DRIVES

Development and Research on Innovative Vocational Education Skills

U2 MACHINE LEARNING OVERVIEW

U2.E1 WHAT IS MACHINE LEARNING?

Machine Learning Engineer

January 2021, Version 1

Co-funded by the Erasmus+ Programme of the European Union

The Development and Research on Innovative Vocational Educational Skills project (DRIVES) is co-funded by the Erasmus+ Programme of the European Union under the agreement 591988-EPP-1-2017-1-CZ-EPPKA2-SSA-B. The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

LEARNING OBJECTIVES

The student is able to

MLE.U2.E1.PC1	Know and understand the definition of machine learning.
MLE.U2.E1.PC2	Understand the need, purpose and impacts of machine learning.
MLE.U2.E1.PC3	Recognize the connection between artificial intelligence and machine learning.
MLE.U2.E1.PC4	Understand how machine learning fits into artificial intelligence.

WHAT IS MACHINE LEARNING?

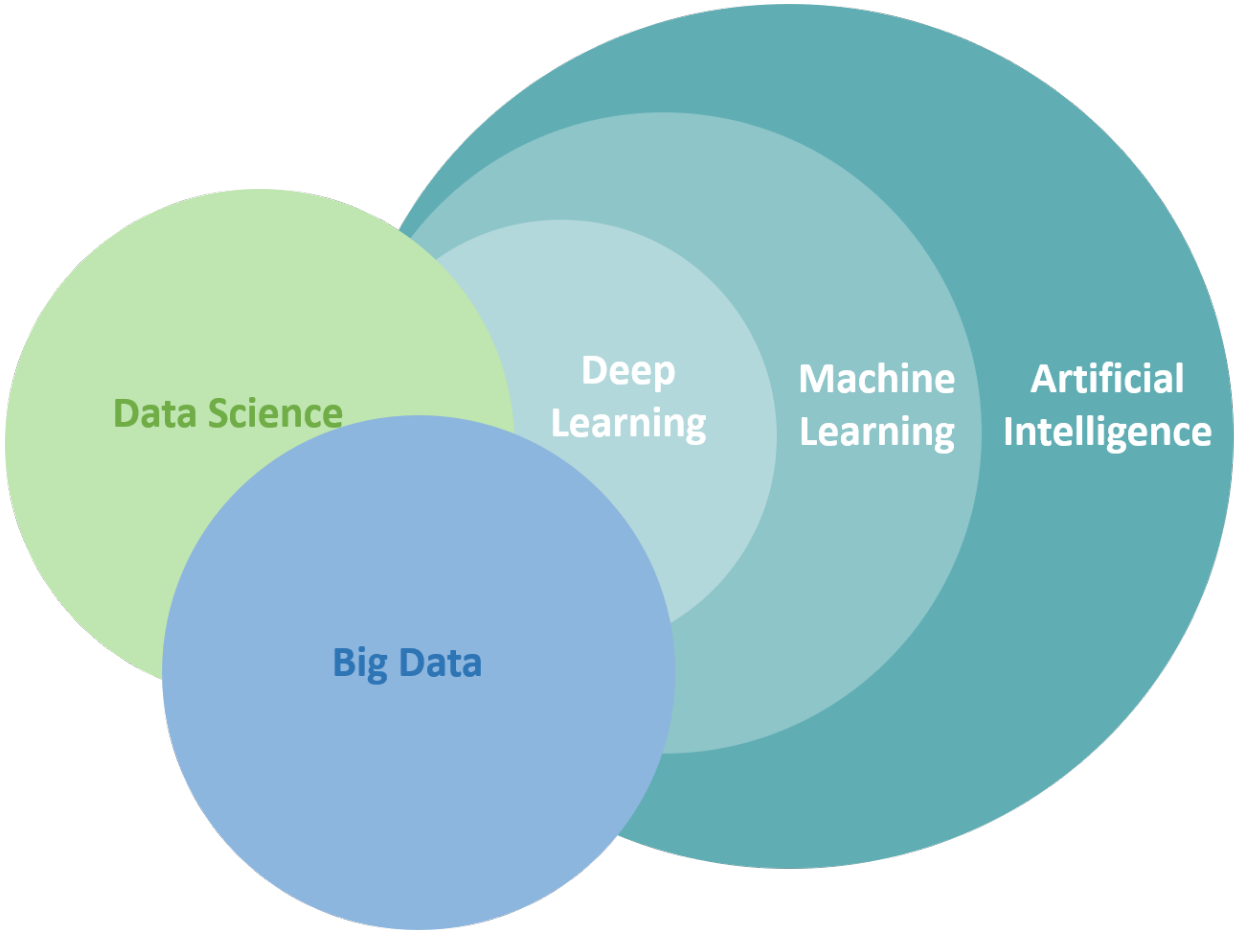
Machine Learning empowers learning systems to act and take data driven decisions to carry out a certain task.

Machine Learning is a subset of Artificial Intelligence that focuses on the development of computer programs (algorithms) that can grant access to data and then use it to learn for themselves.

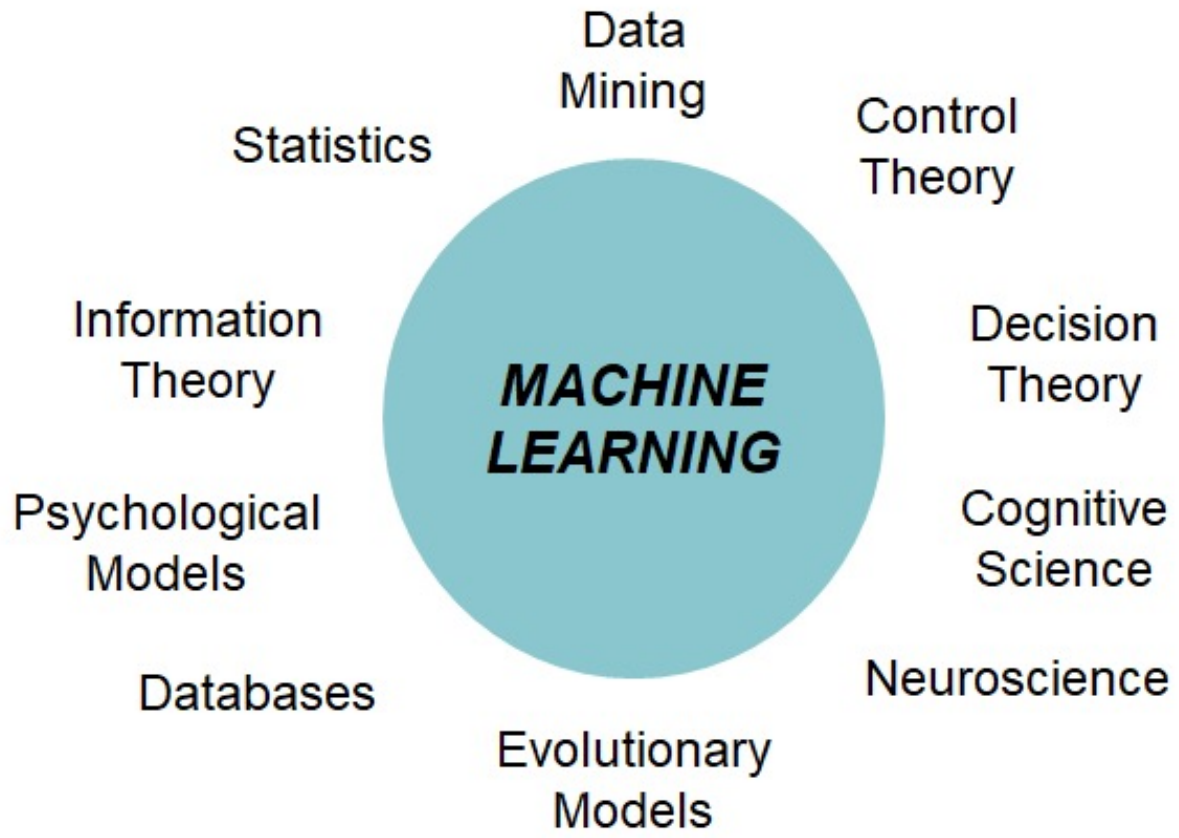
These algorithms can learn and improve over time when exposed to new data, i.e., ML uses statistical methods to enable machines to learn and improve with experience.

Its main goal is to enable computers to learn automatically without human intervention or assistance, as well as, to reduce distance between estimated value and real value (the error).

WHAT IS MACHINE LEARNING?



WHAT IS MACHINE LEARNING?



ARTIFICIAL INTELLIGENCE VS MACHINE LEARNING

AI is defined as acquisition of knowledge intelligence	ML is defined as the acquisition of knowledge or skill
The aim is to increase the chance of success	The aim is to increase accuracy
It works as a computer program that does smart work	It is a simple machine that takes data and learns from it
The goal is to simulate natural intelligence to solve complex problems	The goal is to learn from data a specific task to maximize machine performance on this task

ARTIFICIAL INTELLIGENCE VS MACHINE LEARNING

DRIVES
Development and Research on Innovative Vocational Education Skills

ARTIFICIAL INTELLIGENCE	MACHINE LEARNING
☑ AI is decision making	☑ ML enables systems to learn new things from data
☑ It leads to the development of a system that mimics human behavior to respond to circumstances	☑ It is involved in the creation of self-learning algorithms
☑ AI pursues the optimal solution	☑ ML will only come up with a solution that is either optimal or not
☑ Leads to intelligence or wisdom	☑ Leads to knowledge

WHY DO WE NEED MACHINE LEARNING?

The work that we need to do is increasing day-to-day

Data-based decisions increasingly make a difference between

Saves the manpower of the organization and also increases the productivity

WHY DO WE NEED MACHINE LEARNING?

DRIVES
Development and Research on Innovative Vocational Education Skills

- **No human experts**
  - industrial/manufacturing control
  - mass spectrometer analysis, drug design, astronomic discovery, etc.

- **Black-box human expertise**
  - face/handwriting/speech recognition
  - emotion and activity recognition
  - driving a car, flying a plane

WHY DO WE NEED MACHINE LEARNING?

DRIVES
Development and Research on Innovative Vocational Education Skills

- **Rapidly changing phenomena**
  - credit scoring, financial modeling
  - medical diagnosis, fraud detection
- **Need for customization/personalization**
  - personalized feeds, news reader, etc.
  - products/services recommendation

WHY DO WE NEED MACHINE LEARNING?

input data → analyze data → find patterns → prediction → improvement

WHAT KIND OF PROBLEMS CAN BE TACKLED USING MACHINE LEARNING?

DRIVES
Development and Research on Innovative Vocational Education Skills

- **Text or document classification,** which includes problems such as assigning a topic to a text or a document, spam detection, and determining automatically if the content of a web page is inappropriate or too explicit.

- **Natural language processing (NLP)**, which includes part-of-speech tagging, named-entity recognition, context-free parsing, or dependency parsing.

- **Computer vision applications,** which includes object recognition, object identification, face detection, Optical Character Recognition (OCR), content-based image retrieval, or pose estimation.

WHAT KIND OF PROBLEMS CAN BE TACKLED USING MACHINE LEARNING?

DRIVES
Development and Research on Innovative Vocational Education Skills

- **Speech processing applications**, which includes speech recognition, speech synthesis, speaker verification, speaker identification, as well as sub-problems such as language modeling and acoustic modeling.

- **Computational biology applications**, which includes protein function prediction, identification of key sites, or the analysis of gene and protein networks.

- Other applications such as fraud detection, network intrusion, learning to play games, unassisted vehicle control such as robots or cars, medical diagnosis, design of recommendation systems, search engines, or information extraction systems.

WHAT KIND OF PROBLEMS CAN BE TACKLED USING MACHINE LEARNING?

DRIVES
Development and Research on Innovative Vocational Education Skills

This list is by no means comprehensive!

Most prediction problems found in practice can be cast as learning problems and the practical application area of machine learning keeps expanding.

GROWTH OF MACHINE LEARNING

- **This trend is accelerating:**

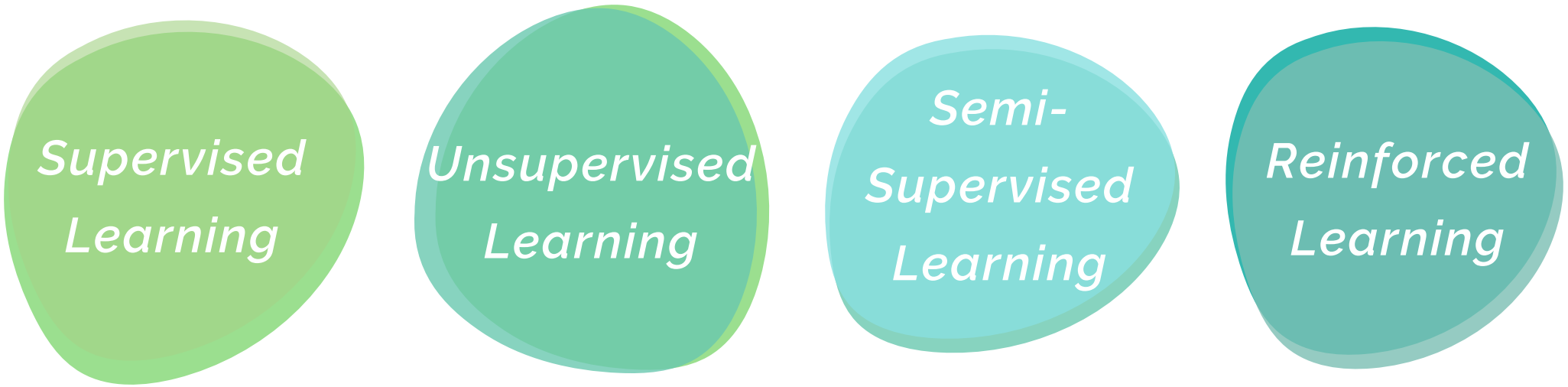
  - Improved machine learning algorithms;
  - Improved data capture, networking and faster computers;
  - Software too complex to write by hand;
  - New sensors / IO devices;
  - Demand for self-customization to user/environment;
  - It turns out to be difficult to extract knowledge from human experts→*failure of expert systems in the 1980's.*

DIMENSIONS OF LEARNING SYSTEMS

DRIVES
Development and Research on Innovative Vocational Education Skills

- **Type of Feedback**
  - supervised (labeled examples)
  - unsupervised (unlabeled examples)
  - semi-supervised (labeled and unlabeled examples)
  - reinforcement (reward)
- **Representation**
  - attribute-based (feature vector)
  - relational (first-order logic)
- **Use of Knowledge**
  - empirical (knowledge-free)
  - analytical (knowledge-guided)

TYPES OF MACHINE LEARNING MODELS



TYPES OF MACHINE LEARNING MODELS

- **Supervised (inductive) learning**

  Training data includes desired outputs

- **Unsupervised learning**

  Training data does not include desired outputs

- **Semi-supervised learning**

  Training data includes a few desired outputs

- **Reinforcement learning**

  Rewards from sequence of actions

COMPONENTS OF A MACHINE LEARNING SOLUTION

DRIVES
Development and Research on Innovative Vocational Education Skills

Examples: Items or instances of data used for learning and testing.

Features: The set of attributes also referred to as variables, often represented as a vector, associated to an example. In datasets, features appear as columns.

Labels: Values or categories assigned to examples. In classification problems, examples are assigned specific categories, while in regression, items are assigned real-valued labels.

Hyperparameters (Fitting Parameters): Free parameters that are not determined by the learning algorithm, but rather specified as inputs to the learning algorithm. The optimization of these parameters can be made through greedy search, gradient descent, linear programming and many variations.

COMPONENTS OF A MACHINE LEARNING SOLUTION

DRIVES
Development and Research on Innovative Vocational Education Skills

Training sample: Examples used to train a learning algorithm.

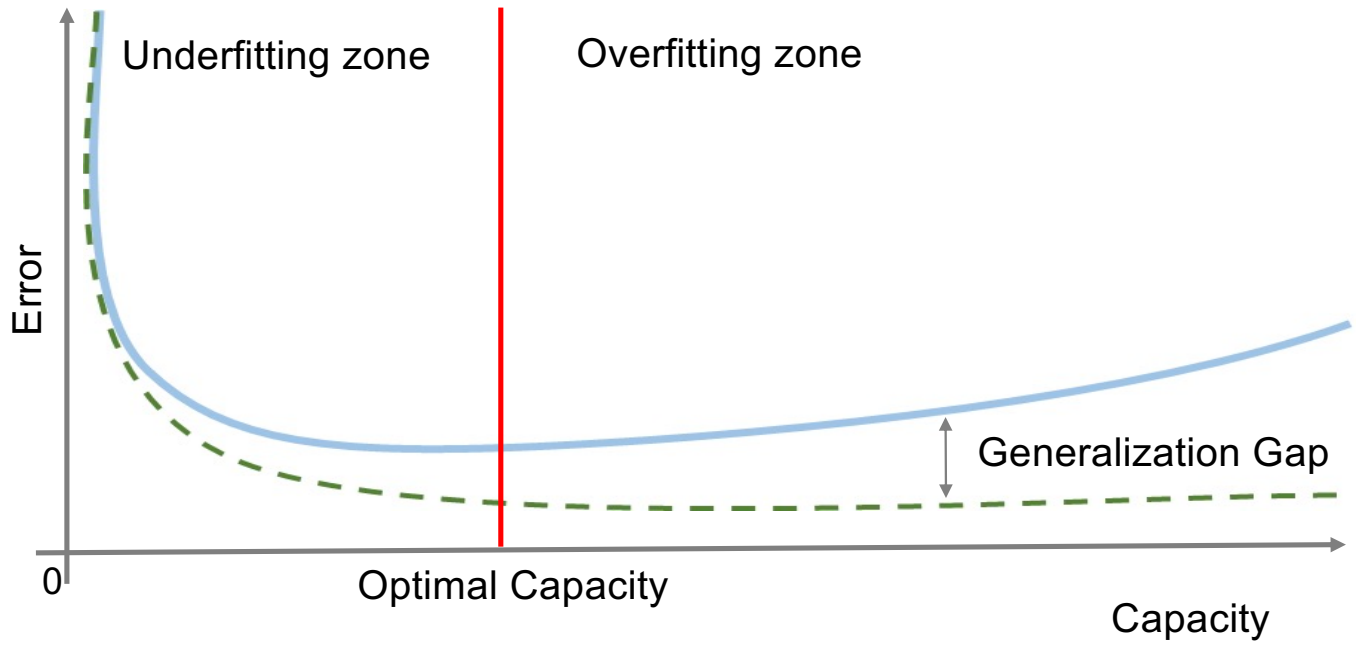
Validation sample: Examples used to tune the parameters of a learning algorithm when working with labeled data. The validation sample is used to select appropriate values for the hyperparameters.

Test sample: Examples used to evaluate the performance of a learning algorithm. The test sample is separate from the training and validation data and is not made available in the learning stage.

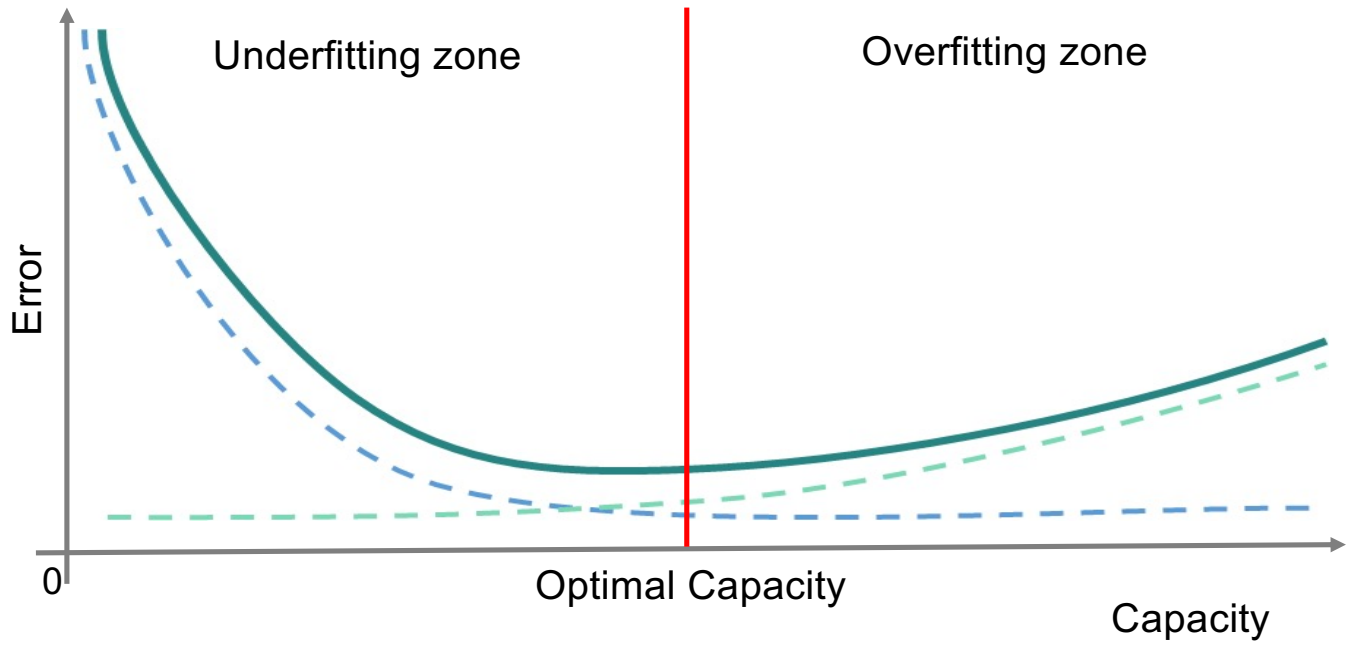
Loss Function: Evaluation of the models through metrics such as acuracy, precision, and squared error.

Deployment/Experimentation Cycle: Dealing with mistakes, adaption over time and maintaining balance.

Generalization and Capacity



Bias and Variance



MACHINE LEARNING AND DATA MINING

Both Data Mining (DM) and Machine Learning (ML) are rooted in Data Science. The limits between the two concepts are often blurred, but there are a few differences between them.

ML is the process of discovering algorithms that can learn from and make predictions on the data. It is the design, study and development of algorithms that allow machines to learn without human intervention.

DM is a process for extracting useful information from a large amount of data. It is used to discover new, reliable and useful patterns in the data, to find meaning and information relevant to the company or to the person who needs it. It is used by humans.

MACHINE LEARNING AND DATA MINING

DRIVES
Development and Research on Innovative Vocational Education Skills

These concepts are often intersected or confused and individuals erroneously use these two terms interchangeably.

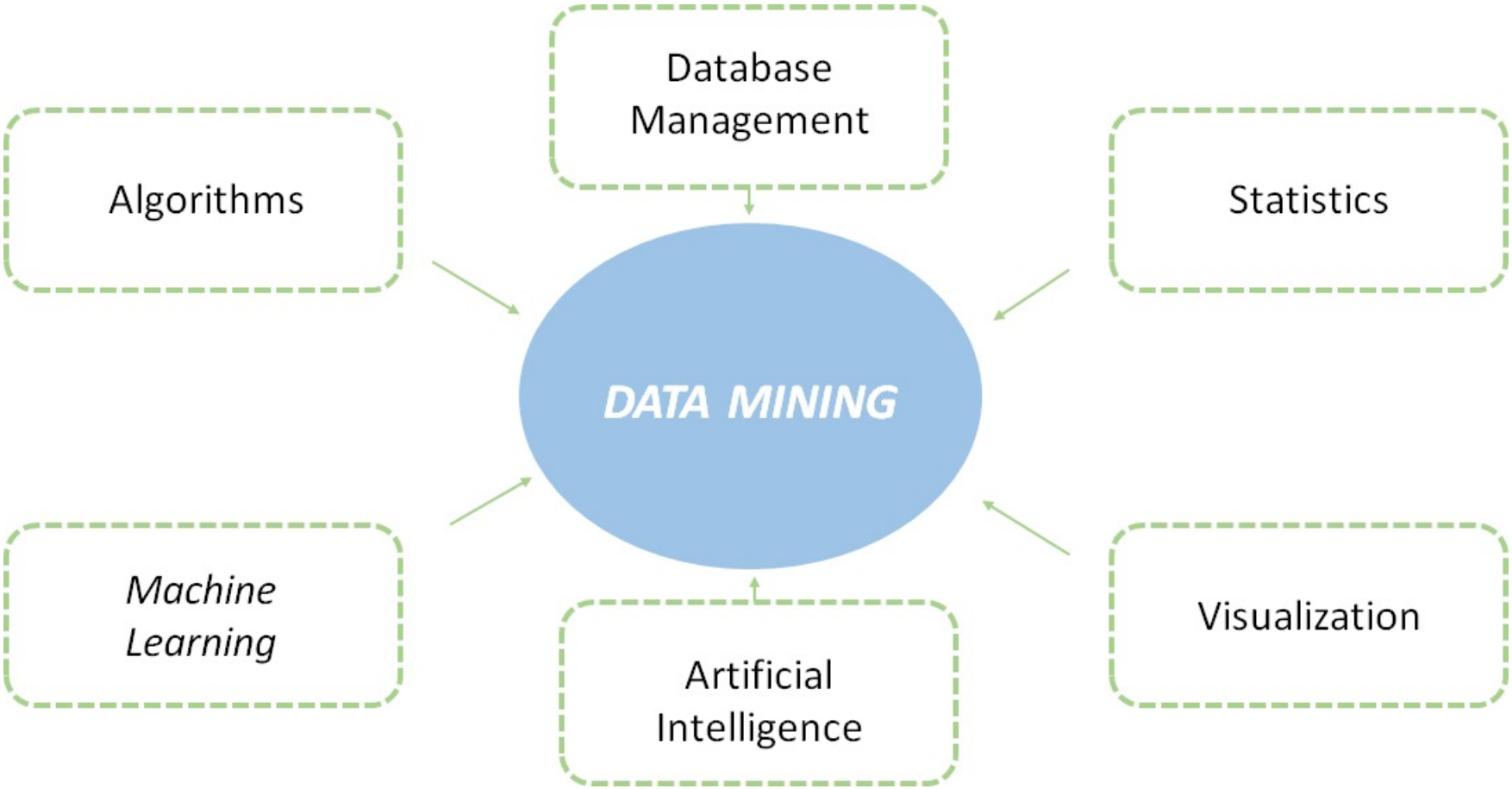
WHY DO HUMANS CONFUSE THESE TERMS?

Machine Learning and Data Mining are both analytics processes.

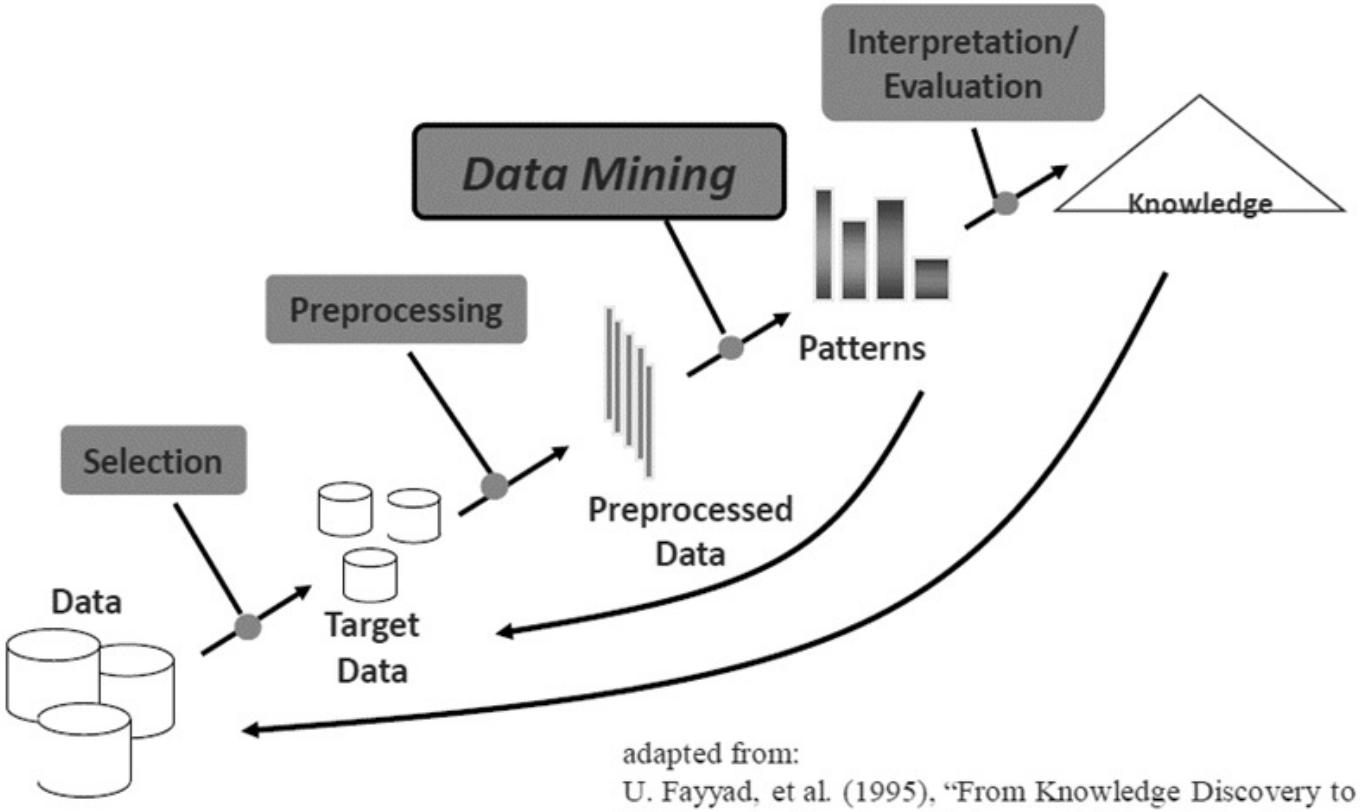
Machine Learning is sometimes used as a means for carrying out useful Data Mining activities.

Both aim to learn from data in order to improve decision-making.

MACHINE LEARNING AND DATA MINING



MACHINE LEARNING AND DATA MINING



adapted from:
U. Fayyad, et al. (1995), "From Knowledge Discovery to Data Mining: An Overview," Advances in Knowledge Discovery and Data Mining, U. Fayyad et al. (Eds.), AAAI/MIT Press

DRIVES
Development and Research on Innovative Vocational Education Skills

DIFFERENCES

DATE OF INVENTION

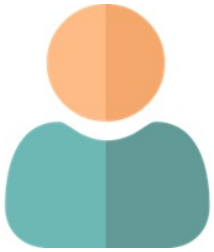
DM predates ML by two decades.

PURPOSE

DM is designed to extract rules from a large amount of data, while ML teaches a computer how to learn and understand information to perform complex tasks.

DRIVES
Development and Research on Innovative Vocational Education Skills

DIFFERENCES



HUMAN FACTOR

DM relies on human intervention and is ultimately designed for human use. Whereas ML teaches itself and not depends on human influence or actions.

GROWTH ABILITY

DM cannot learn or adapt as it follows pre-set rules and is static by nature, while ML adjusts algorithms as circumstances occur.

DRIVES
Development and Research on Innovative Vocational Education Skills

DIFFERENCES

USE OF DATA

Data mining is used on an existing dataset, such as a data warehouse, to discover patterns. Machine learning, on the other hand, is trained on a training data set, which teaches the computer how to make sense of the data, and then makes predictions about new data sets.

SUMMARY
PRACTICE RECOMMENDATIONS

- Machine Learning is a branch of Artificial Intelligence;
- Machine Learning is based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention;
- There are four types of Machine Learning models, namely Supervised, Unsupervised, Semi-supervised and Reinforcement Learning models.
- Machine learning and Data mining are rooted in Data Science but are different;
- Machine Learning is sometimes used as a means for carrying out useful Data Mining activities.

REFERENCES

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

Zhang, X. D. (2020). Machine learning. In *A Matrix Algebra Approach to Artificial Intelligence* (pp. 223-440). Springer, Singapore. https://doi.org/10.1007/978-981-15-2770-8_6.

Sammut, C., & Webb, G. I. (Eds.). (2011). *Encyclopedia of machine learning*. Springer Science & Business Media.

Bonaccorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.

Langley, P. (1996). *Elements of machine learning*. Morgan Kaufmann.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Machine learning basics. *Deep learning*, *1*, 98-164.

REFERENCE TO AUTHORS

Diana Ferreira

- PhD student in Biomedical Engineering
- Research Collaborator of the Algoritmi Research Center

0000-0003-2326-2153

Regina Sousa

- PhD student in Biomedical Engineering
- Research Collaborator of the Algoritmi Research Center

0000-0002-2988-196X

José Machado

- Associate Professor with Habilitation at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0003-4121-6169

REFERENCE TO AUTHORS

António Abelha

- Assistant Professor at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0001-6457-0756

Victor Alves

- Assistant Professor at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0003-1819-7051

REFERENCE TO AUTHORS

This Training Material has been certified according to the rules of **ECQA – European Certification and Qualification Association.**

The Training Material was developed within the international job role committee **"Machine Learning Engineer":**

UMINHO – University of Minho (https://www.uminho.pt/PT)

The development of the training material was partly funded by the EU under Blueprint Project DRIVES.

DRIVES
Development and Research on Innovative Vocational Education Skills

Thank you for your attention

DRIVES project is project under **The Blueprint for Sectoral Cooperation on Skills in Automotive Sector**, as part of New Skills Agenda.

The aim of the Blueprint is **to support an overall sectoral strategy and to develop concrete actions to address short and medium term skills needs.**

Follow DRIVES project at:

More information at:

www.project-drives.eu

Co-funded by the Erasmus+ Programme of the European Union

The Development and Research on Innovative Vocational Educational Skills project (DRIVES) is co-funded by the Erasmus+ Programme of the European Union under the agreement 591988-EPP-1-2017-1-CZ-EPPKA2-SSA-B. The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.