



# Introducere în Machine Learning

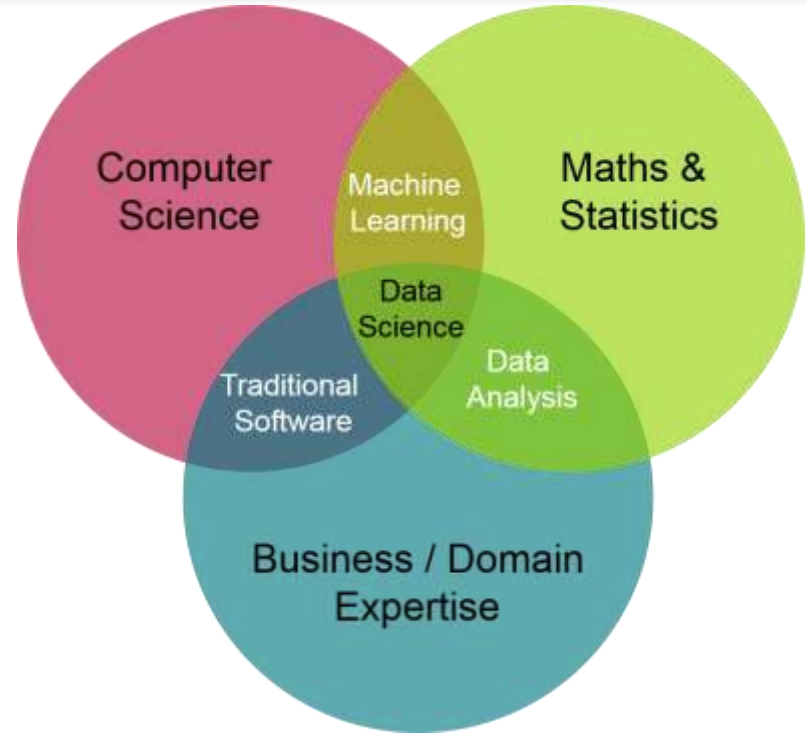
## De la date la informații

*Oprescu Gabriel*



## Data Science

*Domeniu inter-disciplinar care utilizează metode științifice, procese și algoritmi pentru a extrage cunoștințe și informații din date.*





## Matematică

- Algebră (datele sunt reprezentate ca matrici)
- Derivate (pentru algoritmi de optimizare)

## Statistică

- Distribuții și momente
- Testare ipoteze

## Computer Science

- Limbaje de programare (R / Python). SQL
- Algoritmi de programare

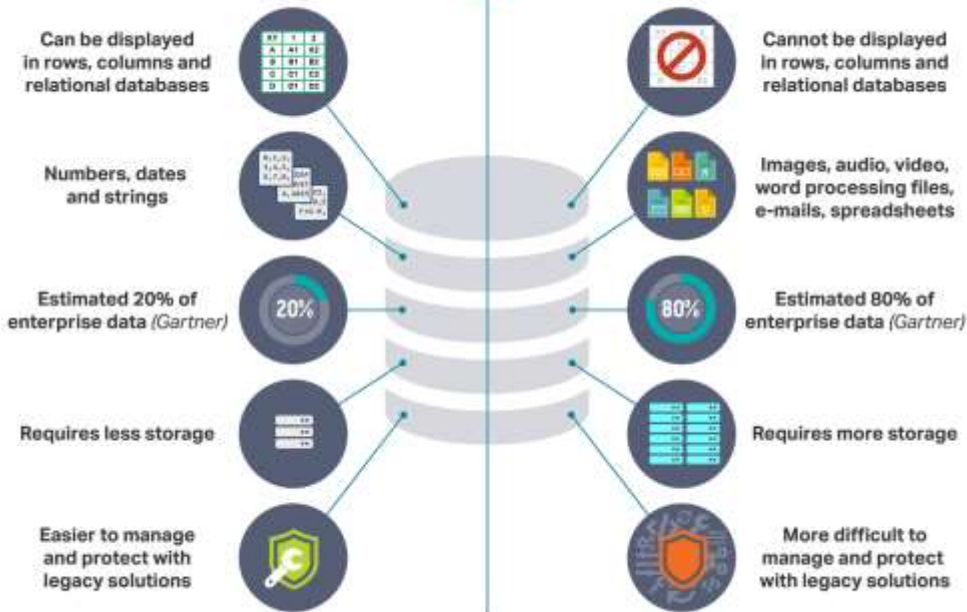
## Business / Domain Expertise

- Înțelegerea în profuzime a problemei analizate
- Consultări cu persoane cu experiență în domeniu



Fără date nu există  
Data Science

## Structured Data VS Unstructured Data





## Date structurate

| EmployeeID | Ename | DeptID | Salary | Dname   | Dlocation |
|------------|-------|--------|--------|---------|-----------|
| 1001       | John  | 2      | 4000   | IT      | New Delhi |
| 1002       | Anna  | 1      | 3500   | HR      | Mumbai    |
| 1003       | James | 1      | 2500   | HR      | Mumbai    |
| 1004       | David | 2      | 5000   | IT      | New Delhi |
| 1005       | Mark  | 2      | 3000   | IT      | New Delhi |
| 1006       | Steve | 3      | 4500   | Finance | Mumbai    |
| 1007       | Alice | 3      | 3500   | Finance | Mumbai    |

## Date nestructurate



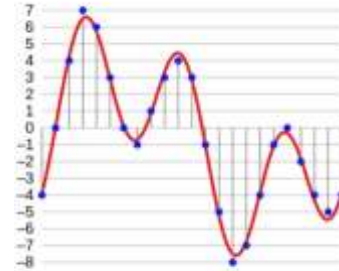
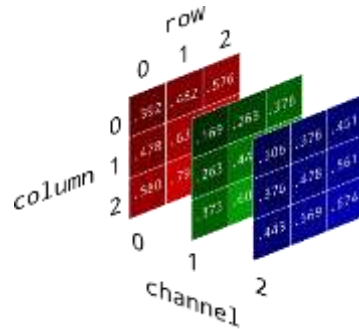
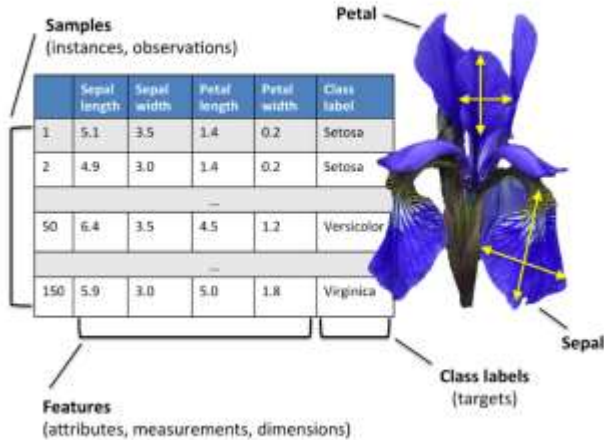
“ Poate că povestea este partea cea mai frumoasă a vieții omenești... cu povești ne leagă lumea, cu povești ne adoarme... Ne trezim și murim cu ele. ”





## Date structurata

## Date nestructurate



[-4, 0, 4, 7, 6, 3, 0, -1, 1, 3, 4, 3, -1, -5, -8, -7, -4, -1, 0, -2, -4, -5, -4]

Array

Vocabulary:  
Man, woman, boy,  
girl, prince,  
princess, queen,  
king, monarch



|          | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|----------|---|---|---|---|---|---|---|---|---|
| man      | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| woman    | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| boy      | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| girl     | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| prince   | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| princess | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| queen    | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| king     | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| monarch  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

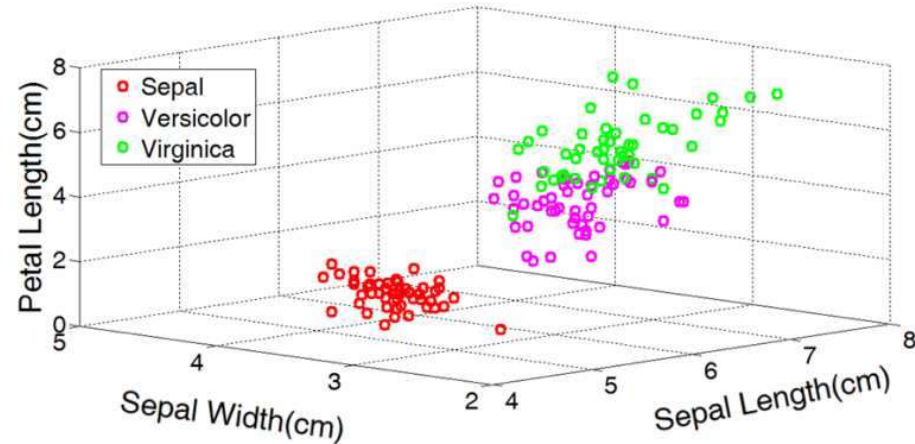
Each word gets a 1x9 vector representation





## Date structurată – reprezentare în spațiu multidimensional

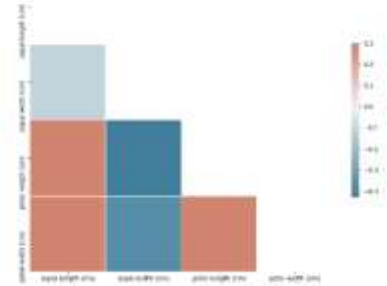
| id  | sepal length | sepal width | petal length | petal width | label      |
|-----|--------------|-------------|--------------|-------------|------------|
| 1   | 5.1          | 3.5         | 1.4          | 0.2         | setosa     |
| 2   | 4.9          | 3           | 1.4          | 0.2         | setosa     |
| 3   | 4.7          | 3.2         | 1.3          | 0.2         | setosa     |
| 51  | 7            | 3.2         | 4.7          | 1.4         | versicolor |
| 52  | 6.4          | 3.2         | 4.5          | 1.5         | versicolor |
| 53  | 6.9          | 3.1         | 4.9          | 1.5         | versicolor |
| 101 | 6.3          | 3.3         | 6            | 2.5         | virginica  |
| 102 | 5.8          | 2.7         | 5.1          | 1.9         | virginica  |
| 103 | 7.1          | 3           | 5.9          | 2.1         | virginica  |



## Data Analytics

Exploarea datelor pentru a observa tendințe, corelații și pentru a extrage informații pe baza cărora se pot lua decizii.

|                   | sepal length |     |      | sepal width |     |      |
|-------------------|--------------|-----|------|-------------|-----|------|
|                   | min          | max | mean | min         | max | mean |
| <b>setosa</b>     | 4.3          | 5.8 | 5.0  | 2.3         | 4.4 | 3.4  |
| <b>versicolor</b> | 4.9          | 7.0 | 5.9  | 2.0         | 3.4 | 2.8  |
| <b>virginica</b>  | 4.9          | 7.9 | 6.6  | 2.2         | 3.8 | 3.0  |



## Machine Learning

Studierea algoritmilor care devin mai performanți o dată cu exepriența.

Algoritmi care pot face predicții fără a fi explicit programați în acest sens.

| id  | sepal length | sepal width | petal length | petal width | label      |
|-----|--------------|-------------|--------------|-------------|------------|
| 1   | 5.1          | 3.5         | 1.4          | 0.2         | setosa     |
| 51  | 7            | 3.2         | 4.7          | 1.4         | versicolor |
| 103 | 7.1          | 3           | 5.9          | 2.1         | virginica  |

Cărei clase îi aparține noul obiect

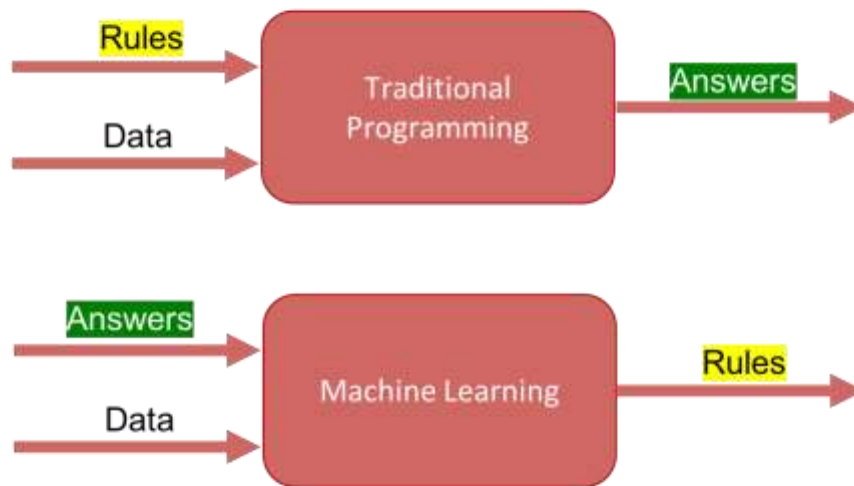
|     |   |     |     |   |     |
|-----|---|-----|-----|---|-----|
| 151 | 8 | 3.5 | 6.2 | 3 | ??? |
|-----|---|-----|-----|---|-----|





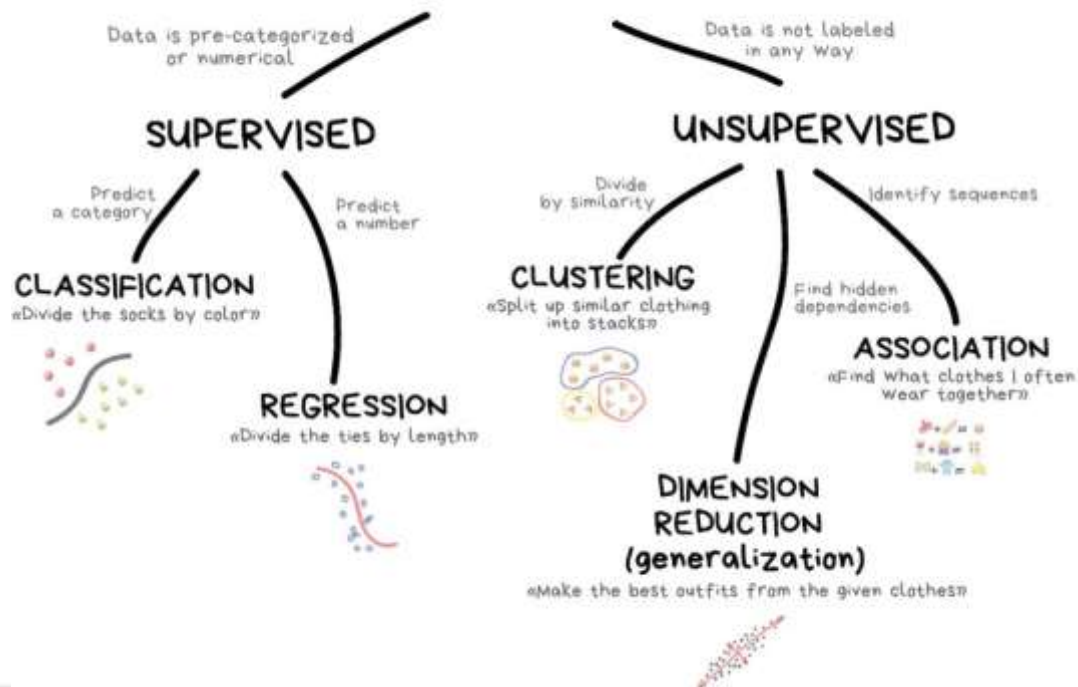


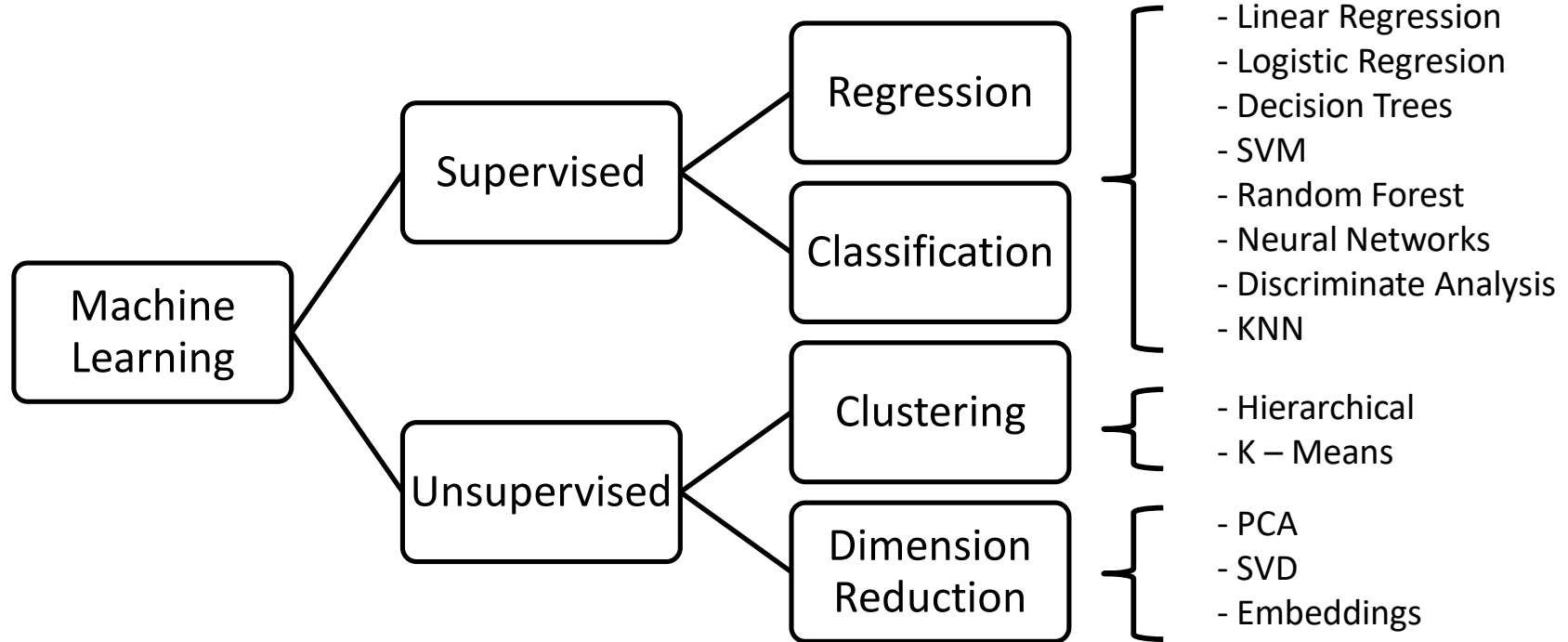
Sistem clasic de  
luare a deciziilor vs.  
Machine Learning





## CLASSICAL MACHINE LEARNING







# Center for Career Development

by LINKgroup

