Introduction

5ARE0: DATA ANALYSIS & LEARNING METHODS (2025 – 2026)

Uzay Kaymak, Jheronimus Academy of Data Science, u.kaymak@tue.nl

Master: Artificial Intelligence & Engineering Systems

Outline

Motivation

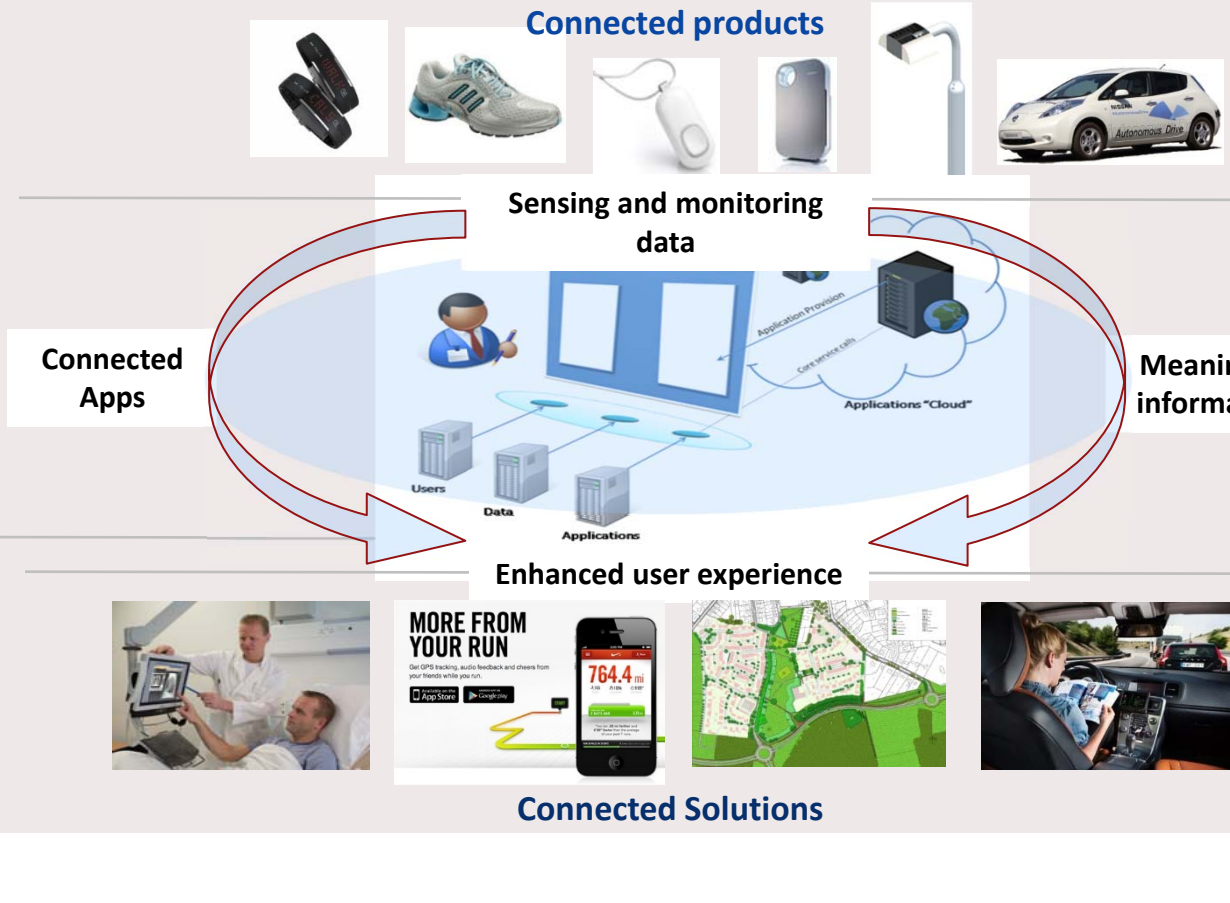
Course organization

Artificial intelligence & engineering systems

Preliminaries: modeling with data

TU/e

The Always-On Society



Everywhere Analytics

From Deloitte

TU/e

Real-world examples

Smart business solutions

Process improvement

Monitors and analyses events in an organization and proposes business improvement actions.

Smart power grids

Measures, monitors, and manages energy production, transport, and consumption is heterogeneous distributed grids.

Clinical decision support

Provides instant clinical decision support by correlating information from different part of uncorrelated sources.

Slide by DSC/e 5

AI systems: self-driving cars

O cmgt<o gvco qty qtmu ~Dtqpxgto grfkpi<Igw{"Ko cigu1kUvqemrjqvq

https://getfello.com/wp-content/uploads/2017/03/Google-self-driving-car-prototype-front-three-quarters1.jpg

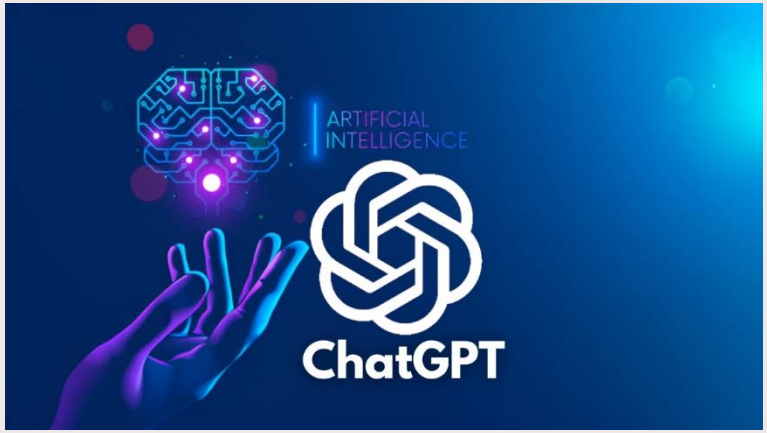
TU/e

AI Systems: IBM Watson

IBM Watson Health focused on value-based care, physician burnout, personalized medicine | Healthcare IT News

IBM Watson moet helpen bij uitlezen röntgenfoto's - ICT&health (icthealth.nl)

AI Systems: ChatGPT



https://hawar.no/2023/04/what-is-chatgpt-how-do-you-use-it-how-do-you-benefit-from-it/

https://www.datacamp.com/blog/a-chat-with-chatgpt-on-the-method-behind-the-bot

8

AI Systems: robot football

TU/e

In this course:

We discuss various approaches to

- **collect,**
- **process,**
- **store, and**
- **analyze**

data for creating AI/ML based engineering systems and solutions.

Remember: data and knowledge do not come cheap!

11

TU/e

Course organization

12

Learning objectives

- design a basic data acquisition system for goal-targeted data collection
- pre-process data for further analysis
- perform feature engineering
- apply basic data analysis techniques
- build machine learning models for clustering, classification and regression
- develop non-linear machine learning models in Python
- present the results of data analysis in writing

Lecturers and teaching assistants

Prof.dr.ir. U. Kaymak
Course Coordinator; Lecturer
u.kaymak@tue.nl

Dr. F. van der Sommen, MSc.
Lecturer
fvdsommen@tue.nl

Dr.ir. L.A. Zavala Mondragon
Lecturer
l.a.zavala.mondragon@tue.nl

Q. Khaled Abo Shama, MSc.
Teaching Assistant
q.k.h.abo.shama@tue.nl

P.C. Gort, MSc.

L. Abdi, MSc.

R. Pavelkin, MSc.

14

TU/e

Meetings

15 sessions

- **typically, 2 x 2 hrs./week lecture, 2 x 2 hrs./week instruction, 8 weeks long, except for the week of 23 September (see schedule for the latest information)**

Tuesday 1 – 4, various rooms, Friday 5 – 8, various rooms
Lectures: introduce and explain main concepts
Instructions for practice exercises and working on the assignment
Q&A session at the end of the quartile

Further questions can be asked through Canvas (preferred method) or during instructions

15

TU/e

Planning (1)

WEEK	DAY/TIME	ROOM	TYPE	TOPIC
1	02-09-2025 / 08:45 - 10:30	Auditorium 04	Lecture	Introduction to the course, basic learning types, basic ML tasks
	02-09-2025 / 10:45 - 12:30	Gemini-Zuid 3A.10 Gemini-Zuid 3A.12	Instruction	Team composition, software installation
	05-09-2025 / 13:30 - 15:15	Atlas -1.715	Lecture	Basic data collection, data representation, pre-processing
	05-09-2025 / 15:30 - 17:15	Flux 1.02	Instruction	Python exercises, introduction to Assignment 1
2*	09-09-2025 / 08:45 - 10:30	Alpha 0.98	Lecture	Feature engineering, feature selection, feature extraction
	09-09-2025 / 10:45 - 12:30	Gemini-Zuid 3A.10 Gemini-Zuid 3A.12	Instruction	Python exercises, work on Assignment 1
	12-09-2025 / 13:30 - 15:15	Atlas -1.715	Lecture	Unsupervised learning, clustering techniques
	12-09-2025 / 15:30 - 17:15	Flux 1.02	Instruction	Python exercises, work on Assignment 1
3*	16-09-2025 / 08:45 - 10:30	Atlas -1.715	Lecture	Mixture models, fuzzy clustering, Kohonen maps, t-SNE
	16-09-2025 / 10:45 - 12:30	Atlas -1.715	Instruction	Python exercises, work on Assignment 1
	19-09-2025 / 13:30 - 15:15	Alpha 0.98	Lecture	Supervised learning, classification, logistic regression, k-NN, decision trees, model evaluation, complexity control
	19-09-2025 / 15:30 - 17:15	Alpha 0.98	Instruction	Python exercises, work on Assignment 1
4	23-09-2025 / 08:45 - 10:30	Alpha 0.98	Lecture	Supervised learning, linear regression, loss functions, regularization, regularized regression
	23-09-2025 / 10:45 - 12:30	Atlas -1.715	Instruction	Python exercises, work on Assignment 1
5*	30-09-2025 / 08:45 - 10:30	Alpha 0.98	Lecture	SVMs, kernels, kernel trick
	30-09-2025 / 10:45 - 12:30	Atlas -1.715	Instruction	Python exercises, work on Assignment 1
	03-10-2025 / 13:30 - 15:15	Alpha 0.98	Lecture	Perceptron, feedforward neural networks, backpropagation
	03-10-2025 / 15:30 - 17:15	Flux 1.02	Instruction	Python exercises, introduction to Assignment 2
	05-10-2025 / 20:00			Deadline for Assignment 1
6*	07-10-2025 / 08:45 - 10:30	Flux 1.02	Lecture	Parameter learning methods: LMS, steepest descent, maximum likelihood, loss functions (cont.)
	07-10-2025 / 10:45 - 12:30	Atlas -1.715	Instruction	Python exercises, work on Assignment 2
	10-10-2025 / 13:30 - 15:15	Alpha 0.98	Lecture	Learning in MLP, dealing with sequence data
	10-10-2025 / 15:30 - 17:15	Flux 1.02	Instruction	Python exercises, work on Assignment 2

16

TU/e

Planning (2)

7*	14-10-2025 / 08:45 - 10:30	Atlas -1.715	Lecture	Fuzzy sets, FIS
	14-10-2025 / 10:45 - 12:30	Atlas -1.715	Instruction	Python exercises, work on Assignment 2
	17-10-2025 / 13:30 - 15:15	Alpha 0.98	Lecture	ANFIS, RBFN
	17-10-2025 / 15:30 - 17:15	Flux 1.02	Instruction	Python exercises, work on Assignment 2
8	21-10-2025 / 08:45 - 10:30	Luna 1.050	Lecture	Application examples
	21-10-2025 / 10:45 - 12:30	Atlas -1.715	Instruction	Python exercises, work on Assignment 2
	24-10-2025 / 13:30 - 15:15	Alpha 0.98	Lecture	Review and Q & A
	24-10-2025 / 15:30 - 17:15	Atlas -1.715	Instruction	Work on Assignment 2
	26-10-2025 / 20:00			Deadline for Assignment 2
	05-11-2025 / 13:30 - 16:30	To be announced	Exam	
	28-01-2026 / 18:00 - 21:00	To be announced	Resit	

Always check CANVAS and OSIRIS for the latest information!

17

TU/e

Course material (literature and tools)

- Material provided by lecturers, e.g. slides, handouts, etc.
- Scientific papers
- Industry white papers
- Self-study (video) tutorials (recommended)
- Python exercises

18

TU/e

Assessment

Components:

Assignment 1 – 20%

- Deadline: October 5th, 2025

Assignment 2 – 20%

- Deadline: October 26th, 2025

Written exam – 50%

Weekly quizzes – 10% (announcements through Canvas)

Assignments will be made in groups of 3.
Register through Canvas as soon as possible

It is not possible to re-sit assignments
Assignments are valid only in the current academic year

A minimum grade of 5.0 for the written exam is needed to pass the course.

19

TU/e

Exam

Type: written, on paper, closed book

Date: 5 November 2025, **13.30 – 16.30**

Re-sit: 28 January 2026, **18.00 – 21.00**

20

TU/e

Code of Conduct

Students are bound by TU/e Code of Conduct. When you submit your work under your own name you are asserting ownership of that work. When using ideas of another person, you must give that person appropriate credit through referencing. Referencing serves multiple purposes: (i) it allows readers to further explore sources you have consulted, (ii) it shows the depth of your own thinking and process of inquiry, (iii) it allows you and your readers to compare and contrast your position with other people's positions, agreeing with some, disagreeing with others, and (iv) it gives proper credit to the hard work that many people have done before you. Make sure you avoid any appearance of plagiarism in your work.

21

TU/e

Policy on the use of AI tools

For the assignment, you may use tools such as ChatGPT to review completed drafts for editing, i.e.

- check narration or subtitles in videos/audio for grammar, clarity, and flow;
- review instructions, labels, or UI text in prototypes for clarity in the prototypes;
- check code comments, variable names, or textual descriptions in scripts, programming, or visualizations for readability and overall coherence.

It is not allowed to have AI generate complete texts, tables, or graphs instead of making them yourself. AI tools may only be used for supportive purposes.

- **Transparency**: document clearly for each tool what you used it for in the technology statement. **You must use the Technology statement student(s) (see text below).** Incomplete transparency or presenting AI output as your own work is not permitted.
- **Data confidentiality**: never enter sensitive or confidential data into AI platforms, as these may not comply with the university's or external organizations' privacy guidelines.

Improper or non-transparent use of AI within the assignment will lead to sanctions and may result in your grade being declared invalid by the Examination Board.

22

TU/e

Technology statement student(s)

For the assignment, you must add the technology statement, using the text below. Replace the text in capital letters with the requested information.

During the preparation of this work, I/We used [NAME TOOL / SERVICE / VERSION OF AI TOOL] in order to [REASON]. The following parts of the assignment were affected/generated by AI tool usage: [INTRODUCTION / METHODS / xxx, DISCUSSION]. After using this tool/service, [NAME STUDENT(S)] evaluated the validity of the tool's outputs, including the sources that generative AI tools have used, and edited the content as needed. As a consequence, [NAME STUDENT (S)] take(s) full responsibility for the content of their work.

23

TU/e

Position in curriculum

Topic: data science

Course name: Data analysis and learning methods

Credits: 5 ECTS

Core course in Master program Artificial Intelligence & Engineering Systems

- **Year 1, Q1**

Input towards multiple courses (core and elective)

24

TU/e

Preliminaries

25

Two MSc programs

Data Science and Artificial Intelligence

Artificial Intelligence and Engineering Systems

26

TU/e

Types of data collection

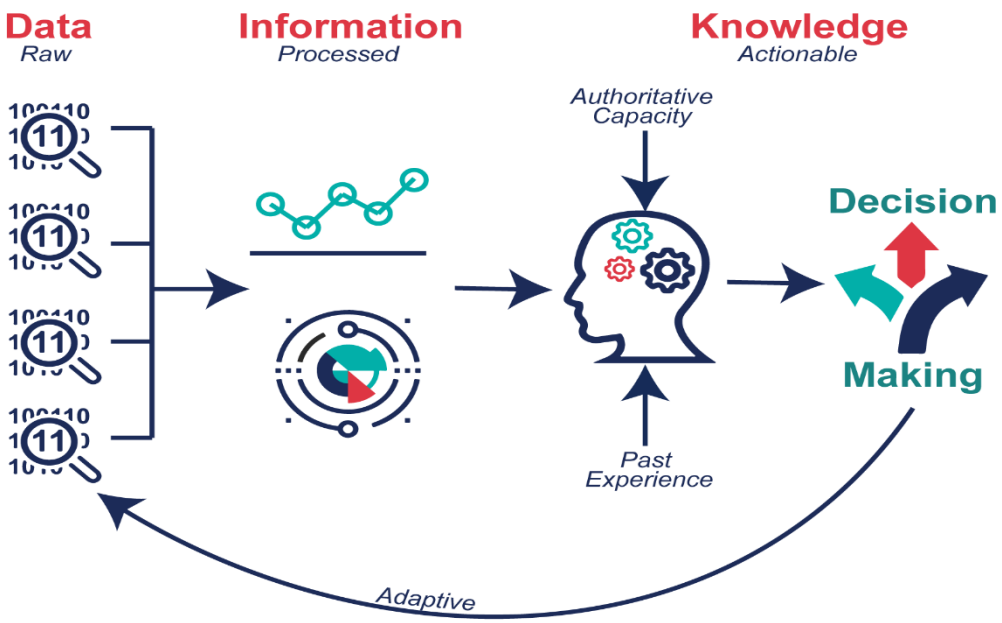
Primary data collection – data collected for the first time, for a particular purpose

Secondary data collection – re-use of data that has already been collected for another purpose

27

TU/e

From raw data to wisdom



Data collection influenced by:

- task
- perspective
- goal

Based on Ackoff, R. L., "From Data to Wisdom", Journal of Applied Systems Analysis, Volume 16, 1989 p 3-9.

Data

- **What is a datum?**
- **When is something data?**
- **What do you need to make sense from the data?**
- **Components to record?**

➔ Data models

29

TU/e

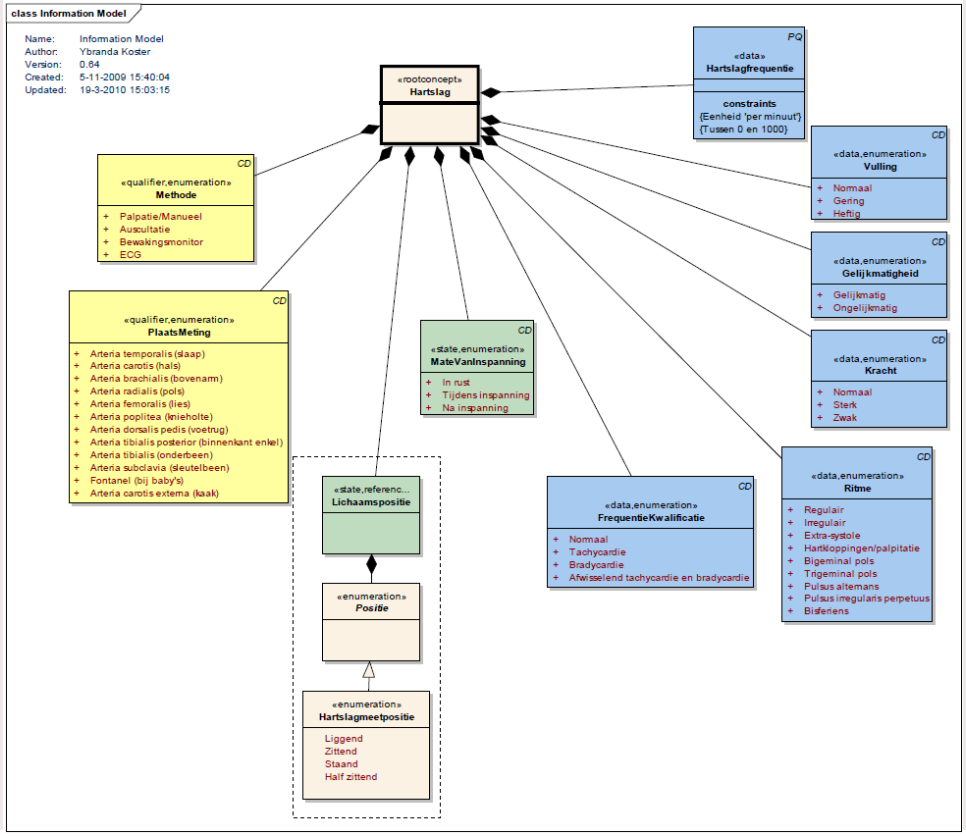
Data model

An abstract model that organizes elements of data and standardizes how they relate to one another and to the properties of real-world entities

Aid communication between domain experts and technical experts

Show the data needed and created by business processes

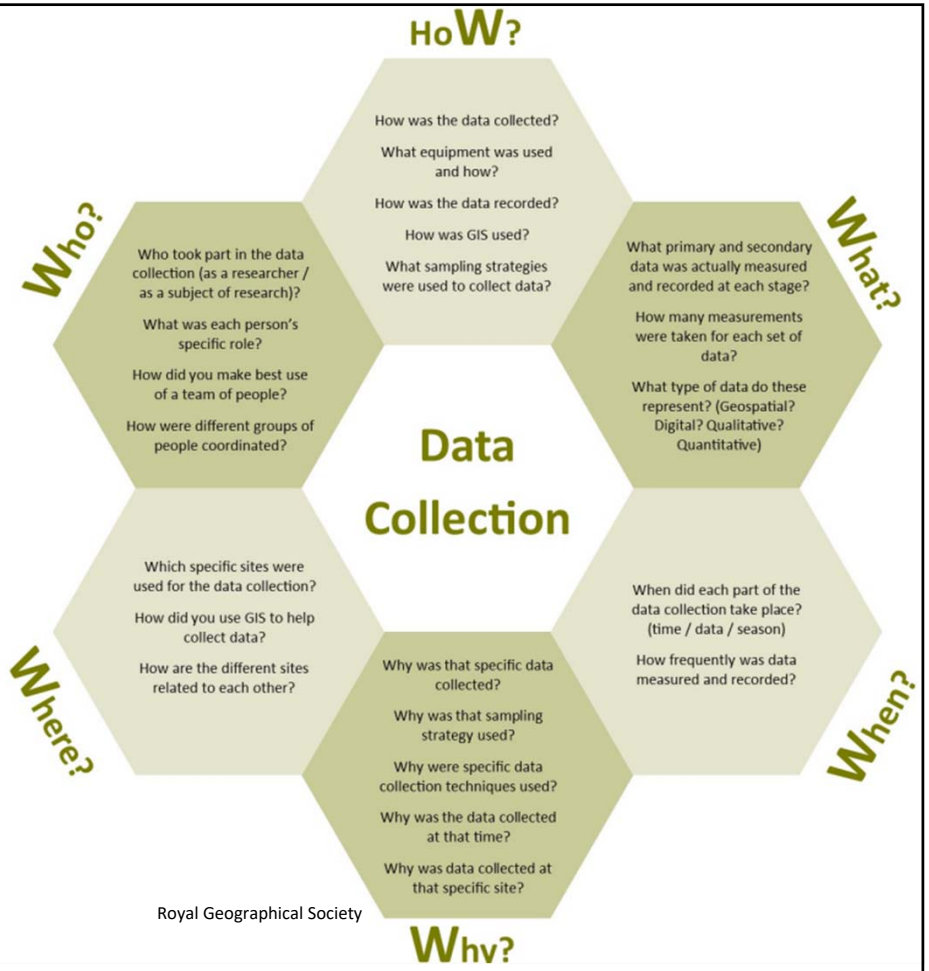
Example: data model for measuring heart rate



Information elements:
- Heart rate frequency
- Heart condition qualification
- Heart rhythm
- Method of measurement
- Place of measurement
- Body posture
- …

Aspects of data collection

Illustration: iStock



Royal Geographical Society

34

TU/e

Machine learning basics

Some slides by G. Manias and D. Di Nucci

35

What is machine learning?

Oxford definition:

the use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyse and draw inferences from patterns in data

Solved through an optimization problem

[36]

Types of learning

Basic learning types:

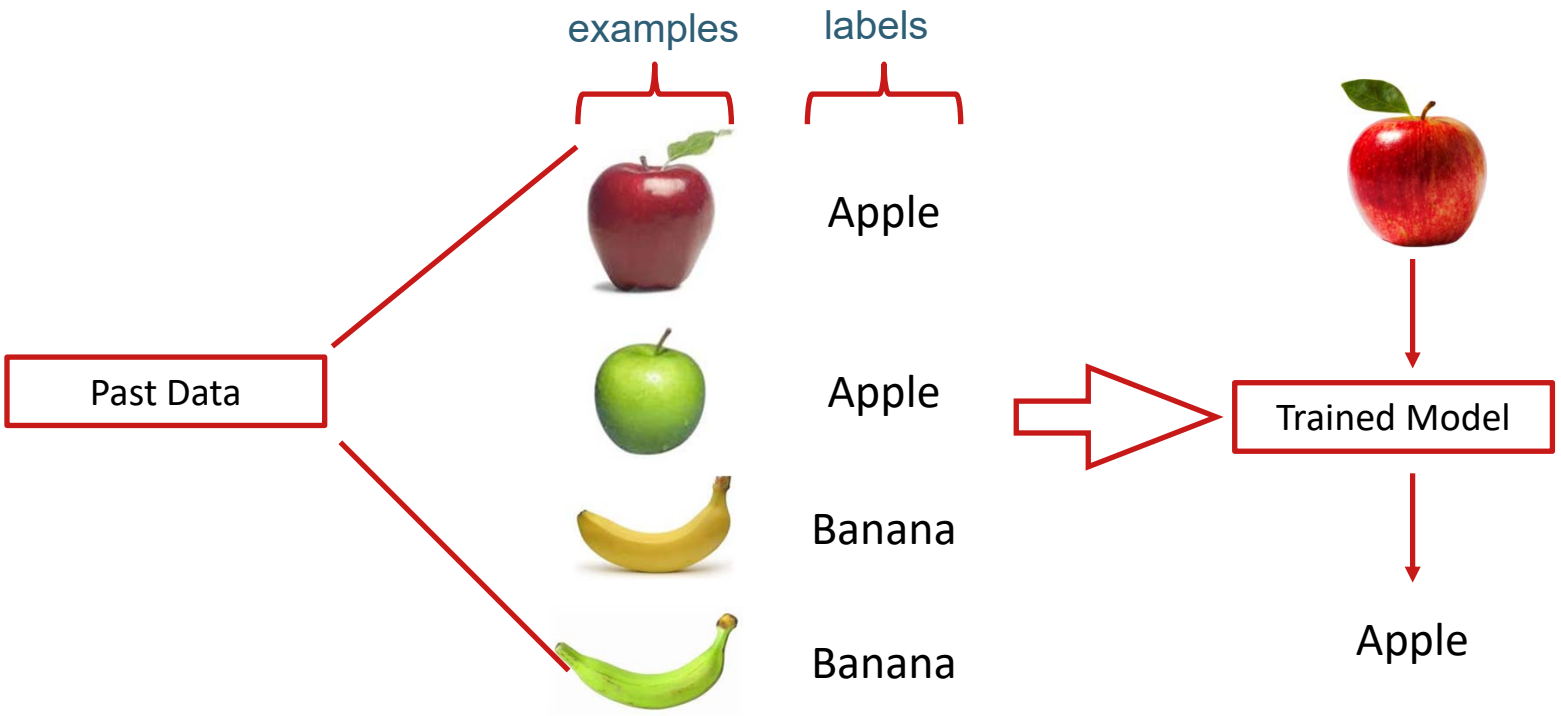
- **Supervised**
- **Unsupervised**
- **Reinforcement learning**

Other learning types:

- **Semi-supervised learning**
- **Transfer learning**
- **Active learning**
- **Etc.**

[37]

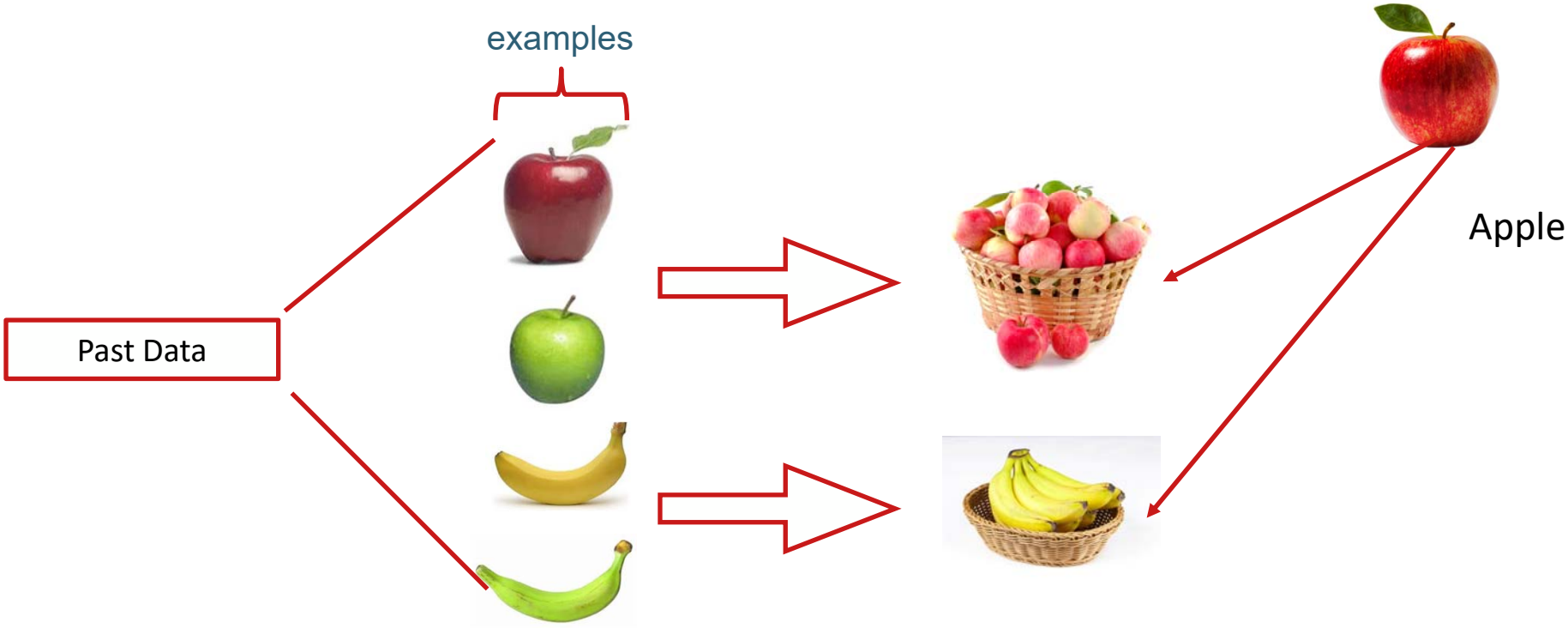
An Example of Supervised Learning



Given a labeled dataset, the model learns to predict new examples.

Slide courtesy of Prof. Dario Di Nucci

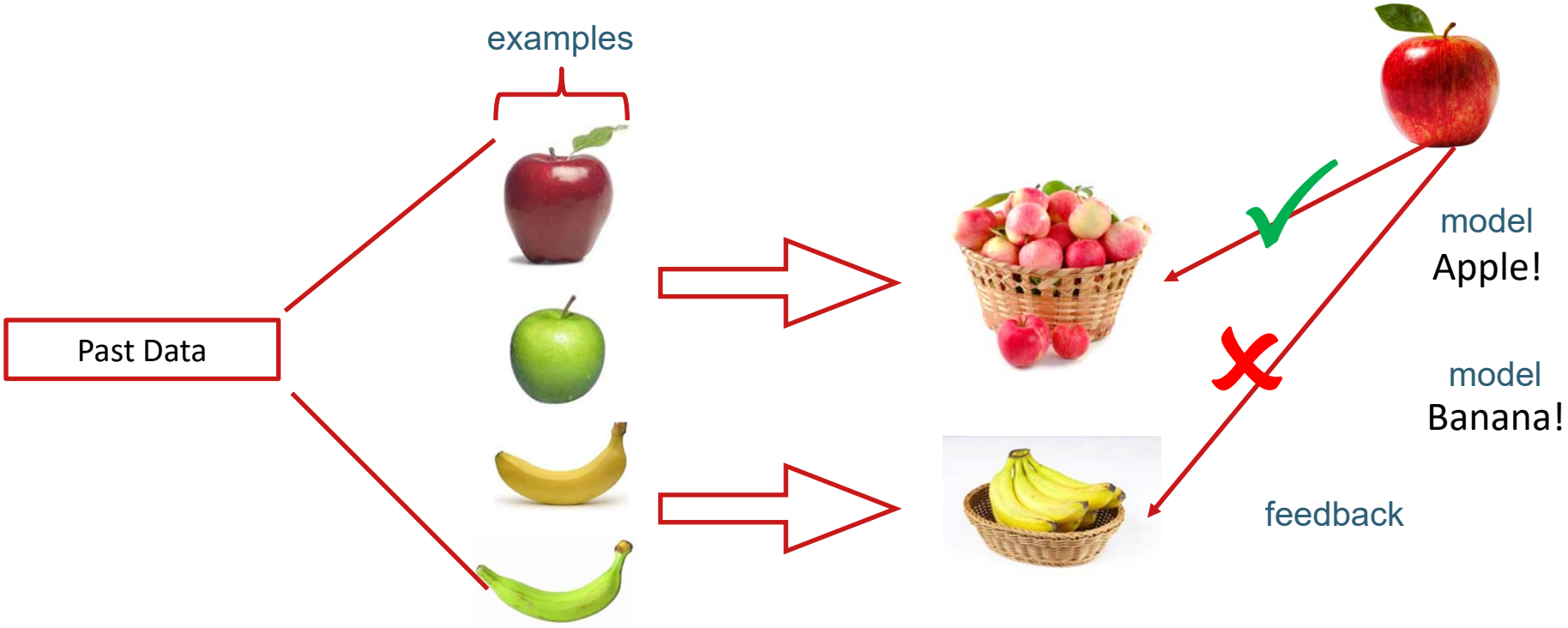
An Example of Unsupervised Learning



Given a dataset, without labels, the model learns to cluster/group similar data.

Slide courtesy of Prof. Dario Di Nucci

An Example of Reinforcement Learning



Given a dataset, the model learns to take correct actions (predictions) from feedback.

Machine learning tasks

- Classification
- Regression
- Similarity matching
- Clustering
- Anomaly detection
- Co-occurrence grouping
- Profiling
- Link prediction
- Data reduction
- Causal modeling

[41]

Common learning methods (models)

Linear regression

Logistic regression

Decision trees

k nearest neighbors classifier

Naïve Bayes classifier

(Nonlinear) input – output mappings

Mixture models

Support vector machines

Neural networks

Fuzzy inference systems

Bayesian networks

[42]

Recap

- **Data are nowadays ubiquitous and have multiple facets**
- **Distinction between primary data collection and secondary data collection**
  - In engineering systems, primary data collection has focus
  - Data science deals also with secondary data collection
- **Learning is related to solving an optimization problem**
- **Three main types of learning**
- **Many different machine learning tasks**