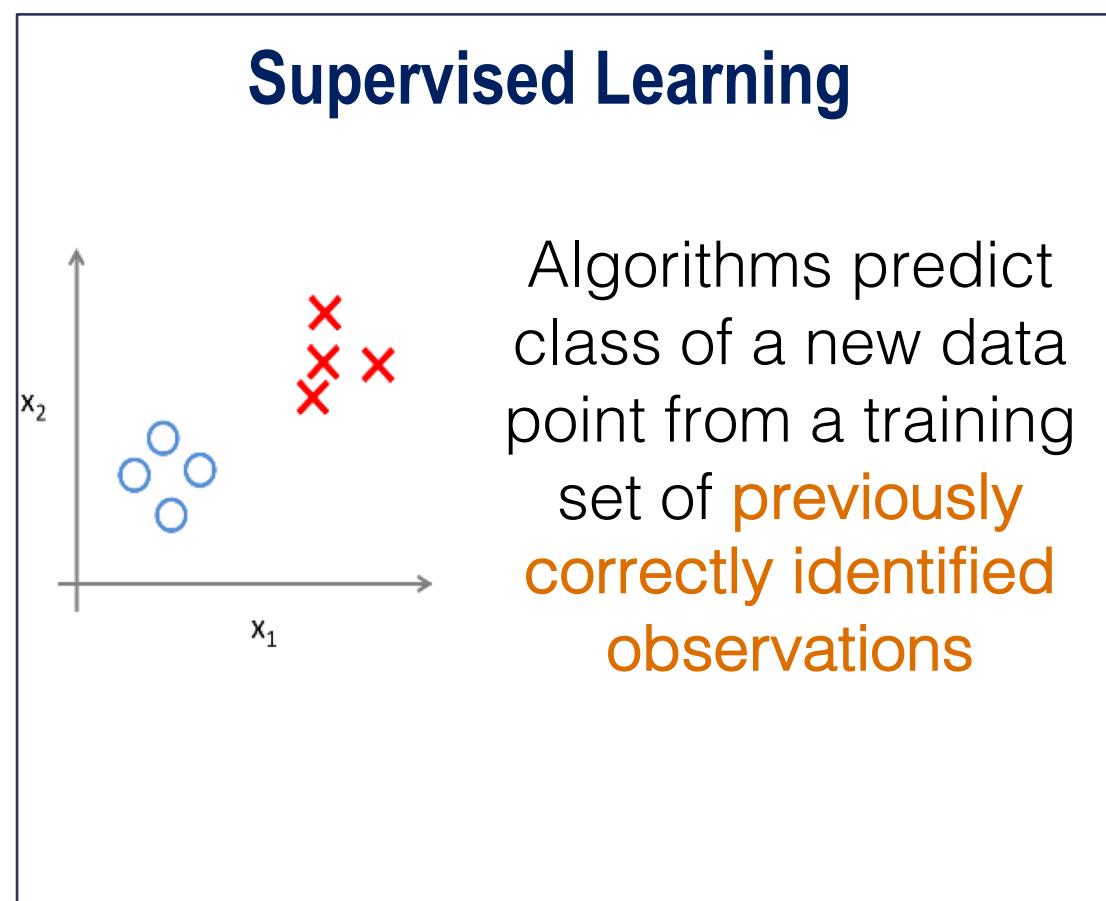
Aim to predict or model a
known target



Reinforcement Learning

Optimise actions in a way that
maximises cumulative reward

Types of algorithms



Algorithm examples

Given examples of classes, the model assigns new input data to classes

- **Decision trees**, k-nearest neighbors (kNN), Logistic regression
- **Random Forest**, Support Vector Machines (SVM), Gradient Boosted Decision trees (GBT)
- Neural networks + **Deep Learning**

Given several classes the model assigns input data to classes

- **Linear regression**, Elastic nets
- **Regression trees**

Given ordered pairs examples the model ranks new data

Divide the input data into groups with similar data points assigned to the same group

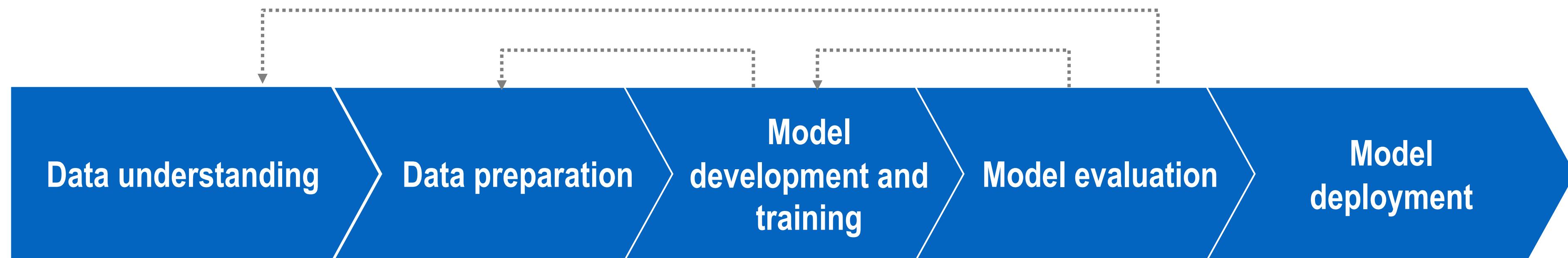
- **k-means**, **spectral**

Mapping input data in lower dimensional space
PCA

Estimates the probability distribution of values

Detecting outliers in data

MODELING PROCESS PIPELINE



Understanding the data at hand, checking the quality

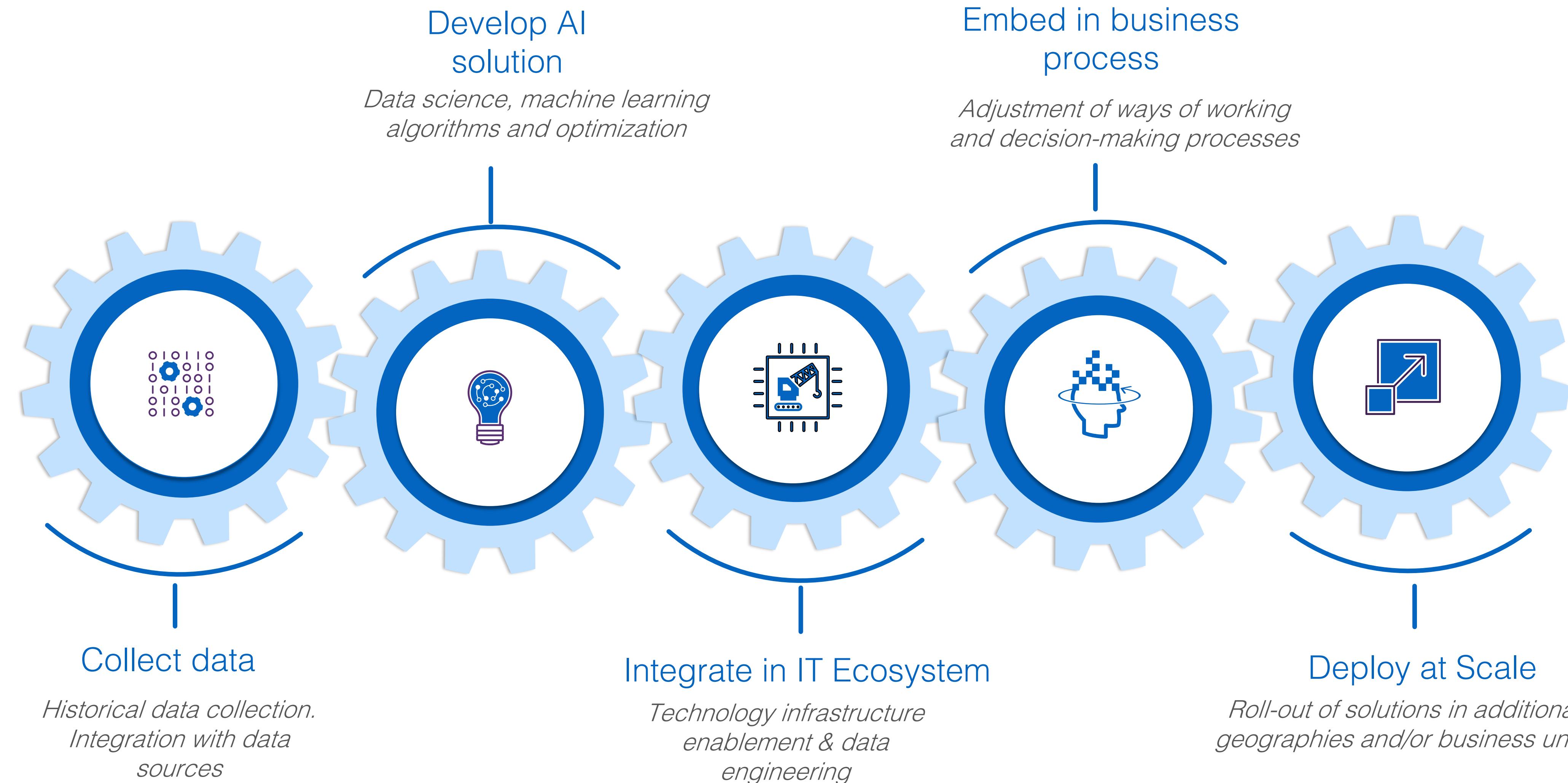
ETL activities to get data into right format, cleaning, aggregation

Developing and training the model on given dataset

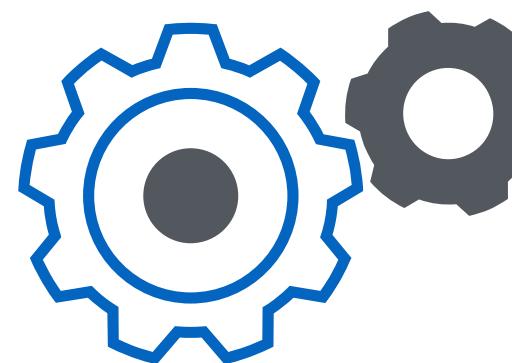
Validating model performance including business metrics on the holdout set

Running the pilot

INTEGRATION OF ANALYTICS PROJECT INTO BUSINESS



ANALYTICS PROJECTS SUCCESS FACTORS



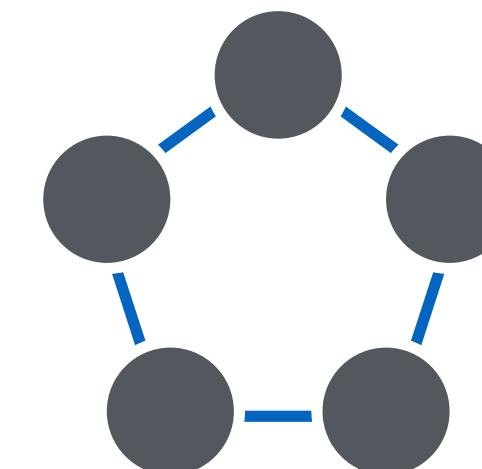
Algorithms & Data

- Data analysis
- Algorithm development



Technology/IT

- Algorithm industrialization
- Digital platforms development



Business transformation

- Business process redesign
- Enablement
- Change management



NATIONAL RESEARCH
UNIVERSITY