UNIVERSIDAD DE CHILE

Minería de Datos

Welcome to the Machine Learning class

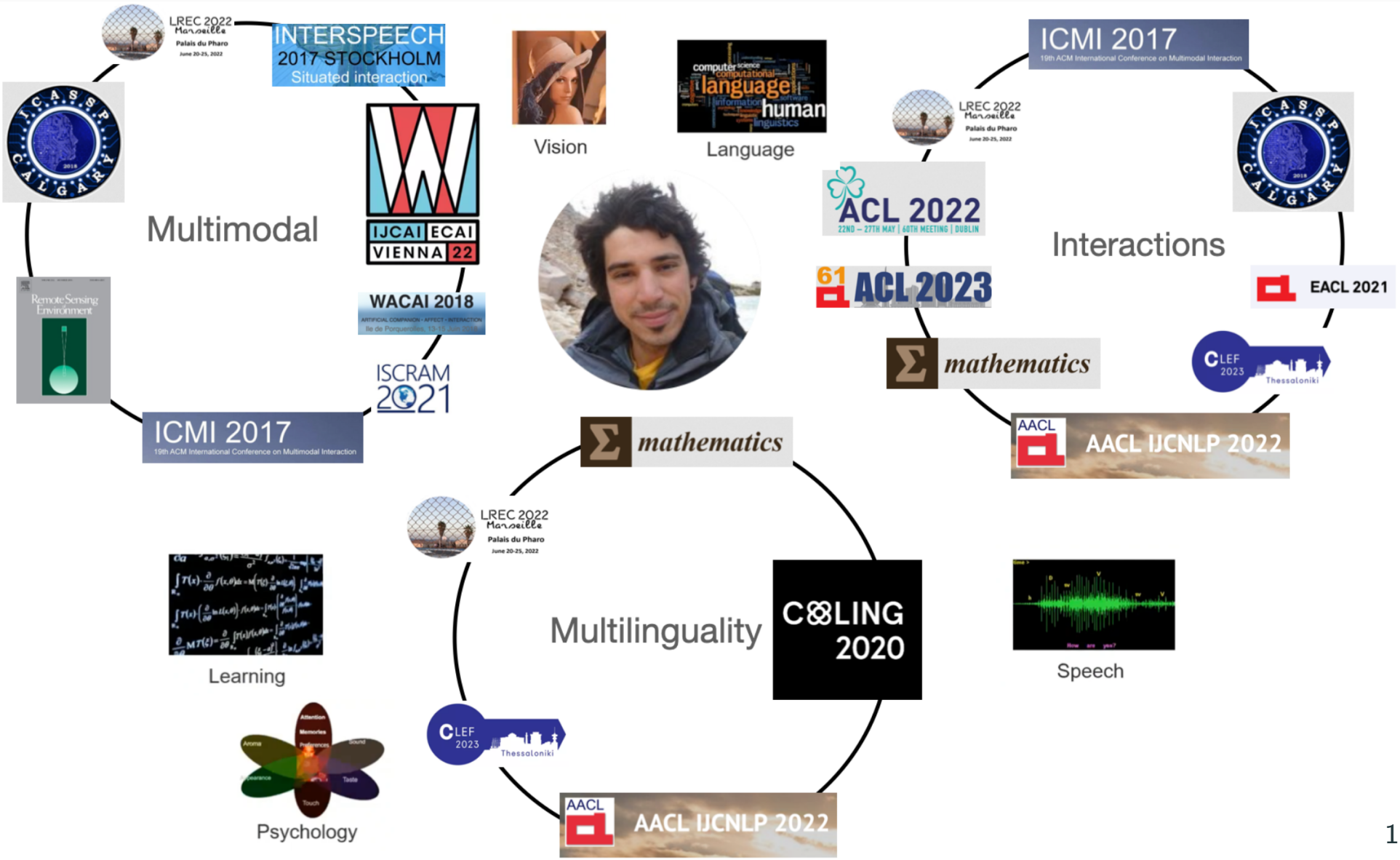
Valentin Barriere

Universidad de Chile – DCC

CC5205, Fall 2025

Introduction

Who am I?


LREC 2022 Marseille Palais du Pharo June 20-25, 2022
INTERSPEECH 2017 STOCKHOLM Situated interaction
ICASSP CALGARY
Multimodal
IJCAI ECAI VIENNA 22
WACAI 2018
ISCRAM 2021
Remote Sensing Environment
ICMI 2017 19th ACM International Conference on Multimodal Interaction
Vision
Language
computer science computational language human linguistics
ACL 2022 22ND – 27TH MAY | 60TH MEETING | DUBLIN
61 ACL 2023
Interactions
EACL 2021
mathematics
CLEF 2023 Thessaloniki
AACL IJCNLP 2022
Multilinguality
COLING 2020
Learning
Speech
Psychology
Attention Memories Preferences Sound Aroma Appearance Taste Touch

Data Mining

Definition

Data mining aims at the extraction of knowledge from large amounts of data using automatic or semi-automatic methods.
It proposes to use a set of algorithms [...] to build models from the data, that is, to find interesting structures or patterns according to predefined criteria, and extract as much knowledge as possible from them.[1]

[1] https://fr.wikipedia.org/wiki/Exploration_de_données

Data

Different types of data:

- Structured data:
  - Social data: Age, Salary, Skin color, Place of residence
  - Metric data: *Likes* on a post, Time spent on a page, Number of mutual friends
- Unstructured data:
  - Text: Sentence, Paragraph, Document
  - Sound: Song, Speech
  - Image: Photo, Video

Mining

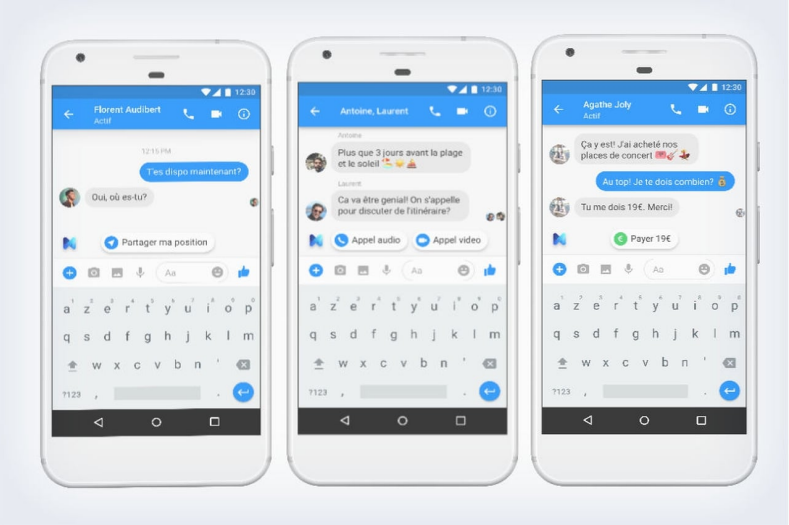
Different types of Mining:

- Data exploration: Detect simple values, biases
- Classification/Regression task: Use data to characterize new data by class or with a value in a supervised manner
- Clustering task: Group data into classes in an unsupervised manner
- Dimensionality reduction: Develop common structures for compressed data representations

Outline : Applications

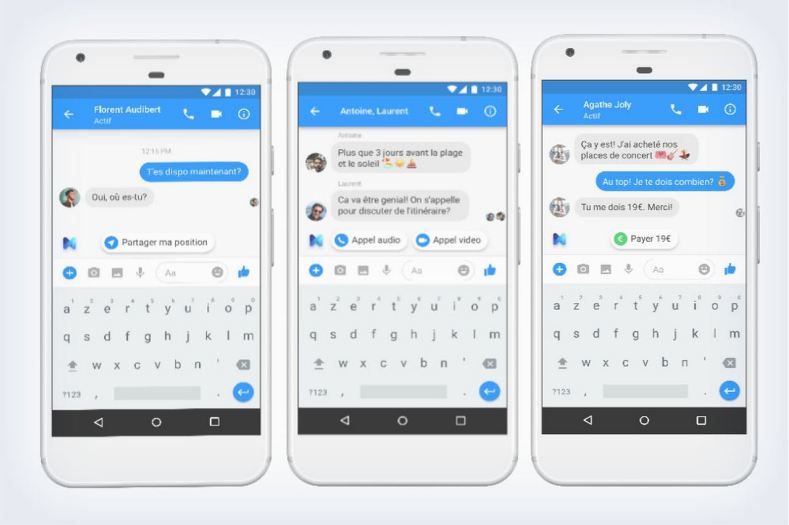
Applications (I/II)

- Event detection in text



Applications (I/II)

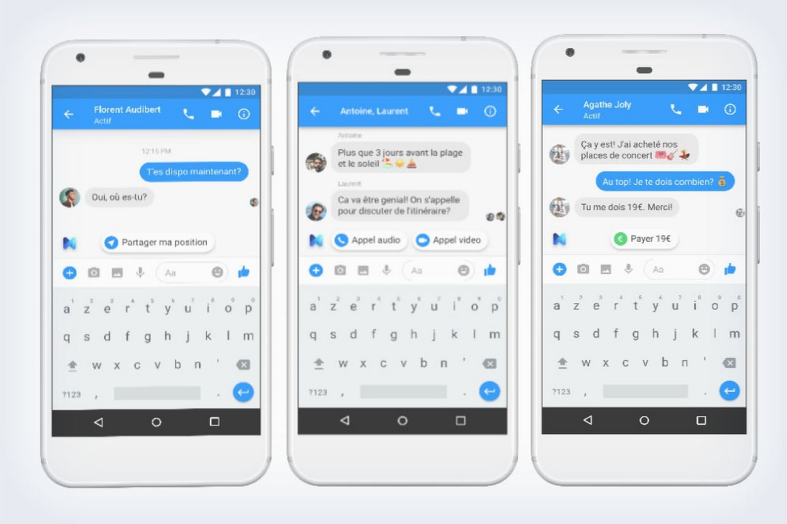
- Event detection in text



- Automatic processing of user opinions

Applications (I/II)

- Event detection in text

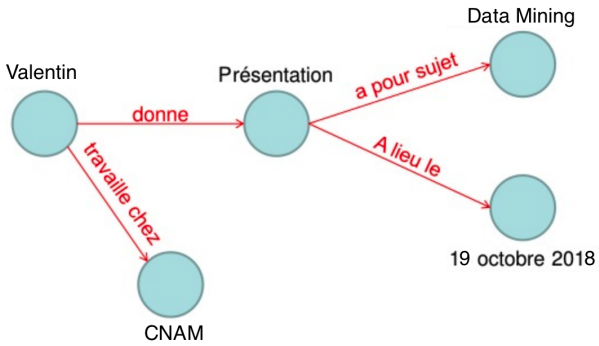


- Automatic processing of user opinions

- Recommending items to a user

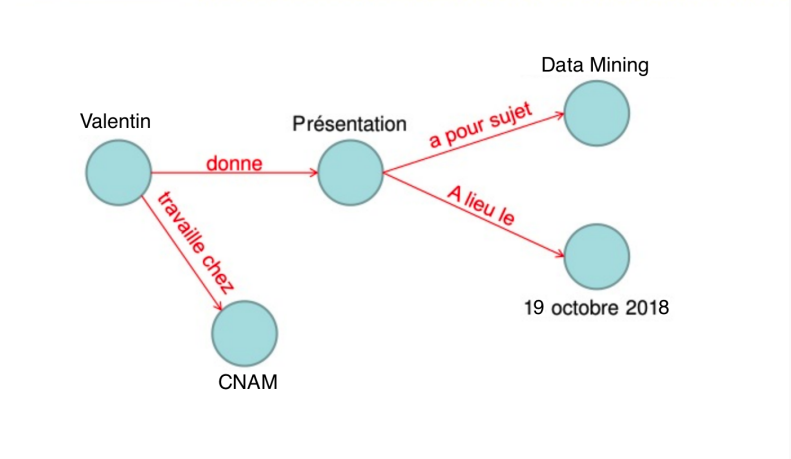
Applications (II/II)

- Detection of relationships between entities in a text

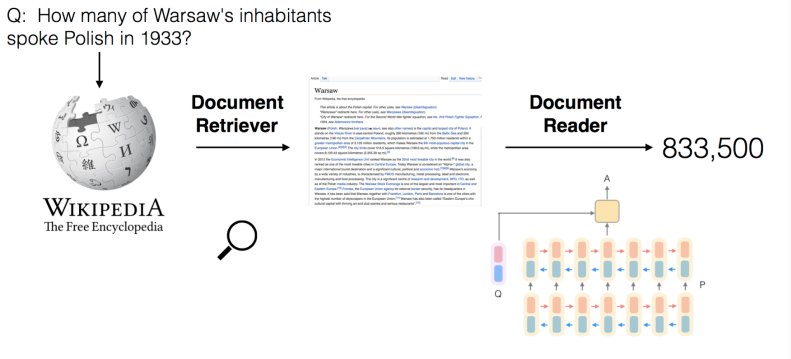


Applications (II/II)

- Detection of relationships between entities in a text

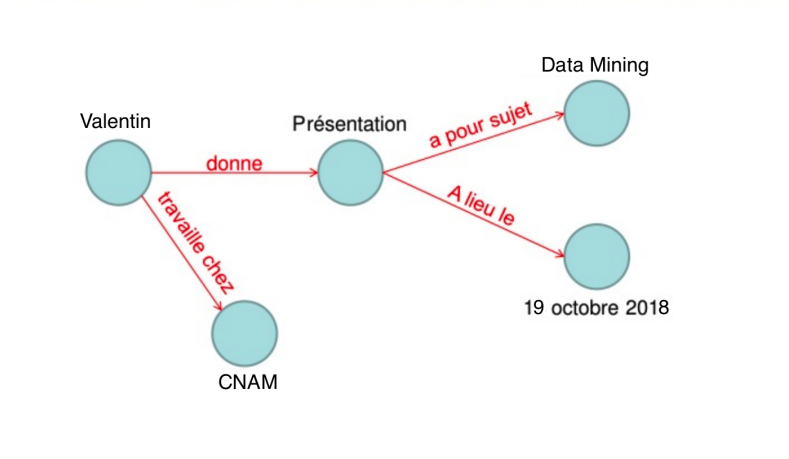


- Question answering

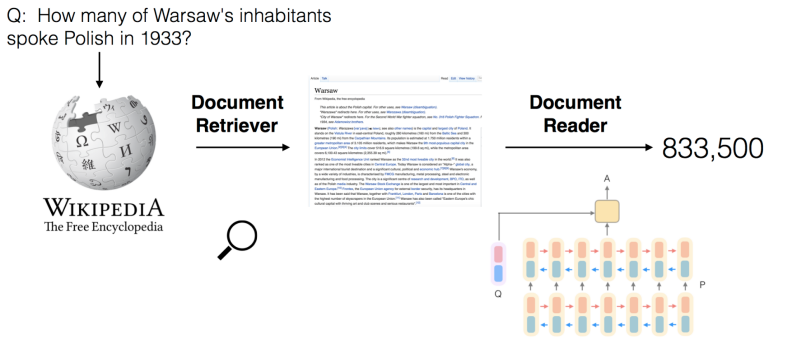


Applications (II/II)

- Detection of relationships between entities in a text



- Question answering



- IE module for a conversational agent

Outline : Meaning of the Terms

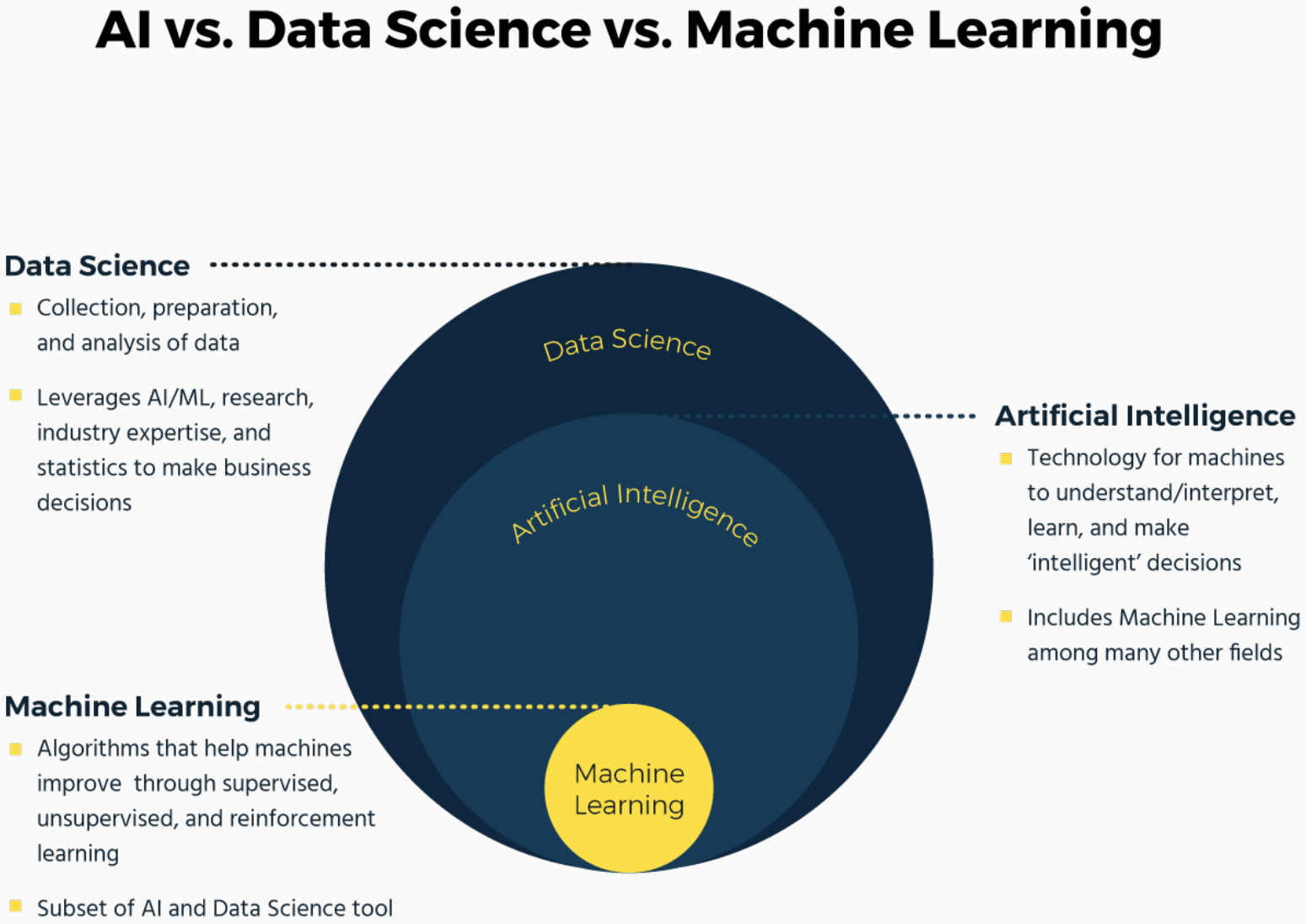


Figure 1: Differences between fields

Data Science

In summary

Data Science focuses on analyzing data to extract knowledge, Machine Learning uses algorithms to make predictions and decisions based on data, and Artificial Intelligence refers to the development of systems that can perform intelligent tasks autonomously.

Overly simplistic definition:

- Data mining generates understanding.
- Machine learning generates predictions.
- Artificial intelligence generates actions.

Example on a Music Platform

Data Scientist

Collects and analyzes user data from music platforms to identify patterns and musical preferences.

Machine Learner

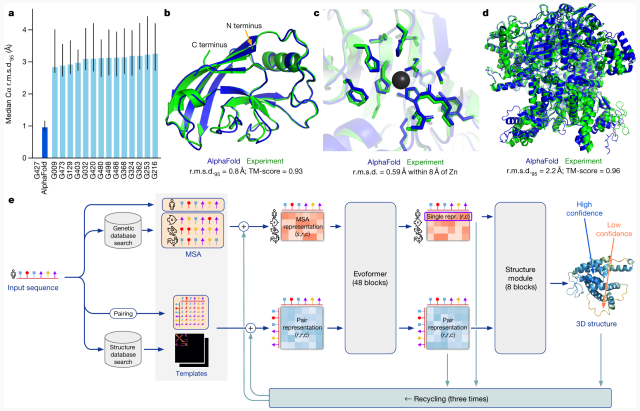
Develops and optimizes a music recommendation model using machine learning algorithms to predict user preferences.

Artificial Intelligence

Implements a social agent that can interact with the user, to enhance the personalization of music recommendations and provide a more accurate and contextualized experience.

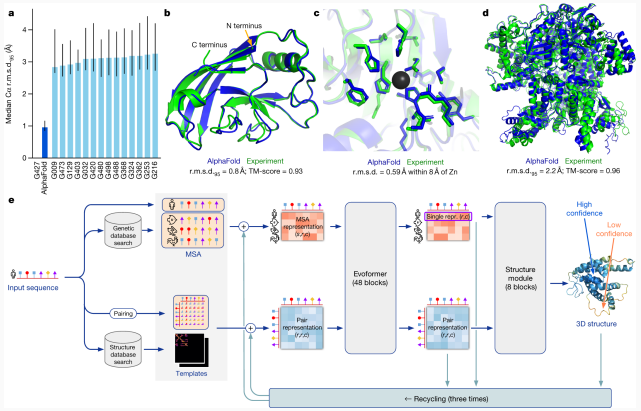
Significance of the Work: Why Do It?

- Scientific advancement



Significance of the Work: Why Do It?

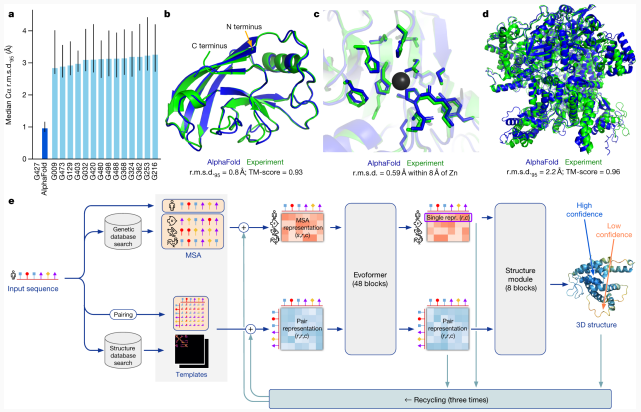
- Scientific advancement



- Prevention and management of natural disasters

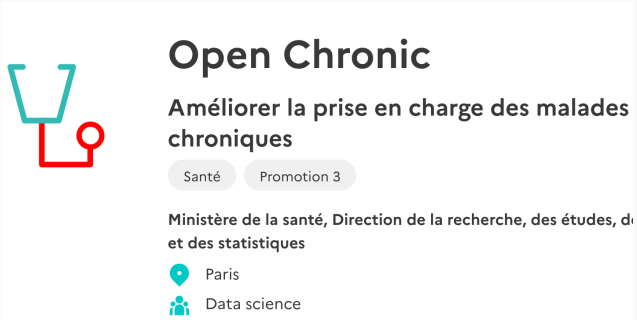
Significance of the Work: Why Do It?

- Scientific advancement



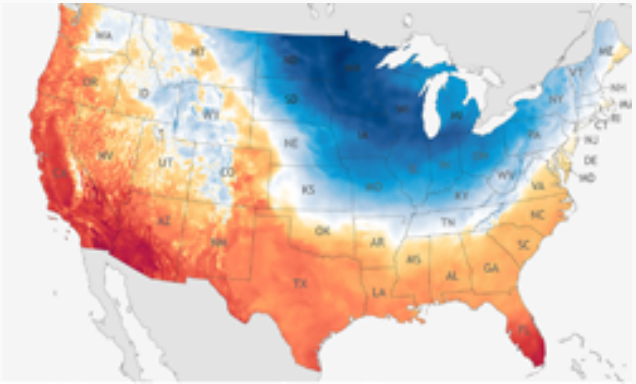
- Prevention and management of natural disasters

- Impact on public health



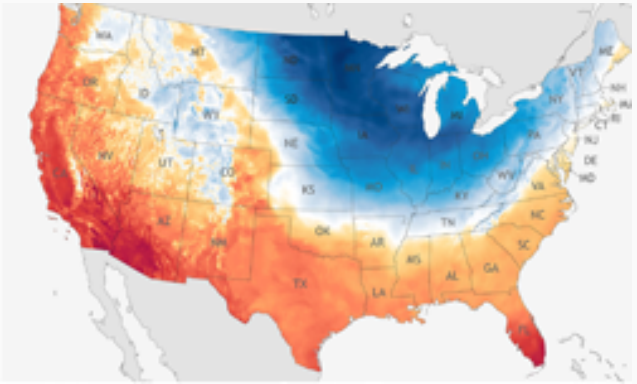
Significance of the Work: Why Do It?

- Environmental sustainability

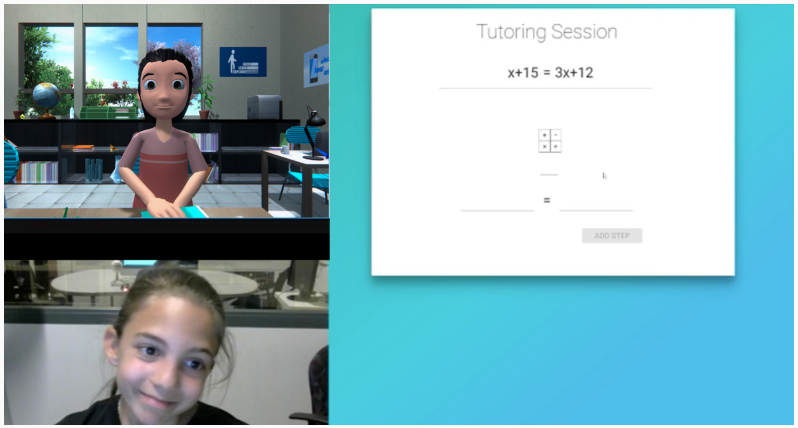


Significance of the Work: Why Do It?

- Environmental sustainability

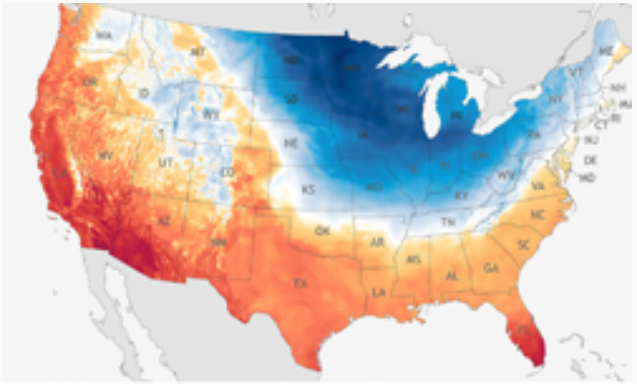


- Boost to education and research

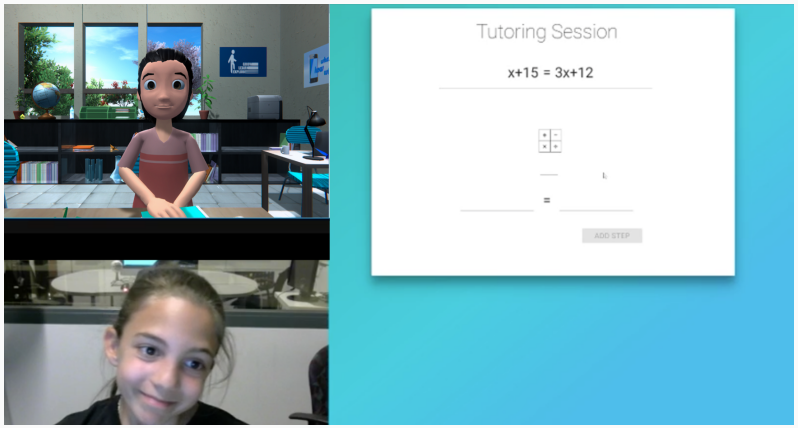


Significance of the Work: Why Do It?

- Environmental sustainability



- Boost to education and research



- Participatory democracy

Outline : Prerequisites

The Program

Theoretical Parts

Foundations in statistics and linear algebra: General overview of statistical learning, mathematical foundations of different models, intuitive approach

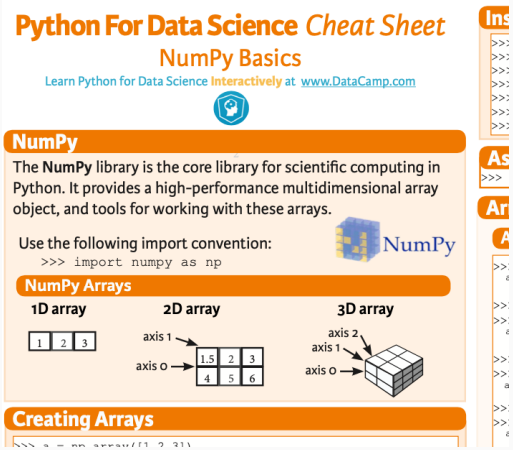
Practical Parts

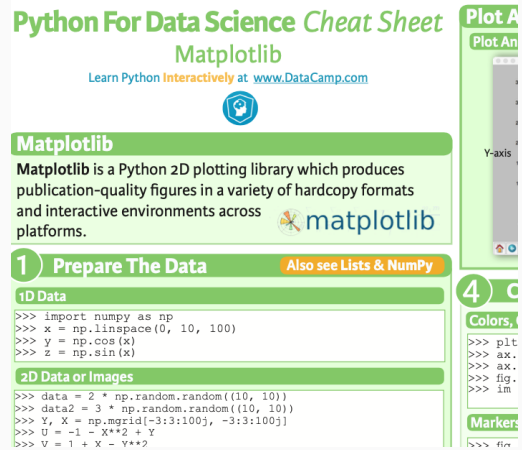
Python basics: Use of data manipulation tools, use of a DL library, use of an ML library, sentiment analysis, ranking based on wine preferences, ...[2]

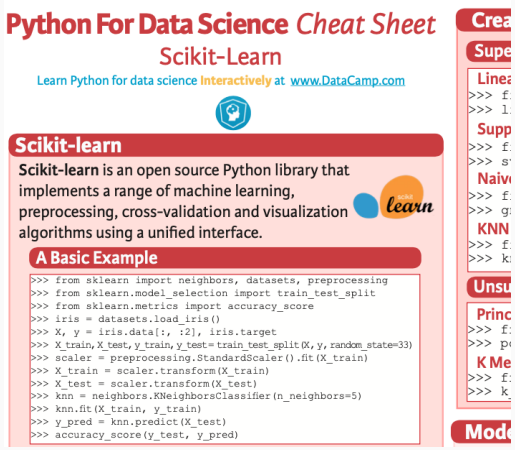
[2] Non-contractual for this course

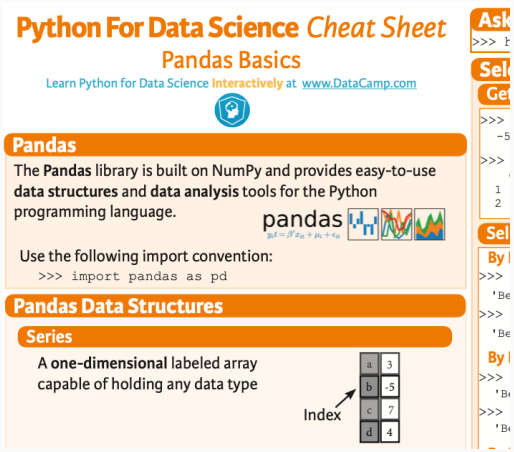
Material

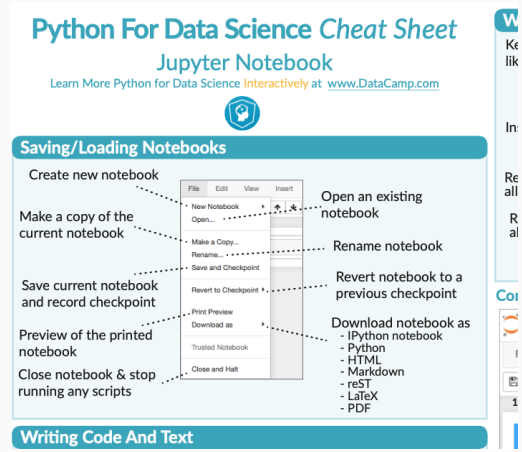
- Computer
- Jupyter Notebook and Anaconda:
  https://www.anaconda.com/download/
- The notebooks and cheatsheets available online:

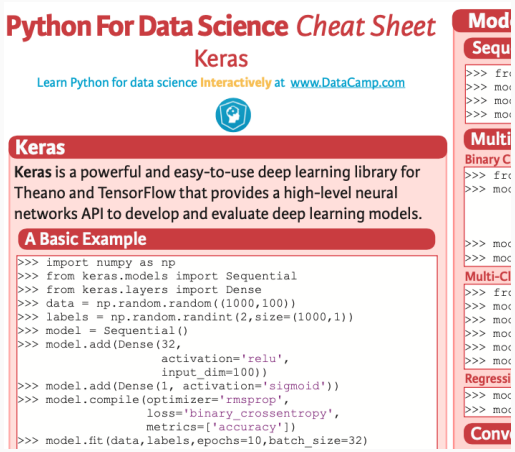












Outline : Data Exploration Practical: MovieLens

Data Exploration: MovieLens

A simple study of a movie review dataset

- 3 million ratings
- Social descriptors: age, gender, ...
- First basic approach to data mining

movielens

Introduction to pandas

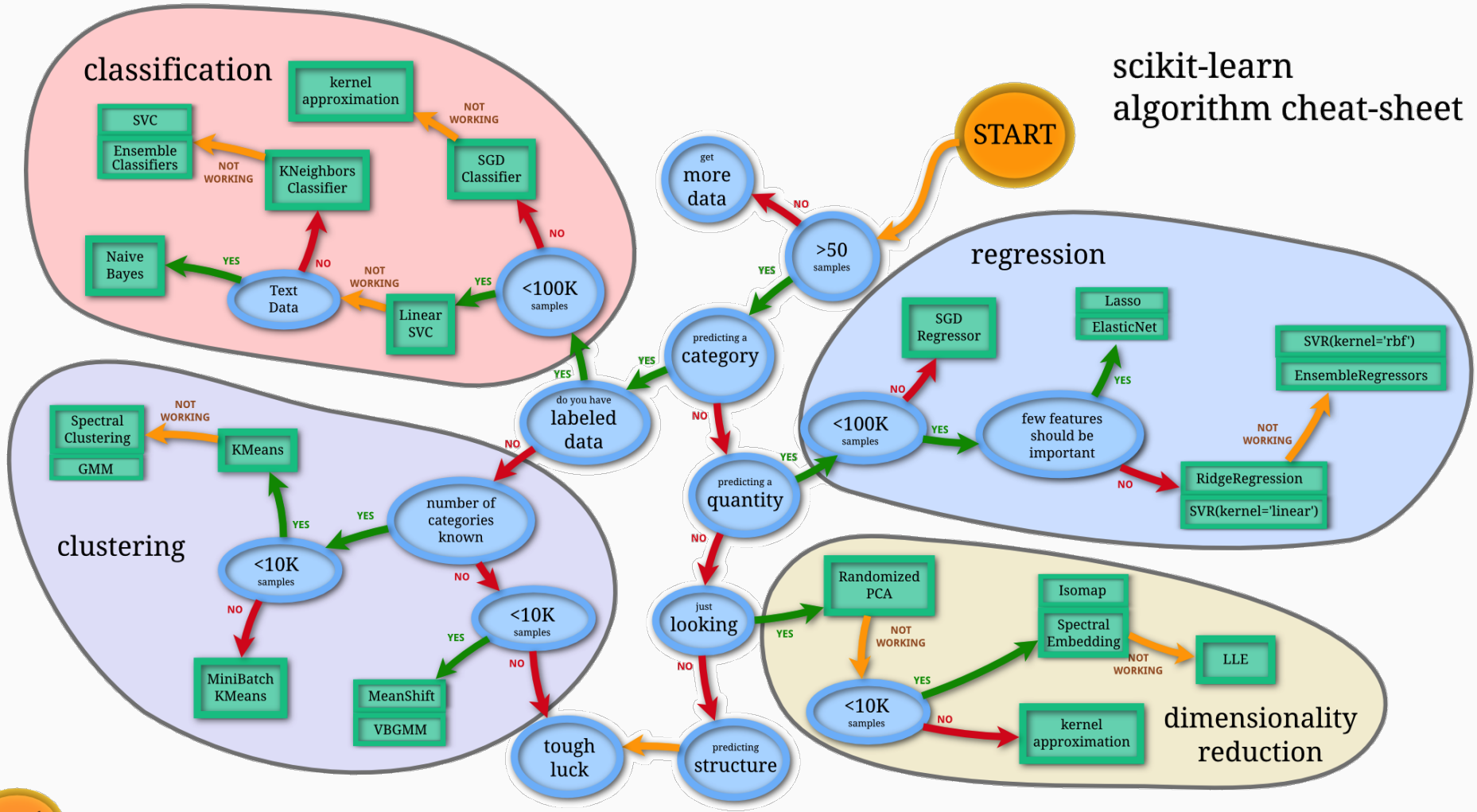
- Python library for data manipulation: https://pandas.pydata.org/
- Enables performing operations and visualizations
- Easy to use

Outline : Overview

The Different Methods

- Classification: predict a given class
- Regression: predict a value
- Clustering: group elements into clusters (undetermined classes)
- Dimensionality reduction: reduce the data representation space
- Anomaly detection

The Different Methods



Examples: Classification

- Speech emotion recognition:
- Animal species classification from images:
- Object detection in medical images:

Examples: Classification

- Speech emotion recognition: **is the person angry or happy?**
- Animal species classification from images:
- Object detection in medical images:

Examples: Classification

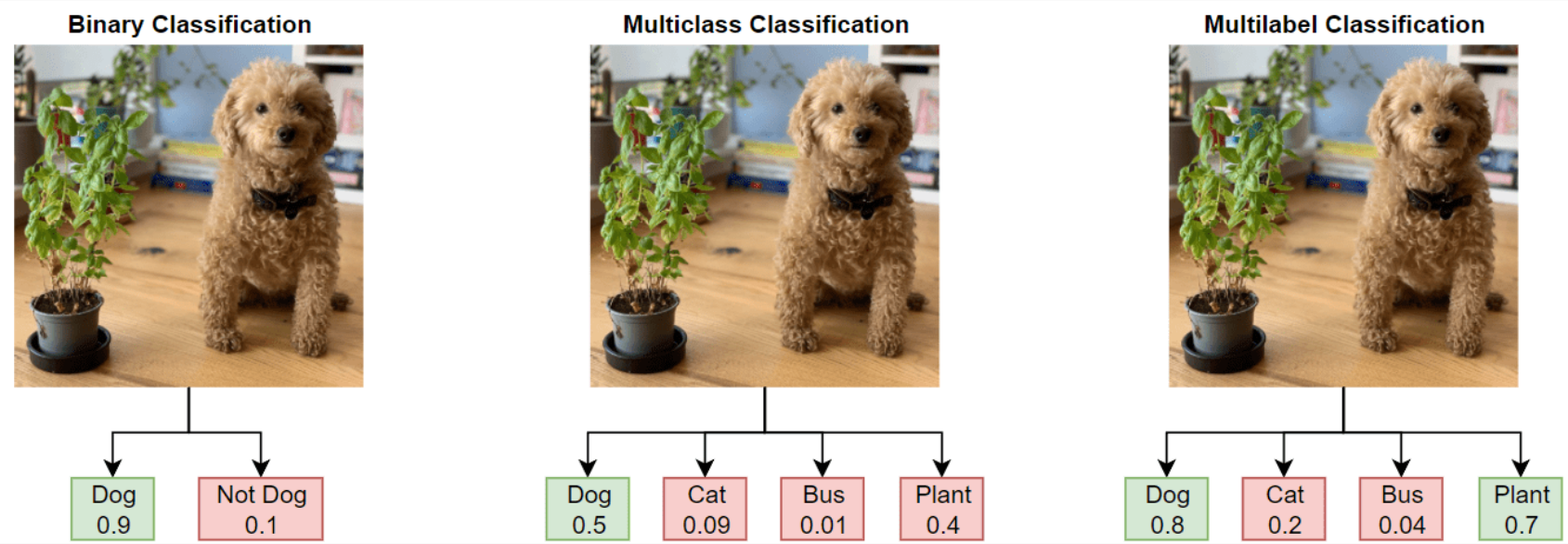
- Speech emotion recognition: **is the person angry or happy?**
- Animal species classification from images: **is it a cat or a puma?**
- Object detection in medical images:

Examples: Classification

- Speech emotion recognition: **is the person angry or happy?**
- Animal species classification from images: **is it a cat or a puma?**
- Object detection in medical images: **is it a tumor?**

Examples: Classification

- Speech emotion recognition: **is the person angry or happy?**
- Animal species classification from images: **is it a cat or a puma?**
- Object detection in medical images: **is it a tumor?**



Examples: Regression

- Speech emotion recognition:
- Damage assessment from post-earthquake images:

- Detection of Alzheimer's severity in voice:

Examples: Regression

- Speech emotion recognition: **what is the intensity of the anger?**
- Damage assessment from post-earthquake images:

- Detection of Alzheimer's severity in voice:

Examples: Regression

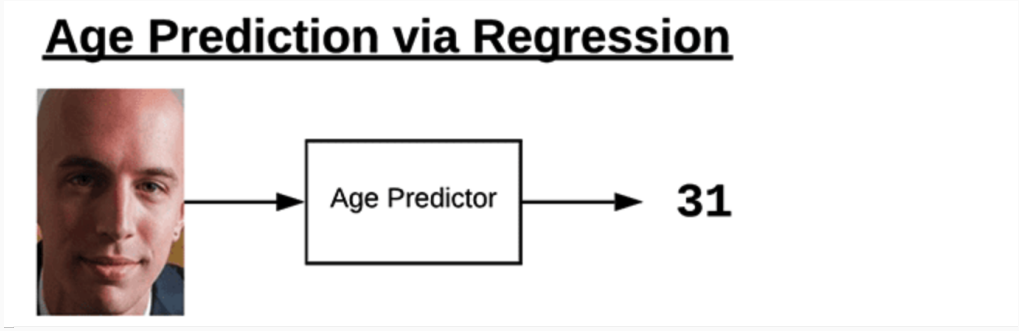
- Speech emotion recognition: **what is the intensity of the anger?**
- Damage assessment from post-earthquake images: **can the ambulance use the bridge?**
- Detection of Alzheimer's severity in voice:

Examples: Regression

- Speech emotion recognition: **what is the intensity of the anger?**
- Damage assessment from post-earthquake images: **can the ambulance use the bridge?**
- Detection of Alzheimer's severity in voice: **how advanced is the condition?**

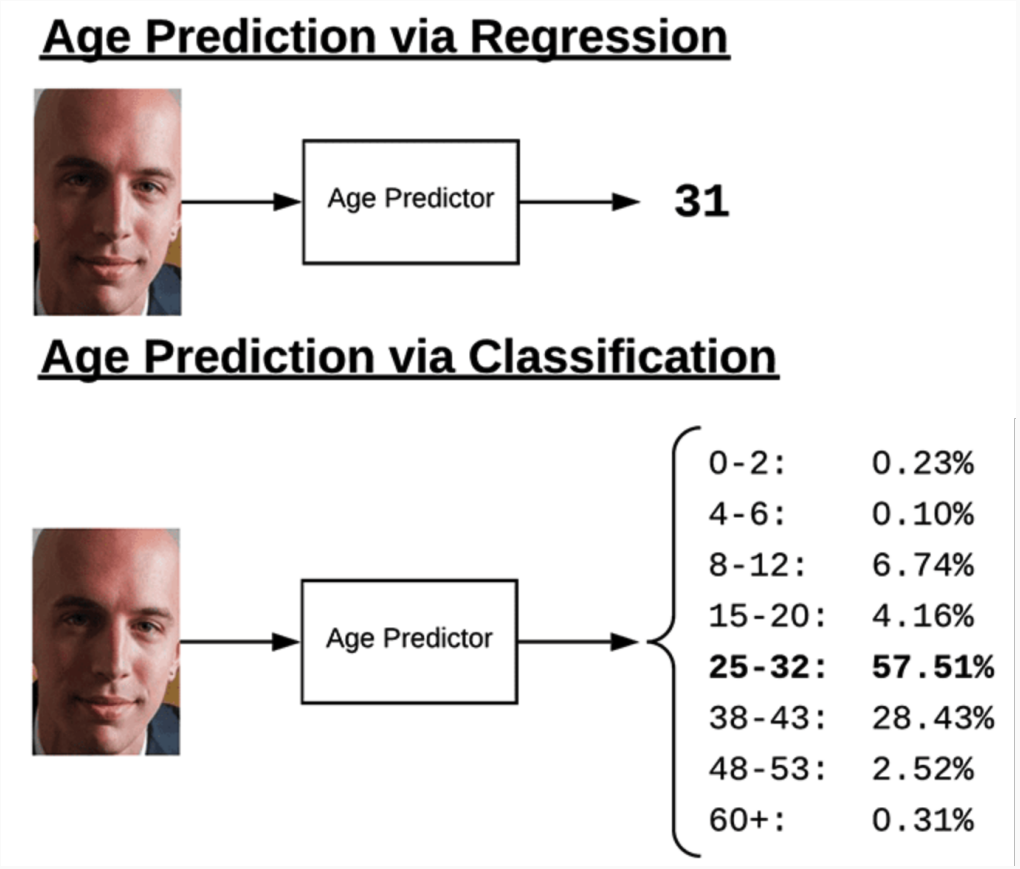
Examples: Regression

- Speech emotion recognition: **what is the intensity of the anger?**
- Damage assessment from post-earthquake images: **can the ambulance use the bridge?**
- Detection of Alzheimer's severity in voice: **how advanced is the condition?**



Examples: Regression

- Speech emotion recognition: **what is the intensity of the anger?**
- Damage assessment from post-earthquake images: **can the ambulance use the bridge?**
- Detection of Alzheimer's severity in voice: **how advanced is the condition?**



Examples: Clustering

- Topic mining in political forums:
- Misinformation classification on social networks:
- Customer segmentation:

Examples: Clustering

- Topic mining in political forums: **what are citizens concerned about?**
- Misinformation classification on social networks:

- Customer segmentation:

Examples: Clustering

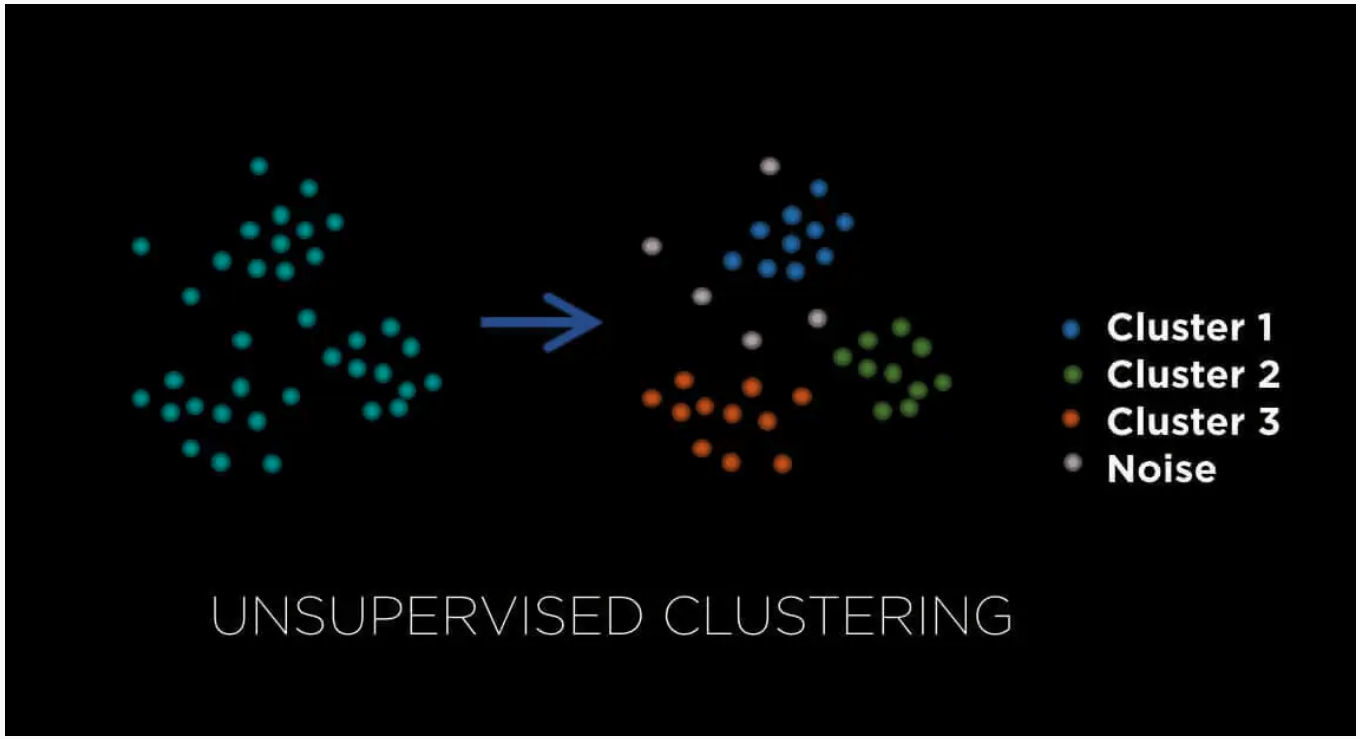
- Topic mining in political forums: **what are citizens concerned about?**
- Misinformation classification on social networks: **we group these news items that seem unusual**
- Customer segmentation:

Examples: Clustering

- Topic mining in political forums: **what are citizens concerned about?**
- Misinformation classification on social networks: **we group these news items that seem unusual**
- Customer segmentation: **people who like beers**

Examples: Clustering

- Topic mining in political forums: **what are citizens concerned about?**
- Misinformation classification on social networks: **we group these news items that seem unusual**
- Customer segmentation: **people who like beers**



Questions?

References i