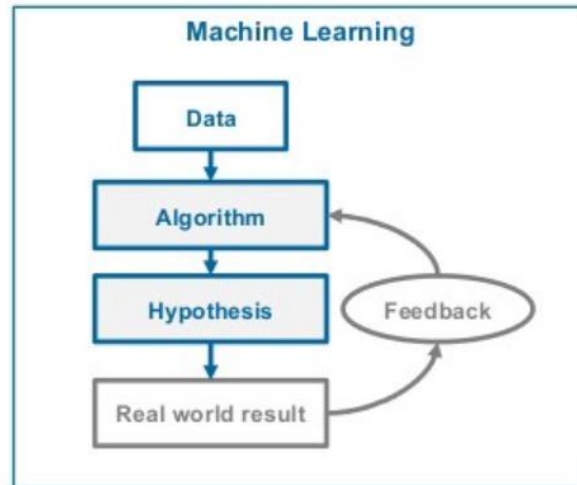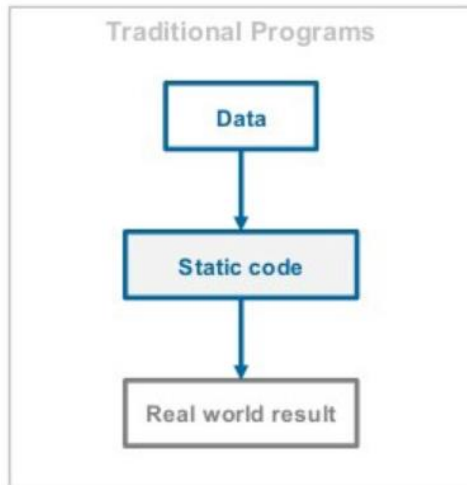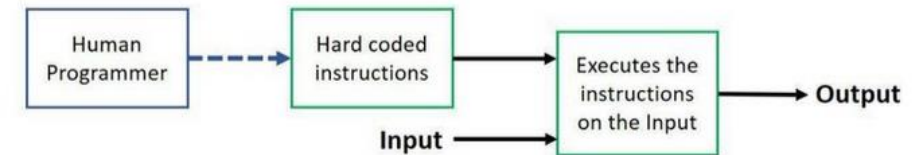**Babeș-Bolyai University**

Faculty of Mathematics and Computer Science

Curs opțional

Modele de inteligență artificială în schimbarea climatică
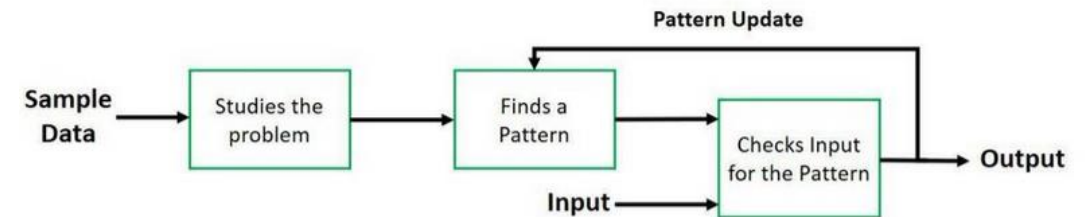
Reminder

Aspecte de baza

Machine Learning vs. Programarea tradițională

ML - utilizare

## ML se folosește:

- Atunci cand experitiza umană nu există sau este greu de aplicat
- Pentru că poate rezolva un task intr-un mod mult mai simplu si mai scalabil

Metodele inteligente ML sunt eficiente dacă:

- Modelele ML - customizate și antrenate pentru un anumit task specific, deci nu pot sa fie generalizate foarte usor
- Modelele ML - un set mare de date pentru a nu overfit/undefit the model

ML – Data type

- Date numerice

- Date categoriale (reprezintă categorii)
  - ordinal date – date categoriale ordonate

- Date temporale

- Date de tip text

- Date binare (yes/no, t/f)

- Date mixte (combinații din oricare se mai sus)
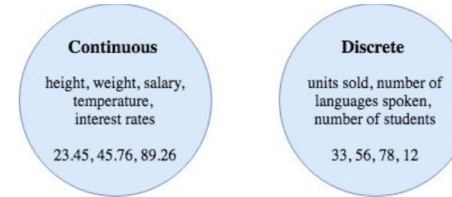
ML - Taxonomie
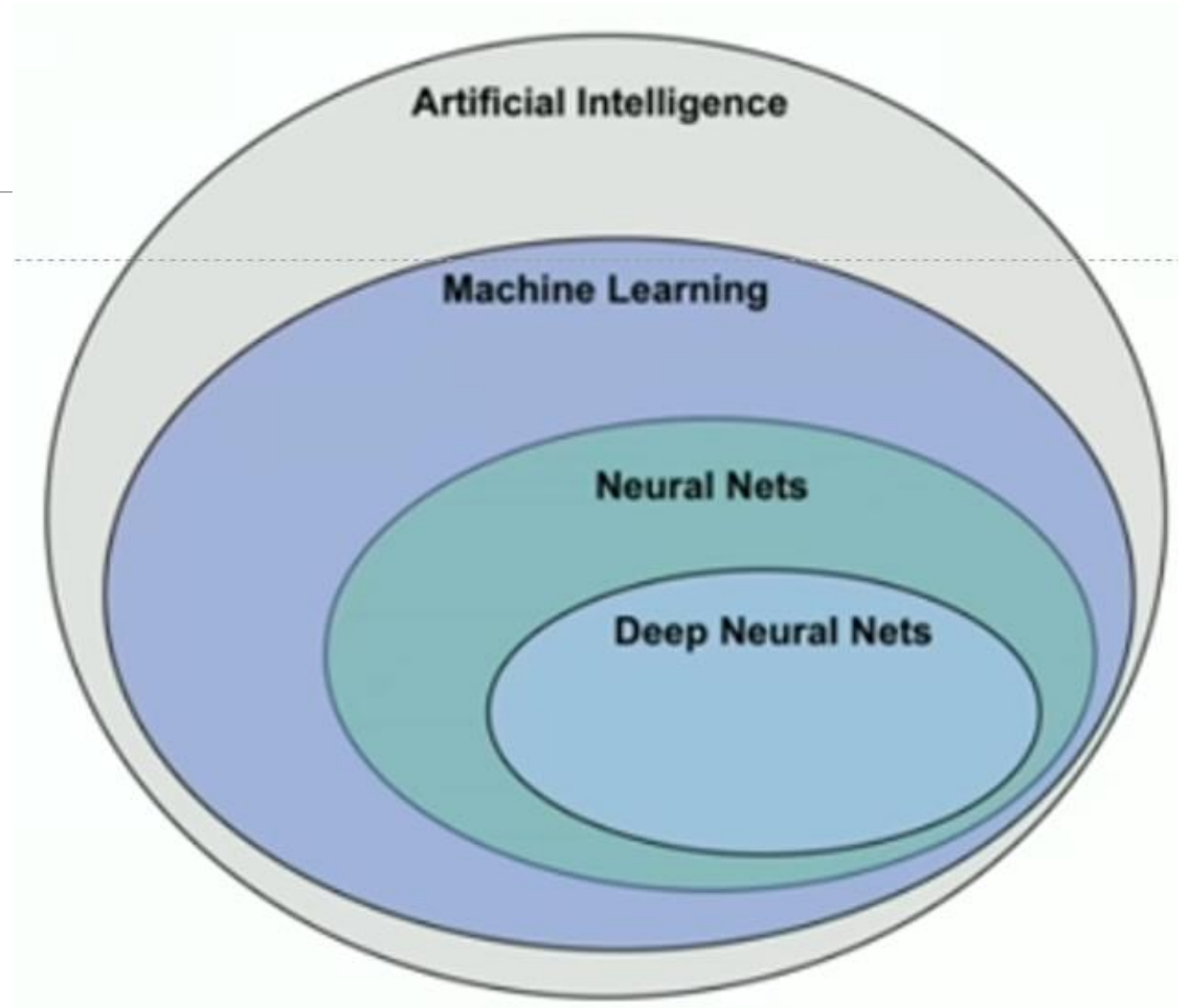
ML - Terminologie

Instance: A single row of data is called an instance. It is an observation from the domain.

Feature: A single column of data is called a feature. It is a component of an observation and is also called an attribute of a data instance. Some features may be inputs to a model (the predictors) and others may be outputs or the features to be predicted.
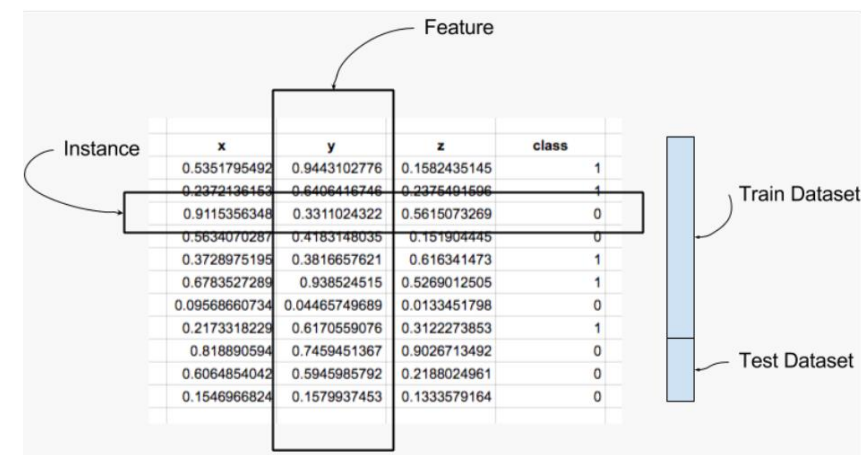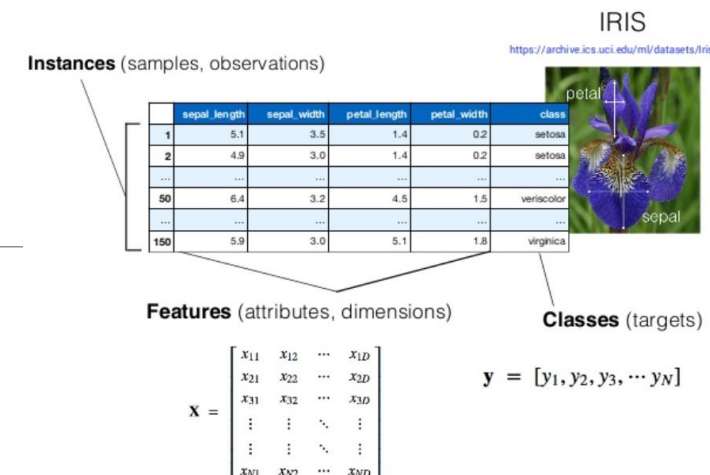
Data Type: Features have a data type. They may be real or integer-valued or may have a categorical or ordinal value. we can have strings, dates, times, and more complex types, but typically they are reduced to real or categorical values when working with traditional machine learning methods.

Datasets: A collection of instances is a dataset and when working with machine learning methods we typically need a few datasets for different purposes.

Training Dataset: A dataset that we feed into our machine learning algorithm to train our model.

Testing Dataset: A dataset that we use to validate the accuracy of our model but is not used to train the model. It may be called the validation dataset.

It is important to mention the distinction between validation and testing datasets. We train our model on our training set, we use a validation set to tune the model parameters, and we use a test set to asses the accuracy of the model.

ML – Etape de lucru recomandate

```
înțelegerea domeniului în care se aplică învățare automată (informații și scop)
  → identificarea datelor utilizate in model
      • colectarea datelor
      • selectarea datelor
  → exploratory data analysis (EDA)
  → aplicarea tehnicilor de preprocesare
  ↓
identificarea algoritmilor pentru crearea modelelor
  → antrenarea modelului
  → testarea/evaluarea modelului
  → validarea rezultatelor
  ↓
vizualizarea rezultatelor
  → interpretarea rezultatelor
  → analiza și sinteza rezultatelor obținute
```