COESIONE ITALIA 21-27
EMILIA-ROMAGNA

Cofinanziato dall'Unione europea

REPUBBLICA ITALIANA

Regione Emilia-Romagna

DATA LAB

GUARDA AVANTI

Big Data, nuove competenze per nuove professioni.

FORM INDUSTRIA EMILIA ROMAGNA

ifoa Sapere utile

REISS ROMOLI la passione della conoscenza

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

Università degli Studi di Ferrara

UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

UNIVERSITÀ DI PARMA

POLITECNICO MILANO 1863 POLO TERRITORIALE DI PIACENZA

UNIVERSITÀ CATTOLICA del Sacro Cuore

"Anticipare la crescita con le nuove competenze sui Big Data" Operazione Rif. PA 2023-19167/RER approvata con DGR n° 843 del 29 maggio 2023 e co-finanziata dal Fondo Sociale Europeo Plus 2021-2027 Regione Emilia-Romagna

ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

Operazione Rif. PA 2023-19167/RER/10/1, "ANTICIPARE LA CRESCITA CON LE NUOVE COMPETENZE SUI BIG DATA", approvata dalla Regione Emilia-Romagna con DGR n° 843 del 29/05/2023 e co-finanziata dal Fondo Sociale Europeo Plus 2021-2027

DATA LAB

Definizione di dati

●I dati sono rappresentazioni originarie, cioè non interpretate, di un fenomeno, evento, o fatto, effettuate attraverso simboli o combinazioni di simboli, o di qualsiasi altra forma espressiva legate a un qualsiasi supporto

● Dati sono rappresentazioni di eventi o fatti

– Non interpretate (originarie)

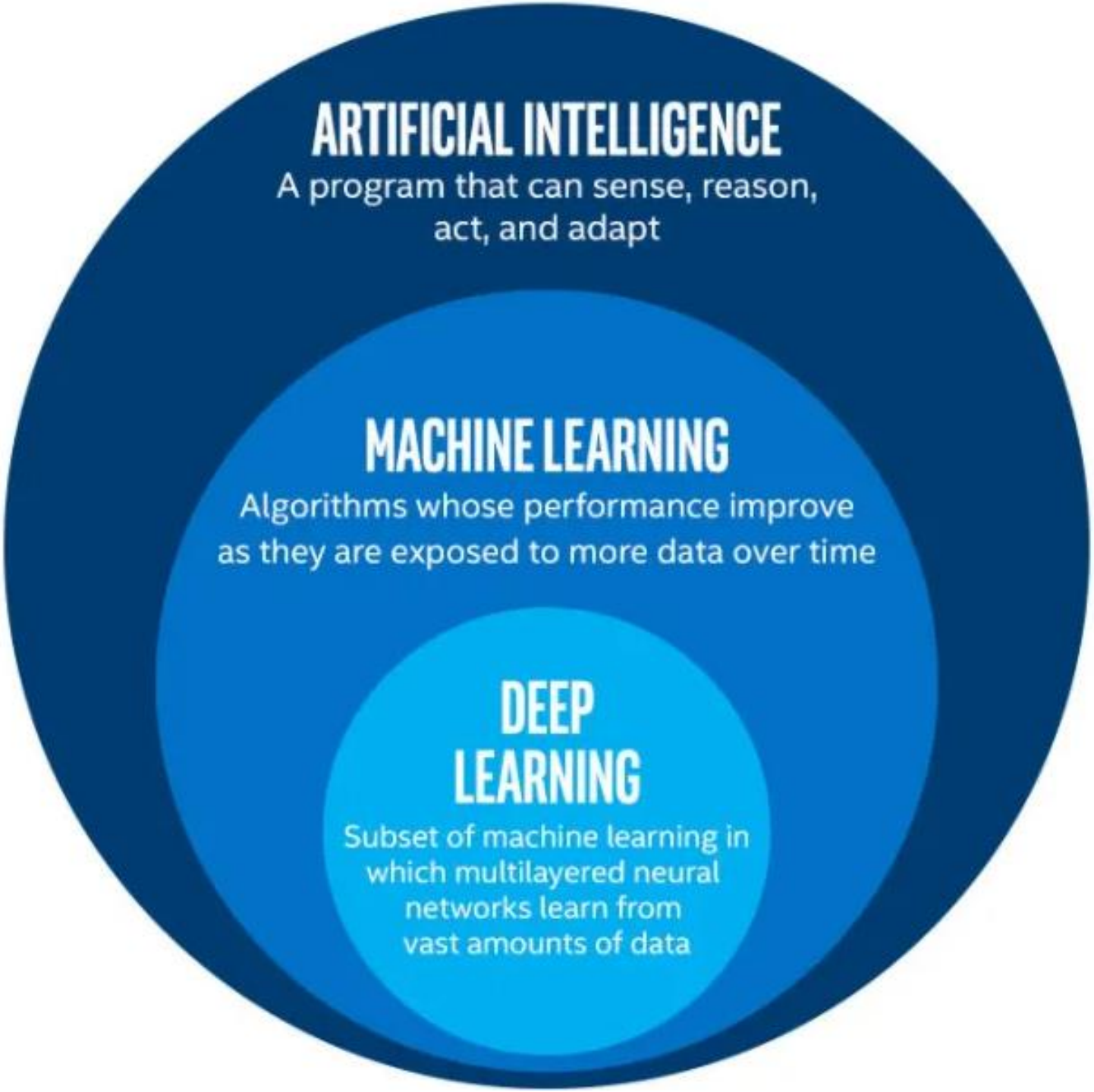
– Attraverso simboli (o combinazioni di simboli)

– Contenute su supporti (forma espressiva)

Definizione di informazione

L’informazione deriva da un dato, o più verosimilmente da un insieme di dati, che sono stati sottoposti a un processo di interpretazione che li ha resi significativi per il destinatario

- L'informazione è:
  - insieme di dati
  - interpretati
  - comprensibili per il destinatario


ARTIFICIAL INTELLIGENCE
A program that can sense, reason, act, and adapt
MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time
DEEP LEARNING
Subset of machine learning in which multilayered neural networks learn from vast amounts of data

Cosa e' **il MACHINE LEARNING ?**

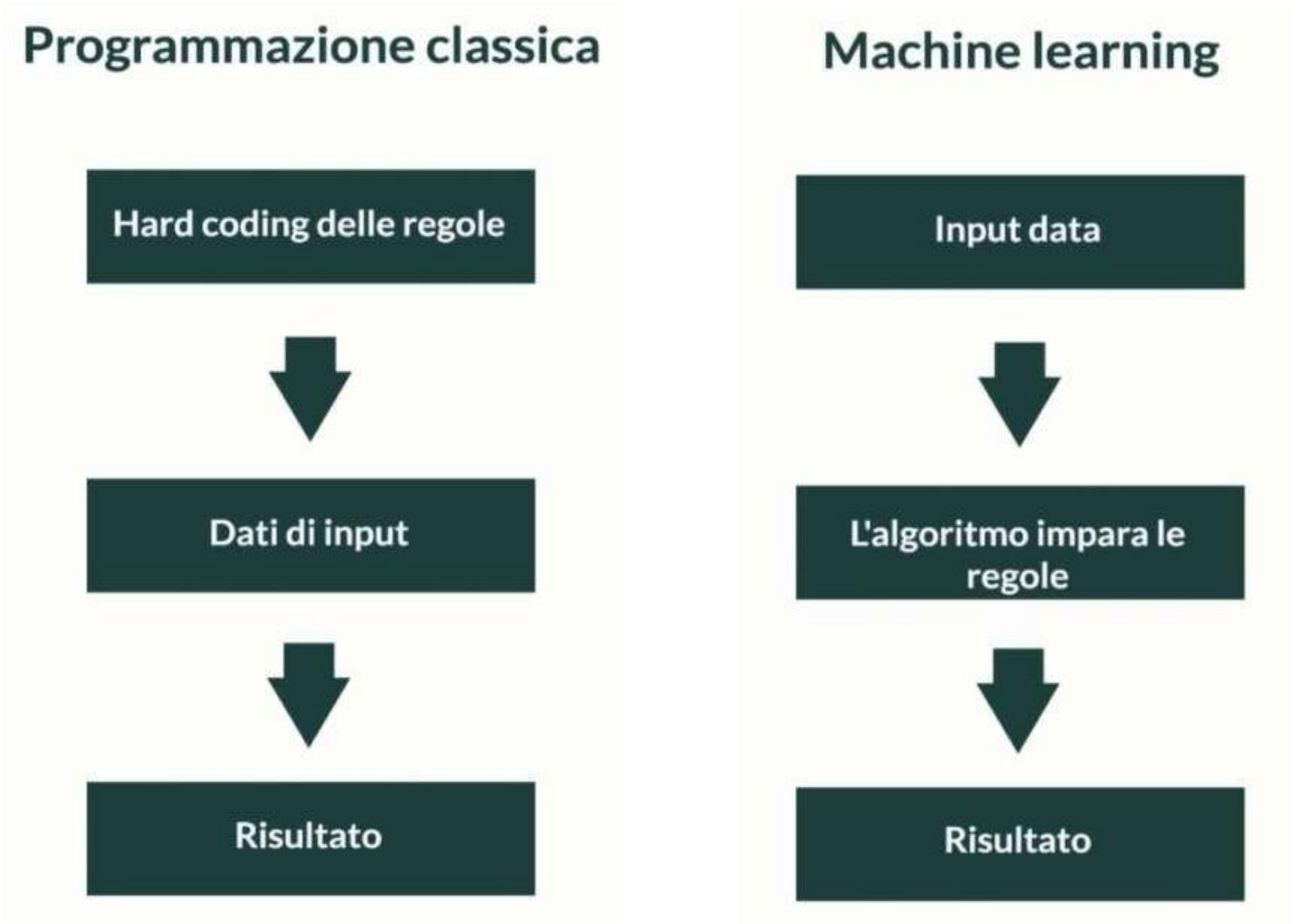
Cosa e' il machine learning

"E' il settore dell'intelligenza artificiale che studia come dare ai computer l'abilità di imparare senza essere esplicitamente programmati" - 1959

ARTHUR SAMUEL

Pioniere dell'AI

Inventore del termine "Machine learning"


Programmazione classica
Hard coding delle regole
Dati di input
Risultato
Machine learning
Input data
L'algoritmo impara le regole
Risultato

Analisi di rischio

Recommender system

Elaborazione del linguaggio parlato

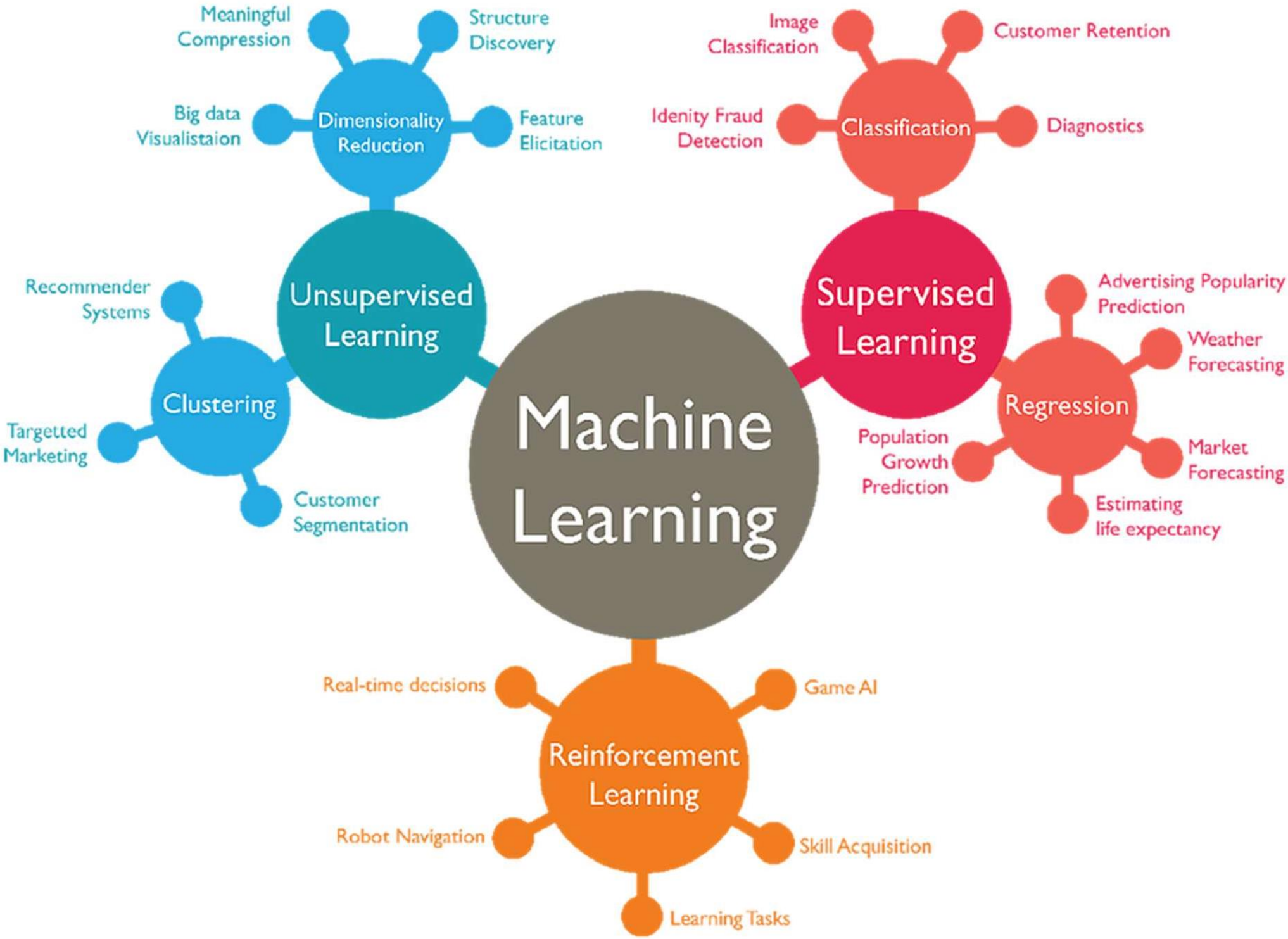
Riconoscimento di oggetti

Veicoli a guida autonoma

Fraud detection

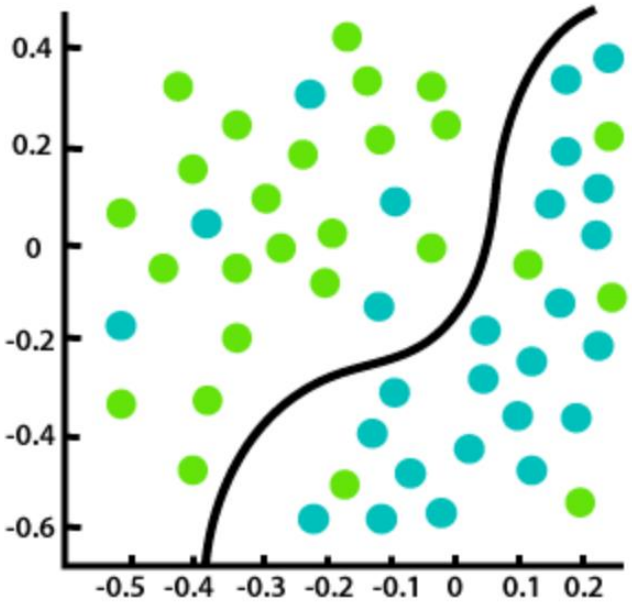
Diagnosi mediche

Customer segmentation

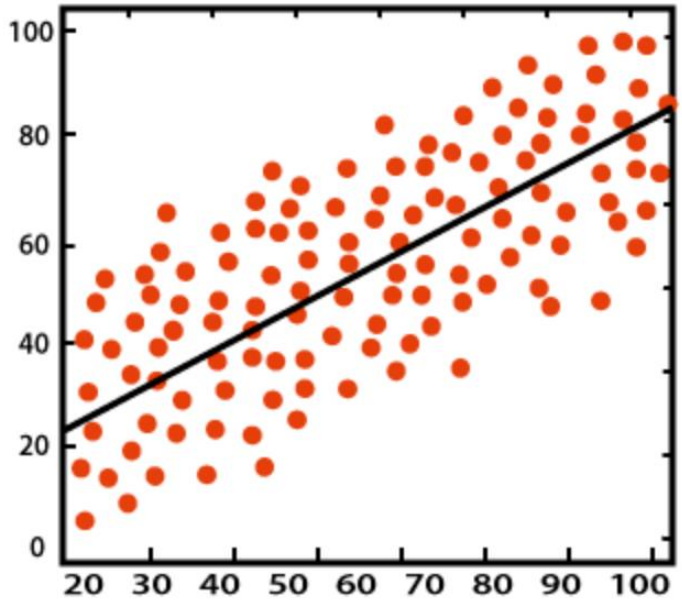

Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

SUPERVISIONATO

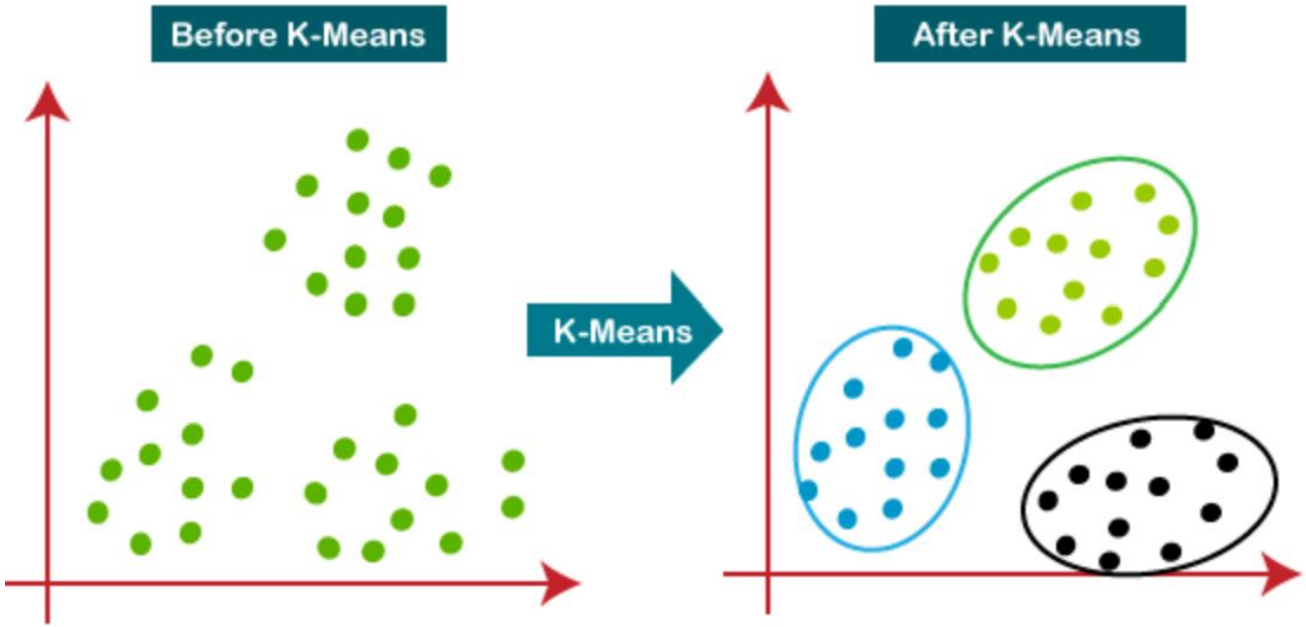
Classification

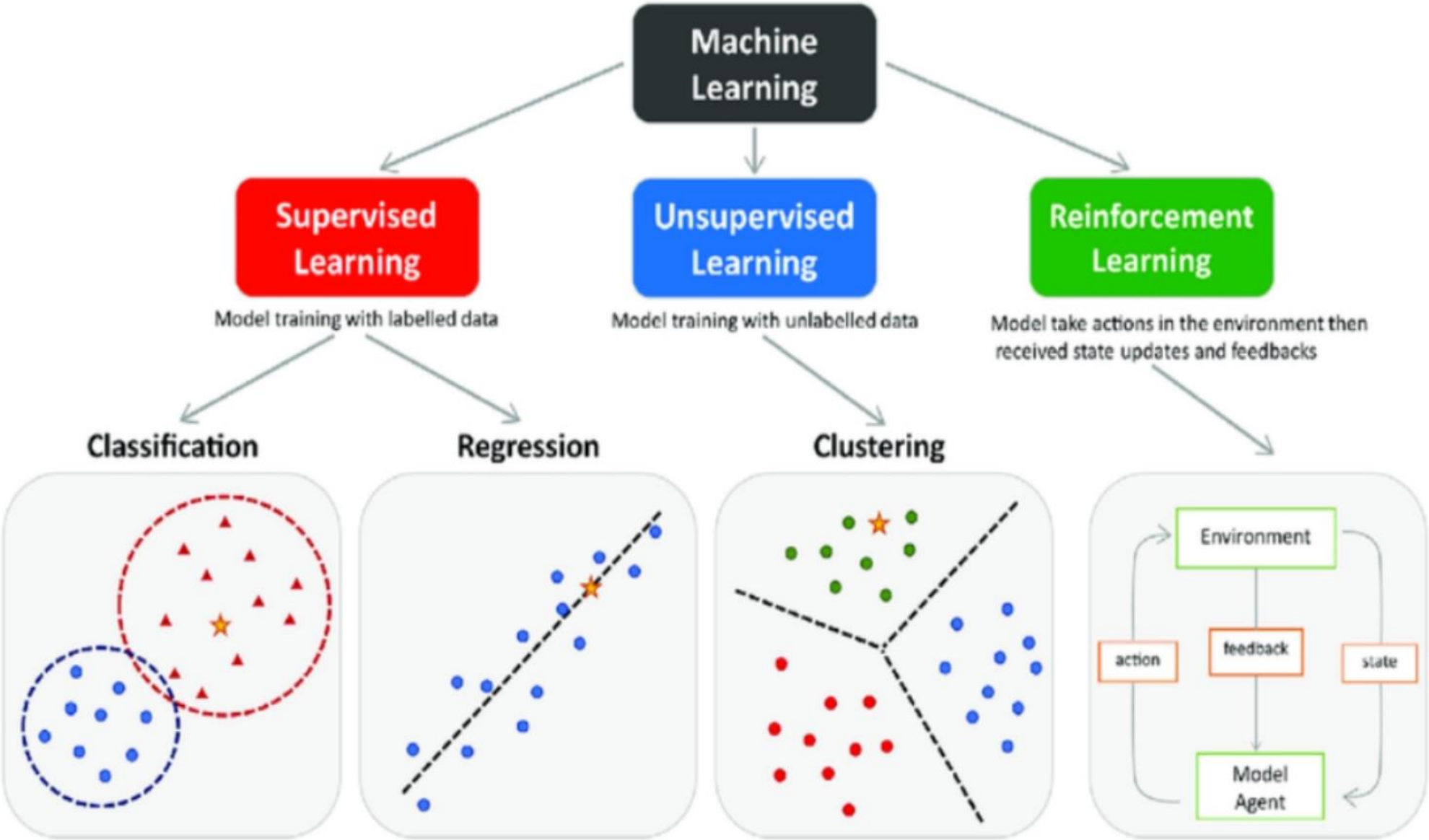


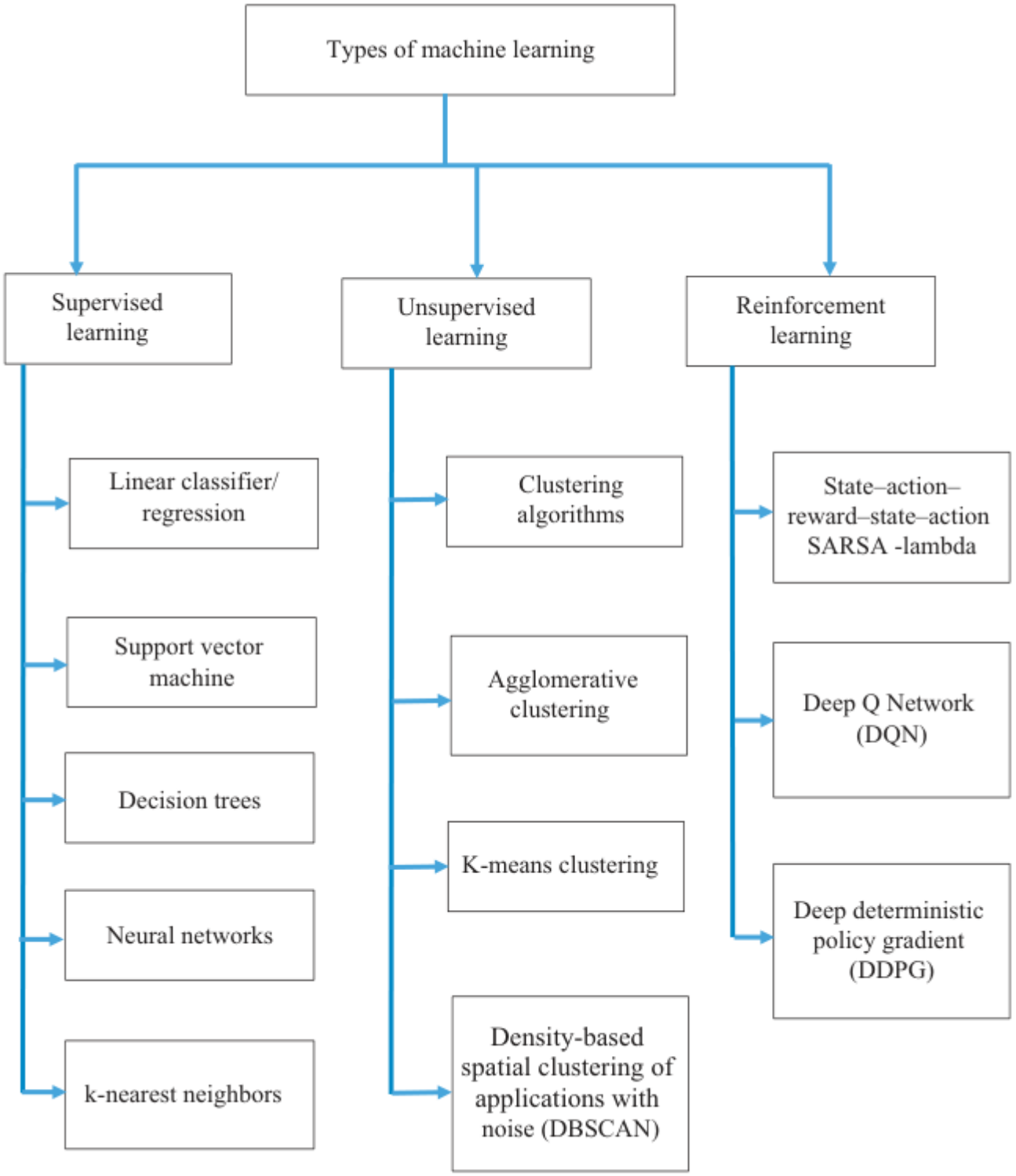
Regression

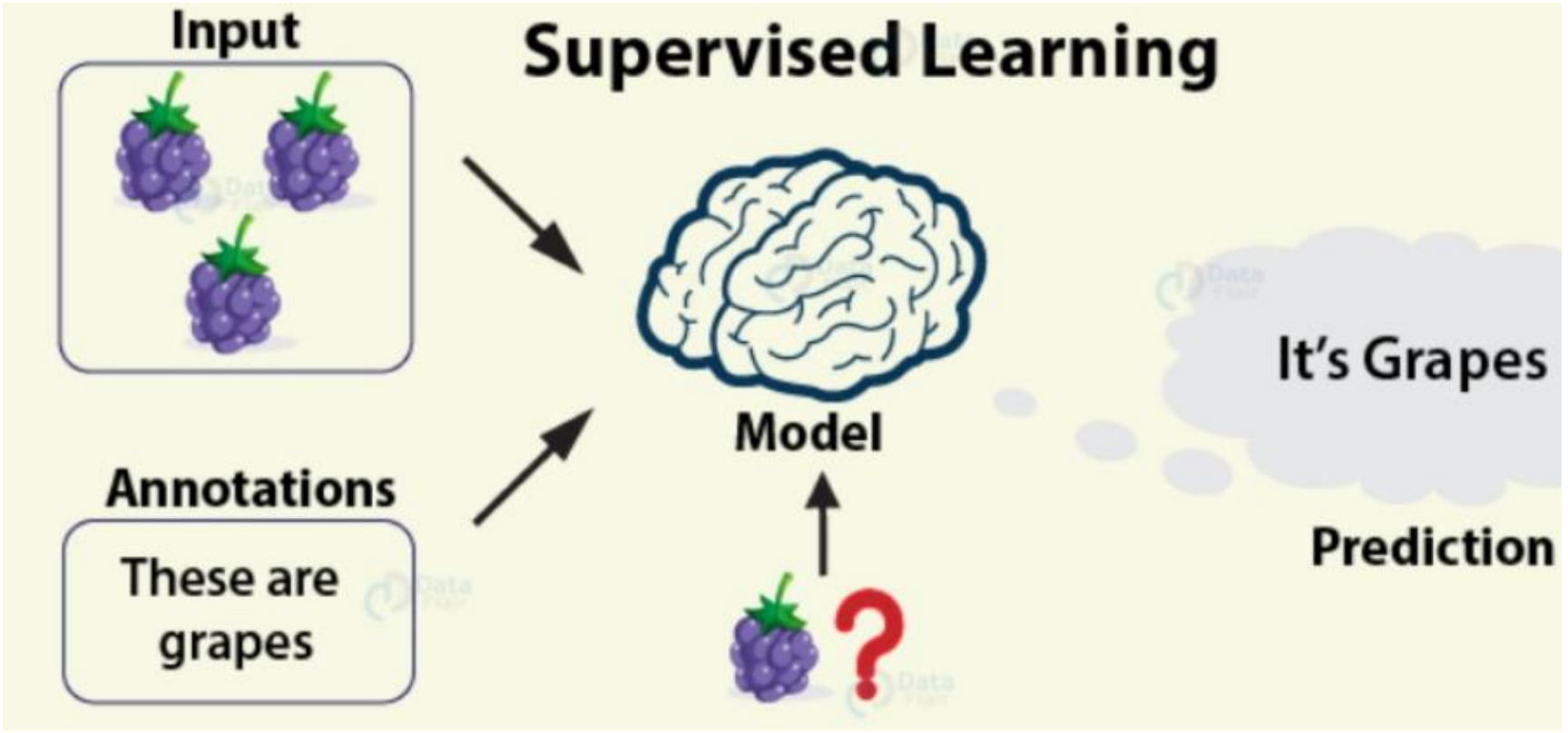


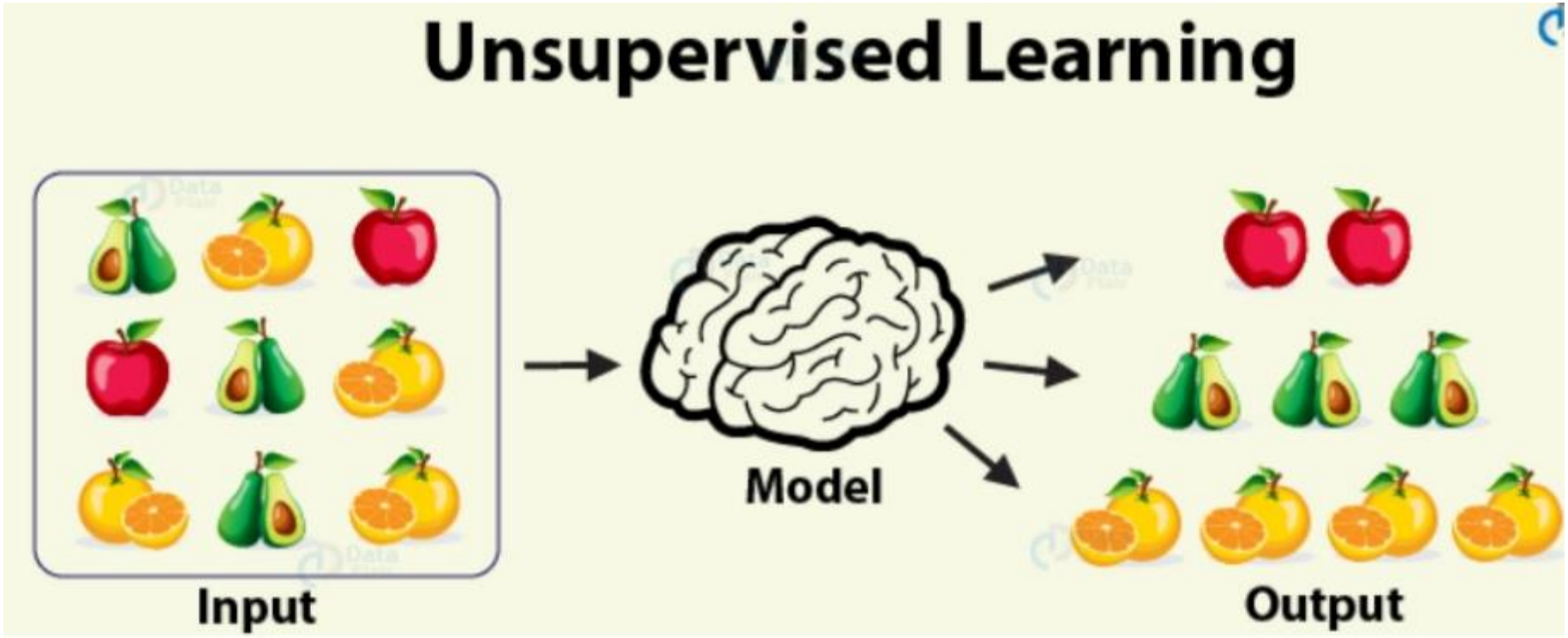
NON SUPERVISIONATO



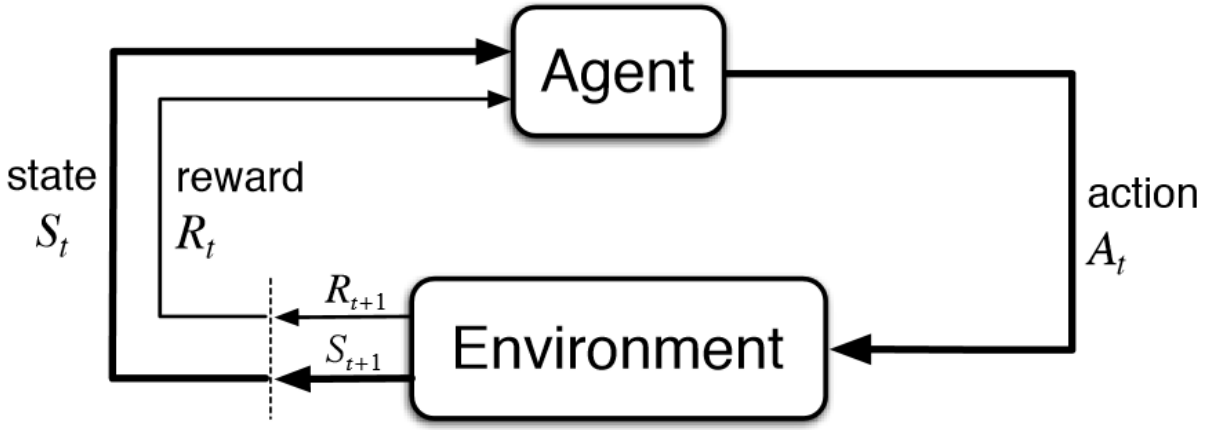

Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Model training with labelled data
Model training with unlabelled data
Model take actions in the environment then received state updates and feedbacks
Classification
Regression
Clustering
Environment
action
feedback
state
Model Agent

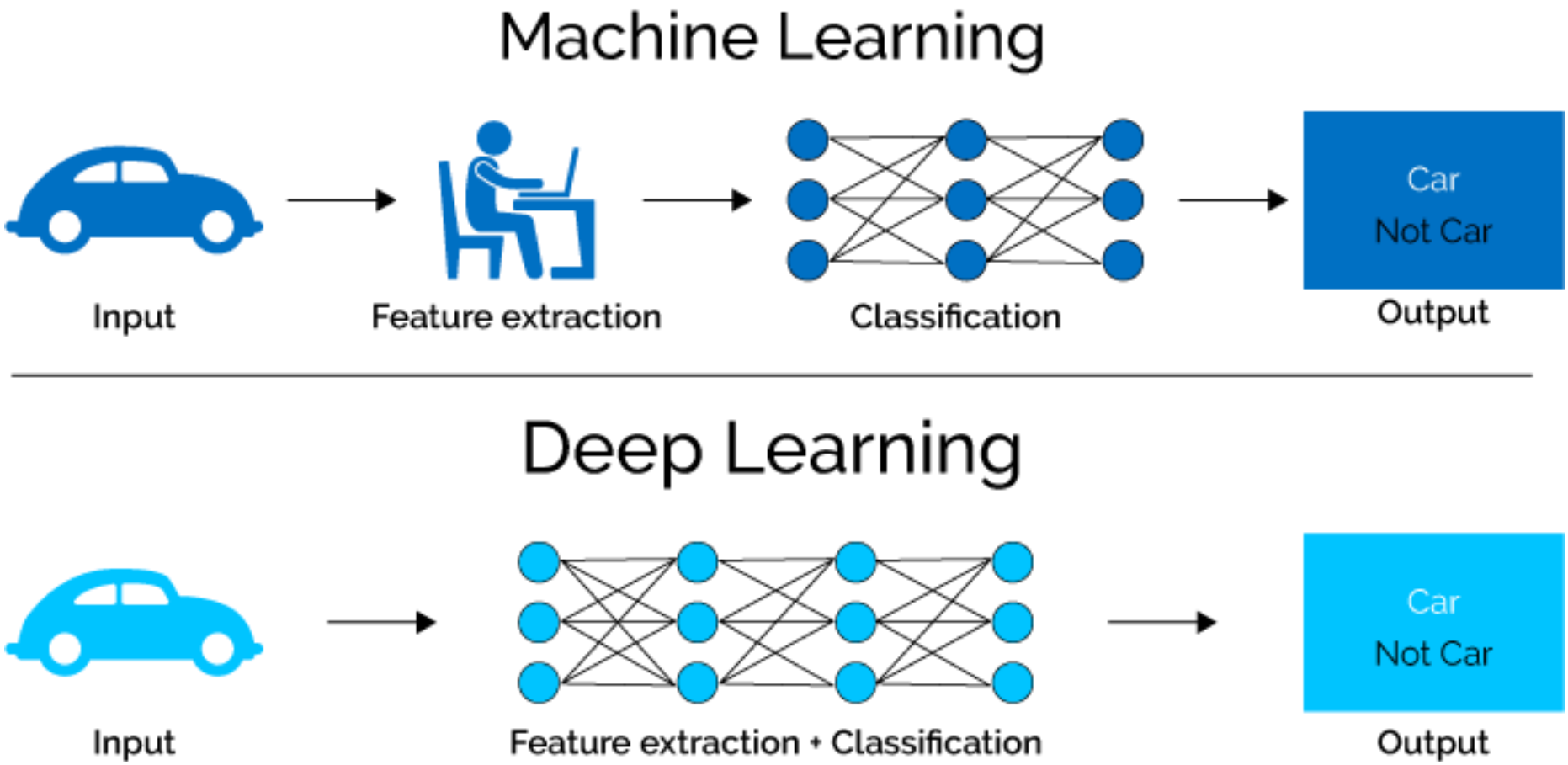

Types of machine learning
Supervised learning
Unsupervised learning
Reinforcement learning
Linear classifier/ regression
Support vector machine
Decision trees
Neural networks
k-nearest neighbors
Clustering algorithms
Agglomerative clustering
K-means clustering
Density-based spatial clustering of applications with noise (DBSCAN)
State–action–reward–state–action SARSA -lambda
Deep Q Network (DQN)
Deep deterministic policy gradient (DDPG)


Input
Supervised Learning
Model
It's Grapes
Annotations
These are grapes
Prediction


Unsupervised Learning
Input
Model
Output


Reinforcement Learning in ML
Input Raw Data
Environment
Output
Reward
Best Action
State
Selection of Algorithm
Agent


Agent
state
S_t
reward
R_t
action
A_t
R_{t+1}
S_{t+1}
Environment


Machine Learning
Input
Feature extraction
Classification
Car
Not Car
Output
Deep Learning
Input
Feature extraction + Classification
Car
Not Car
Output

AI

Unsupervised

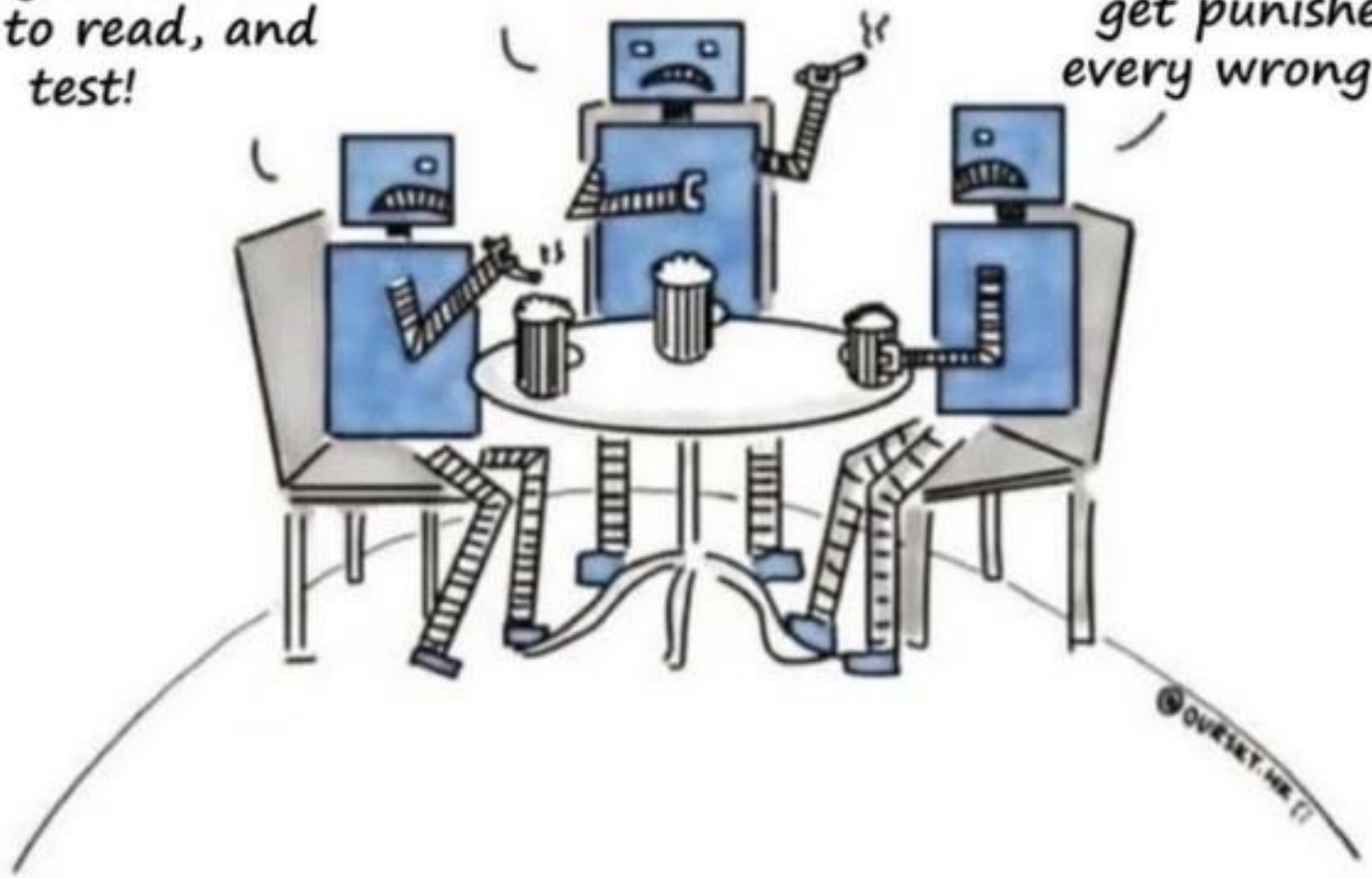
Me too. But at least they told you the answers

Supervised

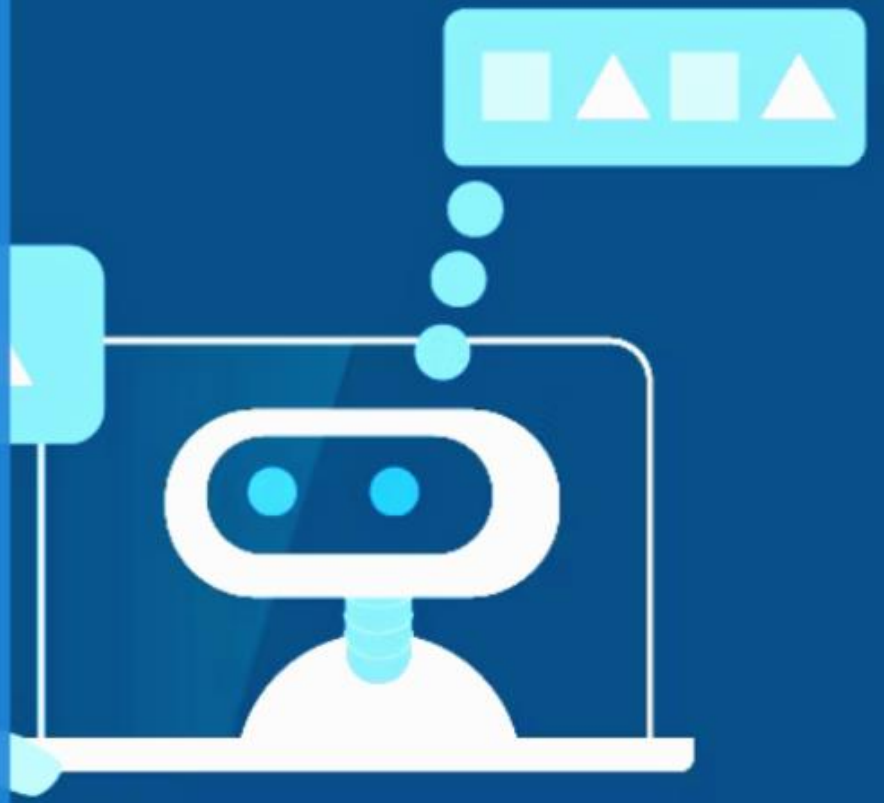
They gave me so much to read, and test!

Reinforcement

At least you all don't get punished for every wrong action



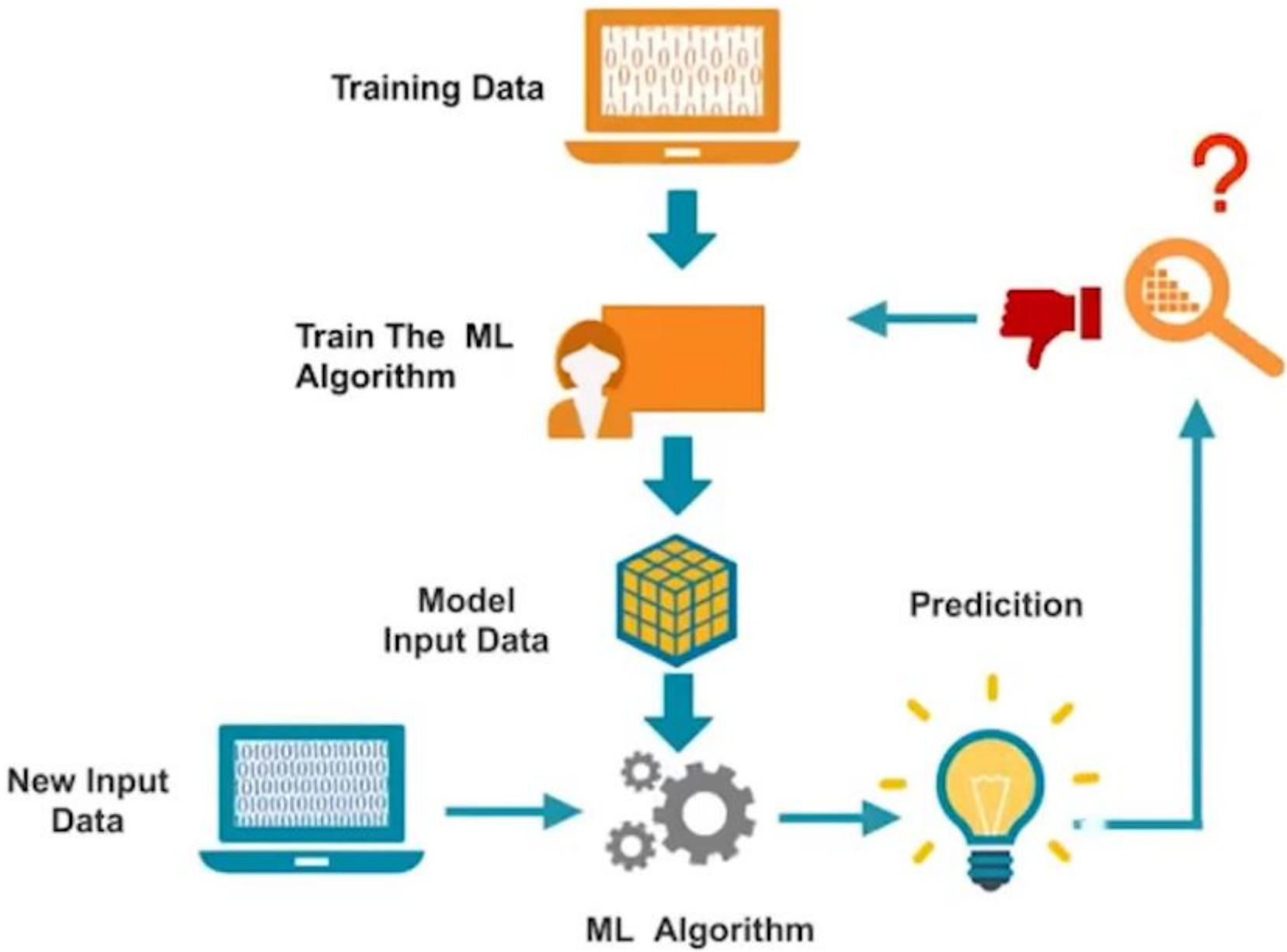
Cos'è un modello di Machine Learning



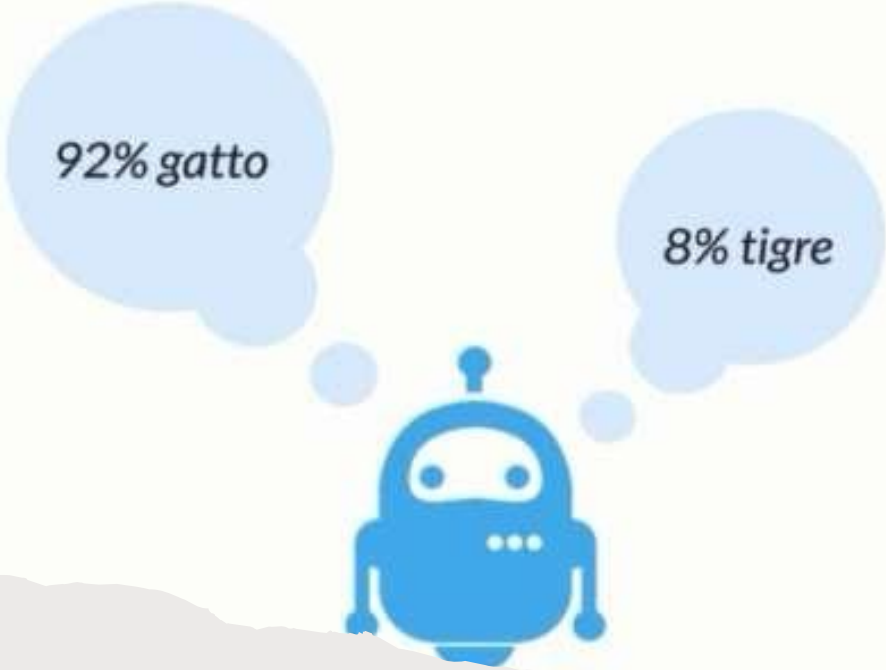
Un modello di machine learning è un'astrazione dei dati e dei pattern presenti in essi.

Funziona come una rappresentazione matematica o statistica delle relazioni intrinseche nei dati, consentendo al sistema di generalizzare i nuovi dati non ancora visti durante il processo di addestramento.

Questa astrazione permette al modello di apprendere dai dati e di adattarsi a diverse situazioni, rendendolo flessibile ed efficiente nell'affrontare problemi complessi


Training Data
Train The ML Algorithm
Model Input Data
Predicition
New Input Data
ML Algorithm

Il machine learning si basa su statistica e probabilità



NON è DETERMINITSTICO
COME GLI ESSERI UMANI

1997 - Deep Blue batte il campione di scacchi Garry Kasparov

2008 - Watson vince a Jeopardy! contro i campioni Ken Jennings e Brad Rutter

2016 - AlphaGo batte Go il 18 volte campio del mondo Lee Seldo

PERCHE' OGGI ?

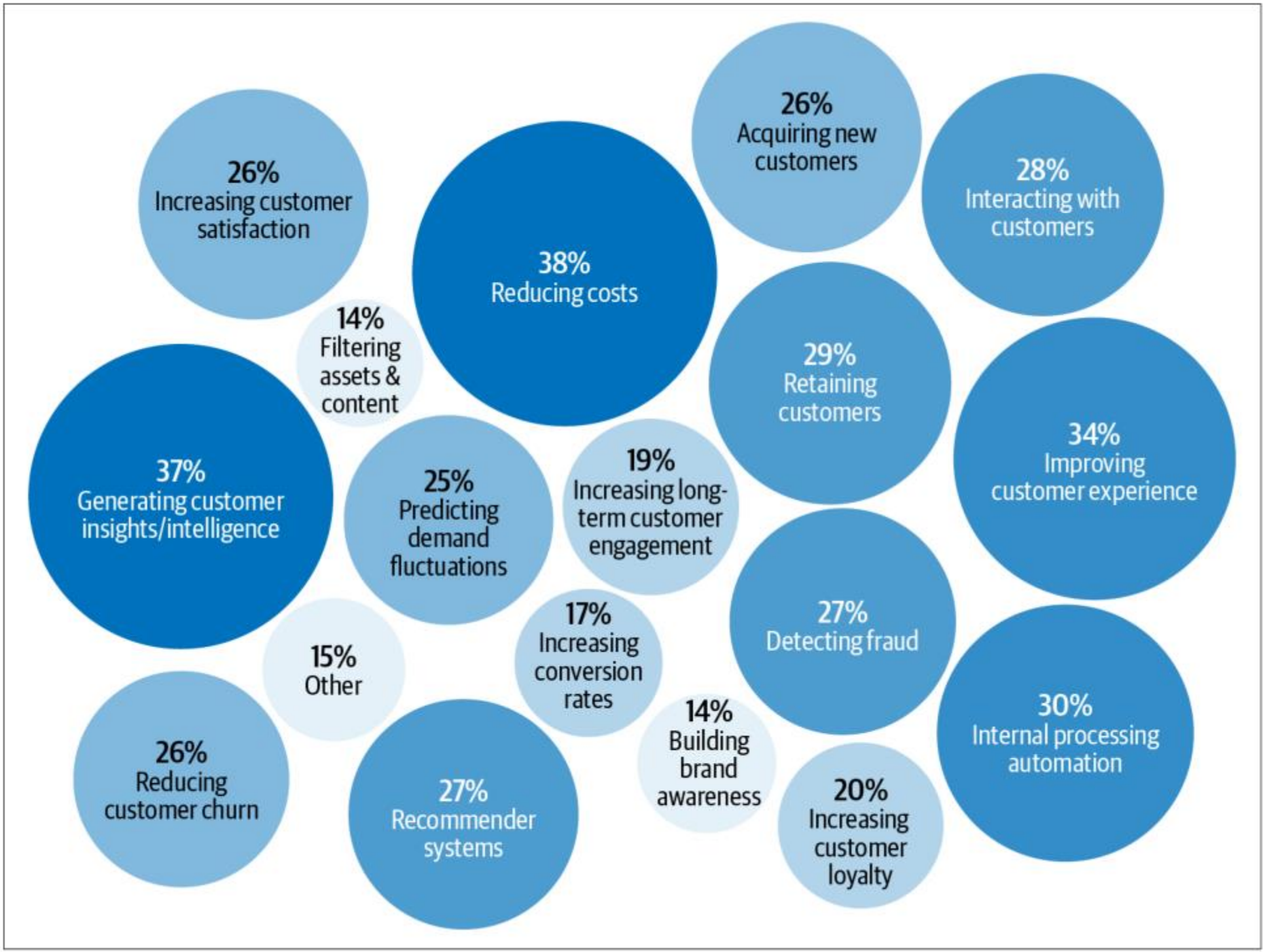
MAGGIORE POTENZA DI CALCOLO DISPONIBILE

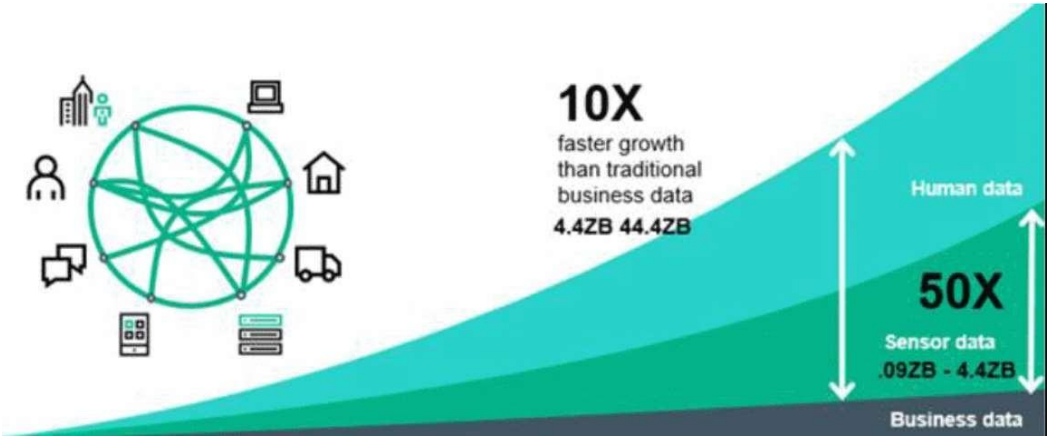
DISPONIBILITA' DI ENORMI QUANTITA' DI DATI

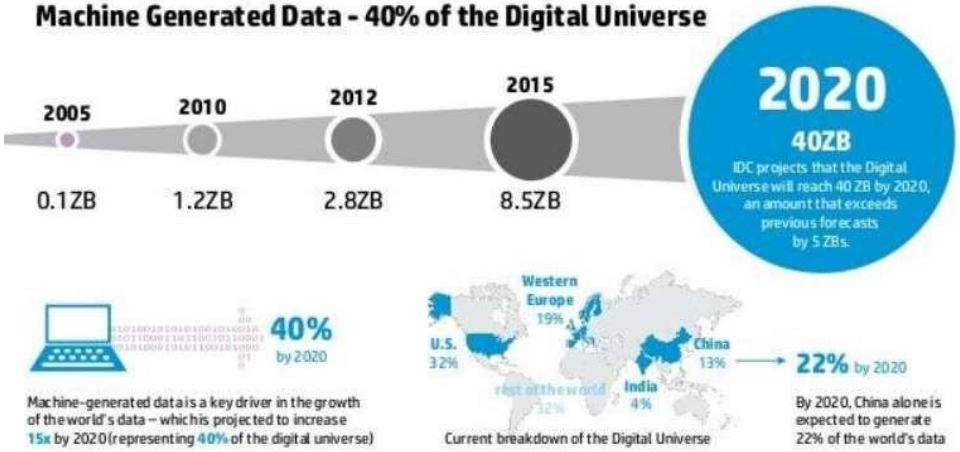
Perché lo si utilizza?

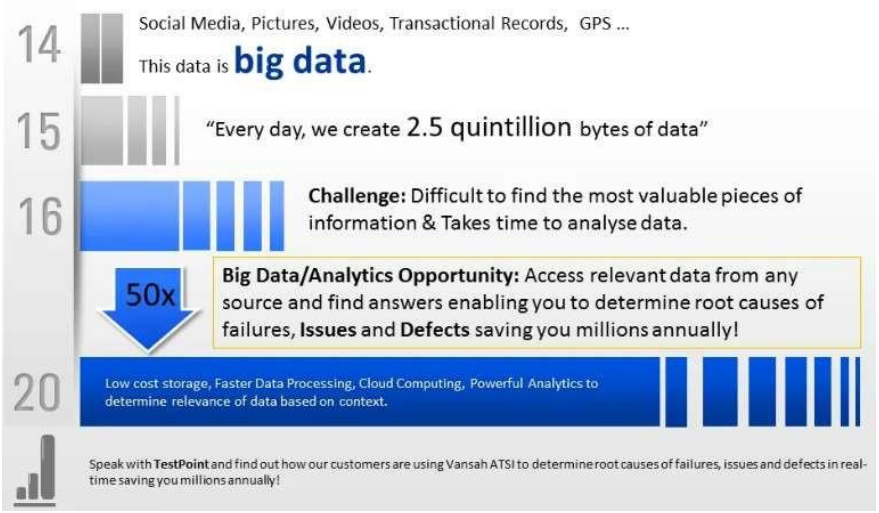
Il machine learning aiuta le aziende a dare un senso ai loro dati, indipendentemente dalla loro dimensione e dal settore di mercato in cui operano.

- Comprendere velocemente le informazioni
- Mostrare i fenomeni che determinano certi andamenti
- Evidenziare i trend emergenti
- Identificare relazioni e pattern nascosti
- Condividere le proprie scoperte con gli altri

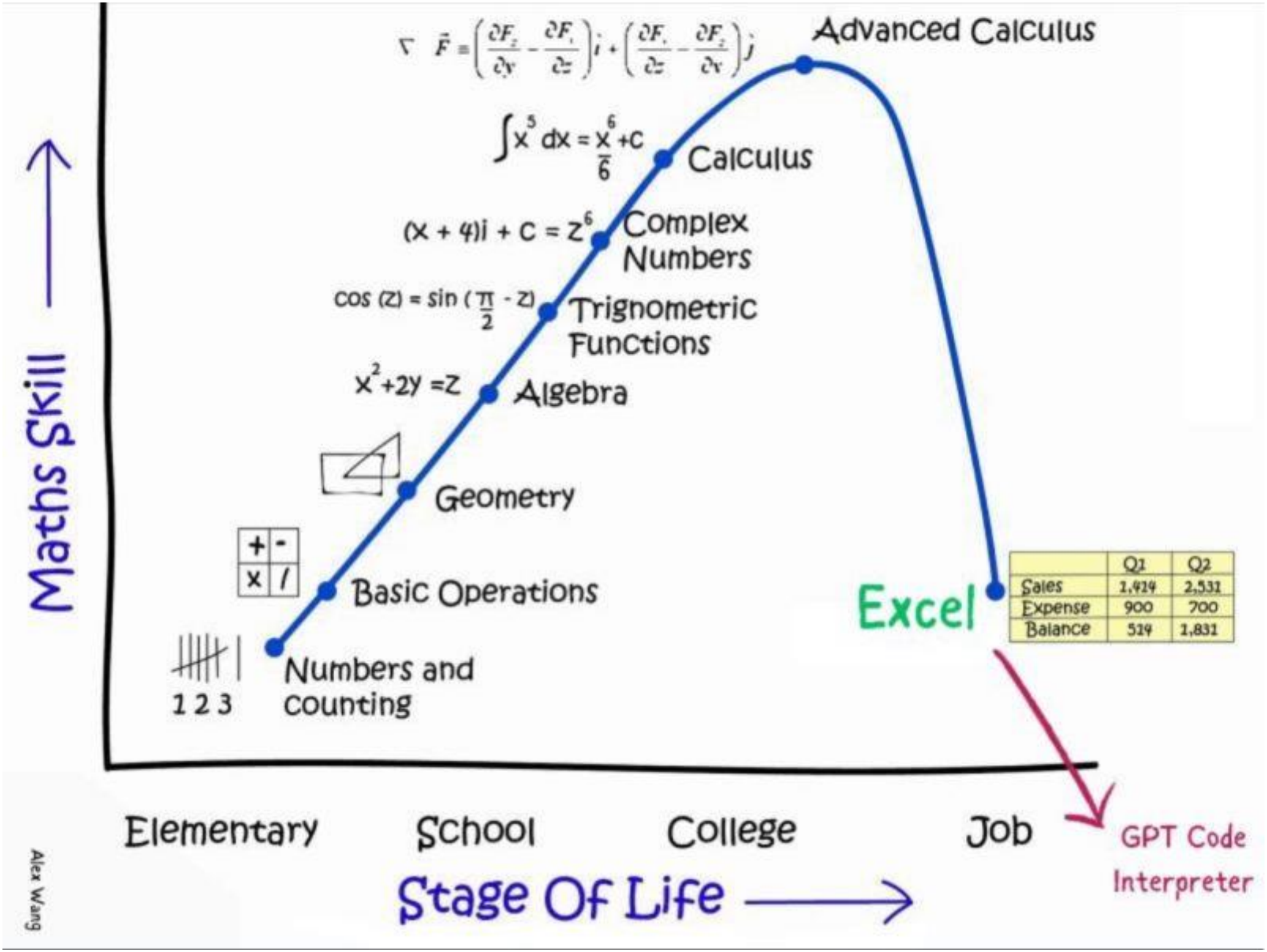

26%
Increasing customer satisfaction
26%
Acquiring new customers
28%
Interacting with customers
38%
Reducing costs
14%
Filtering assets & content
29%
Retaining customers
34%
Improving customer experience
37%
Generating customer insights/intelligence
25%
Predicting demand fluctuations
19%
Increasing long-term customer engagement
27%
Detecting fraud
15%
Other
17%
Increasing conversion rates
30%
Internal processing automation
26%
Reducing customer churn
27%
Recommender systems
14%
Building brand awareness
20%
Increasing customer loyalty


10X
faster growth
than traditional
business data
4.4ZB 44.4ZB
Human data
50X
Sensor data
.09ZB - 4.4ZB
Business data


Machine Generated Data - 40% of the Digital Universe
2005
2010
2012
2015
2020
40ZB
IDC projects that the Digital Universe will reach 40 ZB by 2020, an amount that exceeds previous forecasts by 5 ZBs.
0.1ZB
1.2ZB
2.8ZB
8.5ZB
40%
by 2020
Machine-generated data is a key driver in the growth of the world's data – which is projected to increase 15x by 2020 (representing 40% of the digital universe)
Western Europe 19%
U.S. 32%
China 13%
India 4%
rest of the world 32%
Current breakdown of the Digital Universe
22% by 2020
By 2020, China alone is expected to generate 22% of the world's data


14
Social Media, Pictures, Videos, Transactional Records, GPS ...
This data is big data.
15
"Every day, we create 2.5 quintillion bytes of data"
16
Challenge: Difficult to find the most valuable pieces of information & Takes time to analyse data.
50x
Big Data/Analytics Opportunity: Access relevant data from any source and find answers enabling you to determine root causes of failures, Issues and Defects saving you millions annually!
20
Low cost storage, Faster Data Processing, Cloud Computing, Powerful Analytics to determine relevance of data based on context.
Speak with TestPoint and find out how our customers are using Vansah ATSI to determine root causes of failures, issues and defects in real-time saving you millions annually!

?


Maths Skill
Advanced Calculus
$\nabla \vec{F} \equiv \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right)i + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right)j$
$\int x^5 dx = \frac{x^6}{6} + C$
Calculus
$(x + 4)i + C = z^6$
Complex Numbers
$\cos(z) = \sin\left(\frac{\pi}{2} - z\right)$
Trignometric Functions
$x^2 + 2y = z$
Algebra
Geometry
+ - x /
Basic Operations
1 2 3
Numbers and counting
Excel
Q1 Q2
Sales 1,919 2,531
Expense 900 700
Balance 519 1,831
Elementary
School
College
Job
GPT Code Interpreter
Stage Of Life
Alex Wang

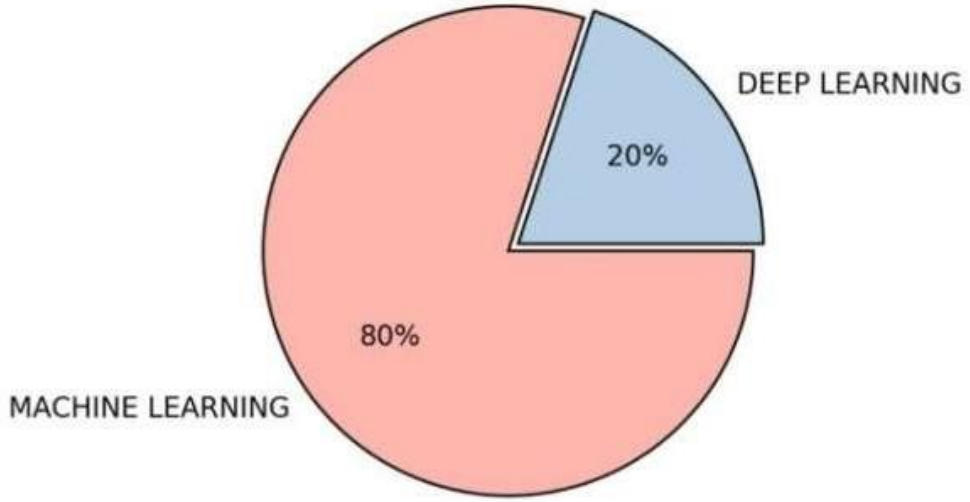
La migliore maestra di vita è l'ESPERIENZA...
Ti costa cara, ma ti SPIEGA BENE.

Alda Merini


Data science is most of the time preprocessing data

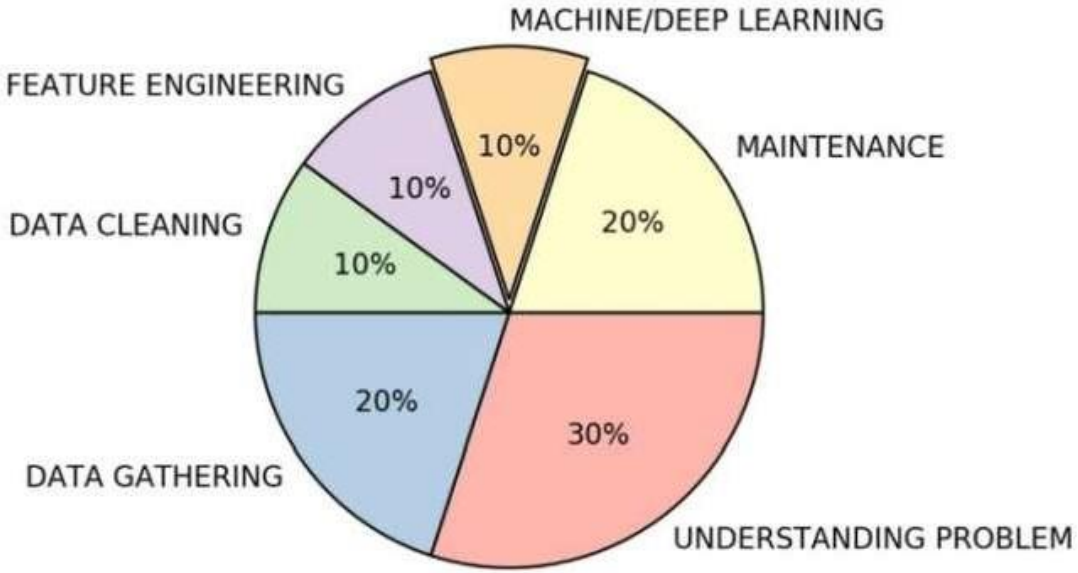
DATA SCIENTIST JOB - EXPECTATION

@drangshu



Follow: Dr. Angshuman Ghosh

DATA SCIENTIST JOB - REALITY



Understanding the business problem

- You should ask relevant questions which makes you understand the problem which you are going to solve
- You should ask multiple WHY? questions and get answers from the client or the stakeholder or the person who told you to do the project.

Data acquisition

- After deciding what features or metrics to use to solve the business problem.
- Next step is to gather the data.
- You may use sources like Databases, API's, Web scraper, online repositories etc...

Data preparation

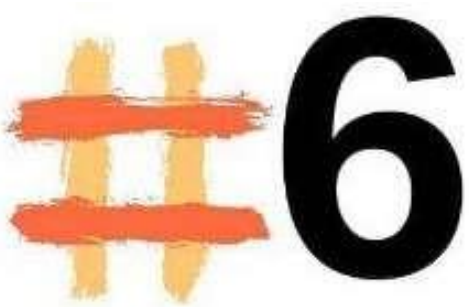
- This step involves 2 important things Data cleaning, Data transformation.
- Data cleaning is like check missing values, inconsistency datatypes, duplicate values etc.. (Check our post on data pre-processing to see what are the most used techniques)
- Data transformation is a process of modifying the data based on predefined rules.

Exploratory data analysis

- EDA helps you to understand what exactly you can do with the data.
- This is the most important step.
- Through EDA you can find what features are the most important in the model building.
- You can also find useful insights through EDA.

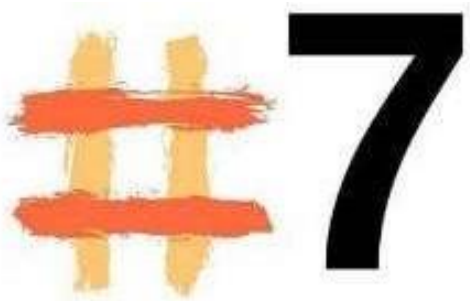
Data modeling

- This is the most important part where you will be finding the model the best fits the business requirement.
- You will be doing multiple iterations on the test and train data to find the best performing model.



Visualization and communication

- This is where you will show all the things which you did and fond during the previous steps to your client, stakeholders or the person who gave you the project.
- You will be creating reports or dashboards to show your business finding in a powerful way (visualizations) to make them understand easily.



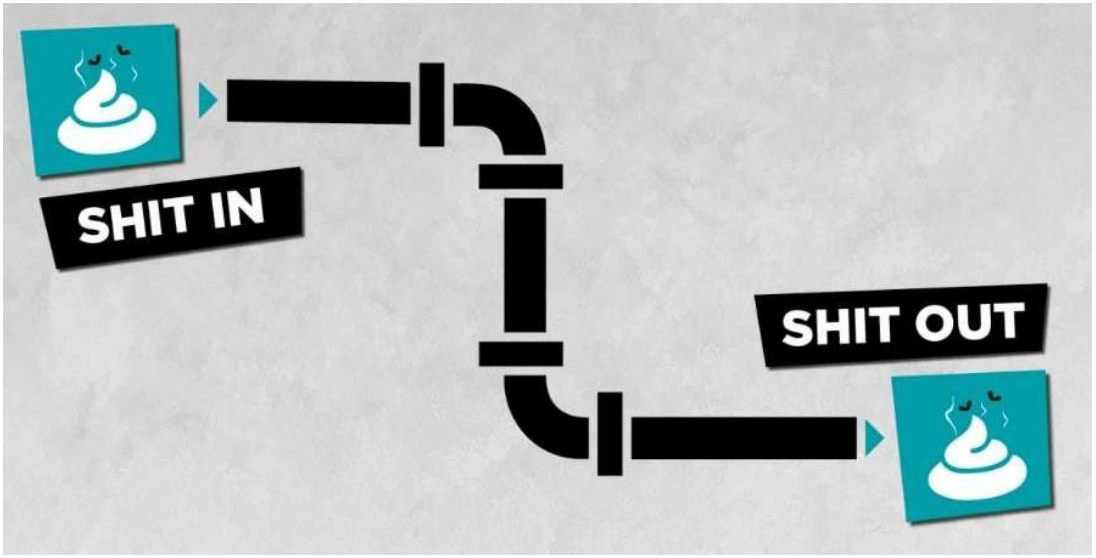
Deploy & maintenance

- Test your best performing model multiple times before deploying it into production.
- You will be using reports and dashboards for real-time analytics.
- It is also important to monitor the model performance in the real world and retraining it if the performance degrades.

Requirements:

- Comprendere i dati che desideri elaborare, incluse le loro dimensioni e cardinalità (l'unicità dei valori dei dati in una colonna).
- Determinare cosa stai processando e quale tipo di informazioni vuoi estrapolare.
- Conoscere la tua audience e comprendere come elabora l'informazione, o cosa si aspetta.
- Utilizzare una visualizzazione che trasmetta le informazioni nella maniera più appropriata e semplice per la tua audience.


AI
COMPANIES
AI
POOR
DATA QUALITY
COMPANIES
imgflip.com

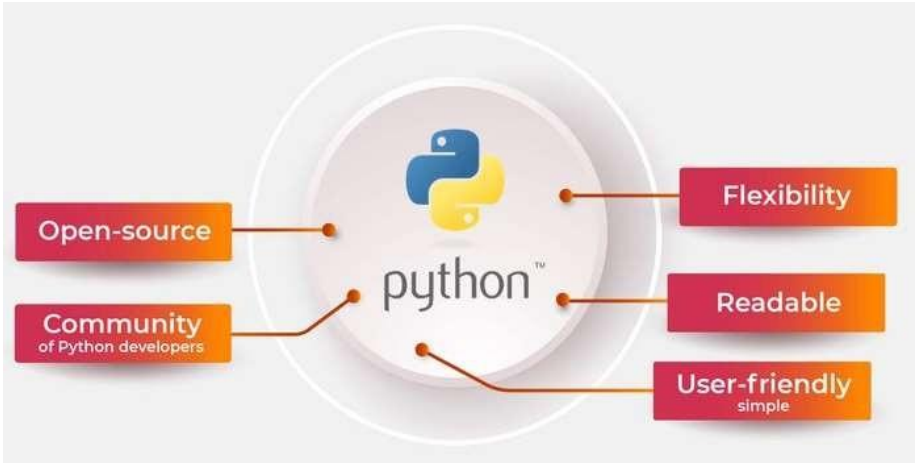

SHIT IN
SHIT OUT


CAUTION: BAD DATA
BAD DATA QUALITY MAY RESULT IN FRUSTRATION AND LEAD TO DROP KICKING YOUR COMPUTER

Evitare Errori Grossolani!

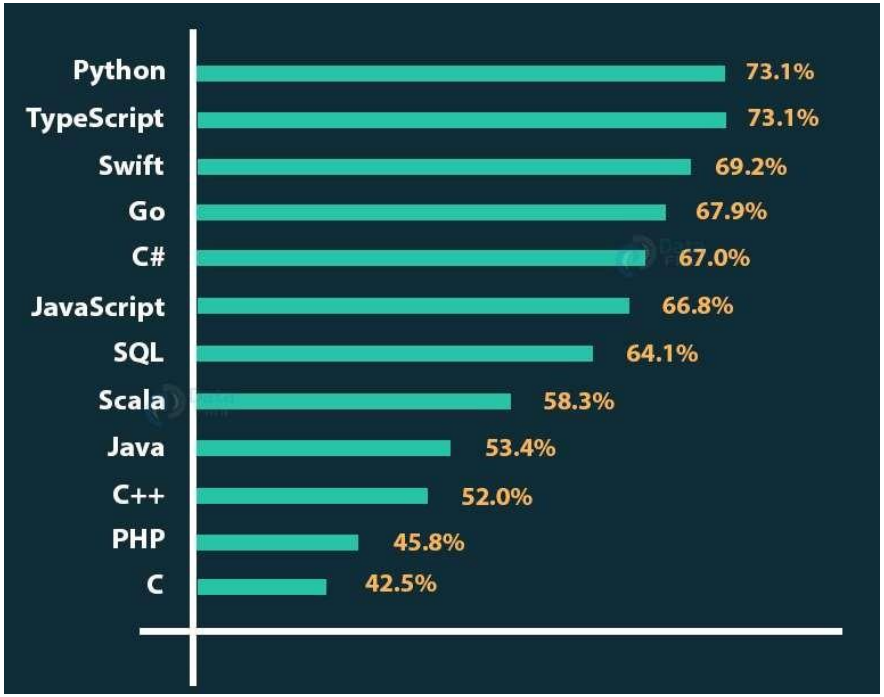

Digitalizzazione dati
Generazione automatica dati
Organizzazione dati
Integrazione metadati
Individuazione Outlier
Stima dati mancanti
Analisi dati Estrazione informazioni
Presentazioni informazioni
Sistemi decisionali

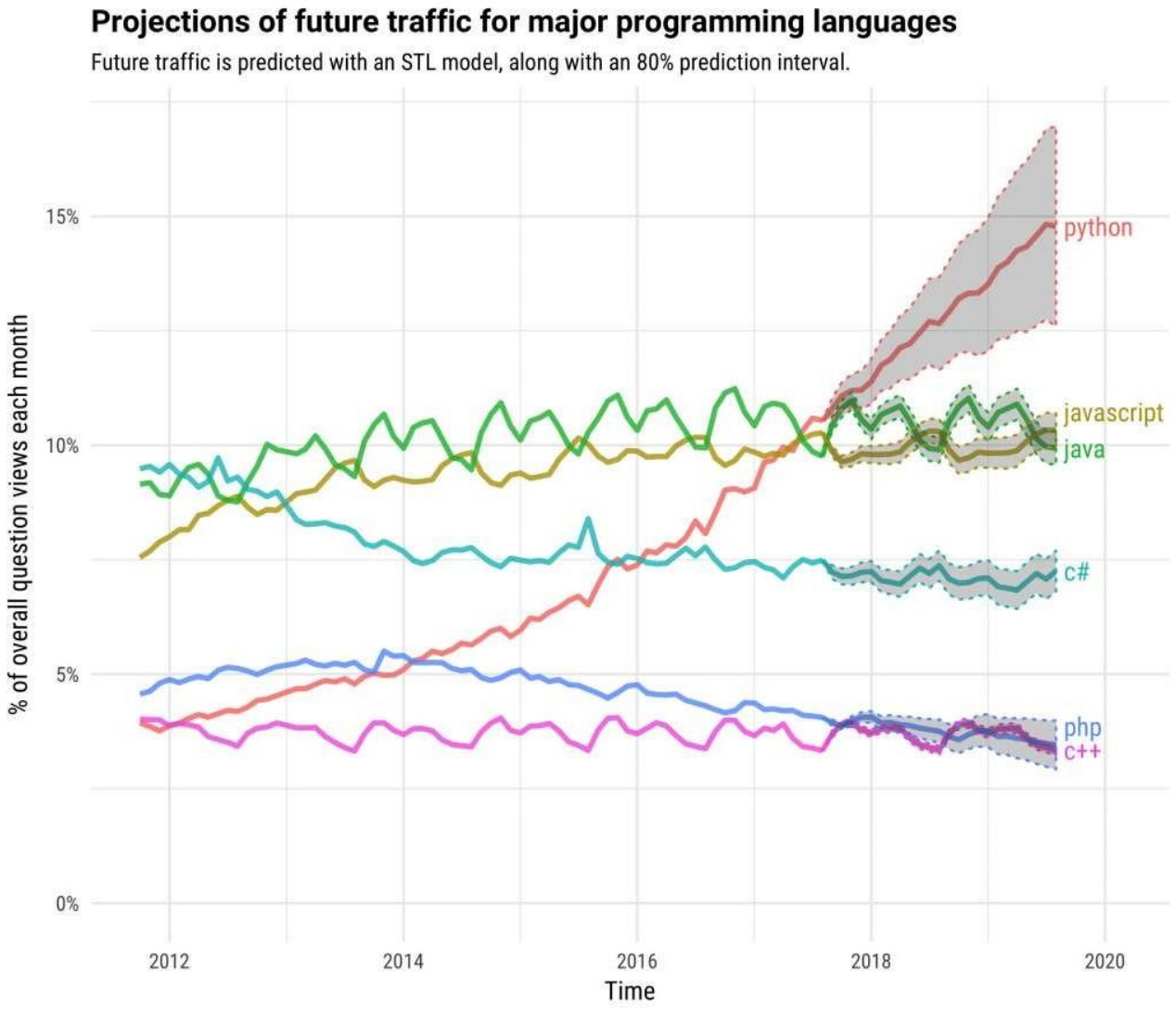

WHAT IS PYTHON?
A back end programming language
High-level & approachable for beginners
Has a welcoming & established community
Used for tasks like:
Web Development
Scripting
Web Scraping
Data Analysis
Automation
Used by companies like:
Used with frameworks like:
django
Flask
jupyter

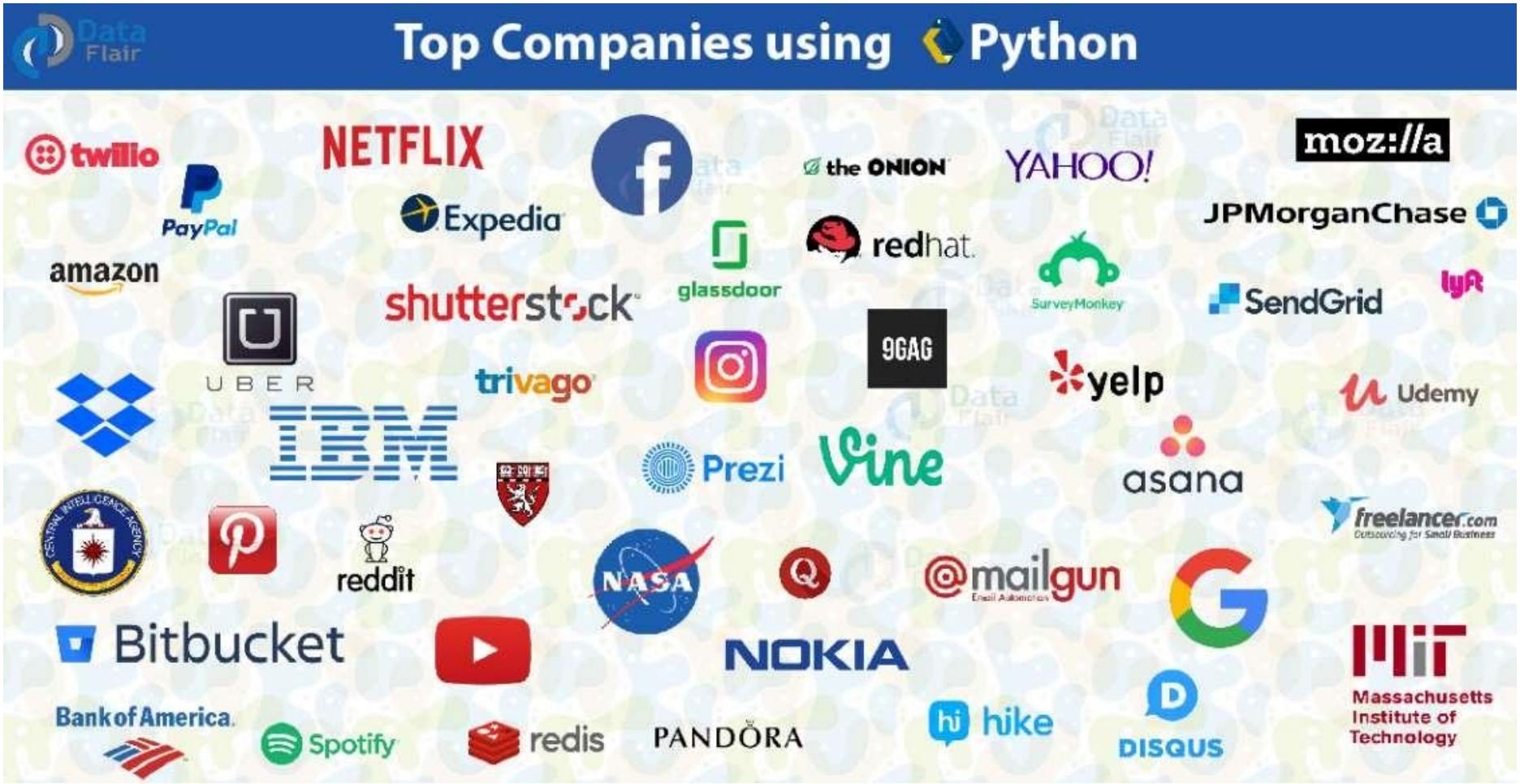

python™
Open-source
Community
of Python developers
Flexibility
Readable
User-friendly
simple

PYTHON

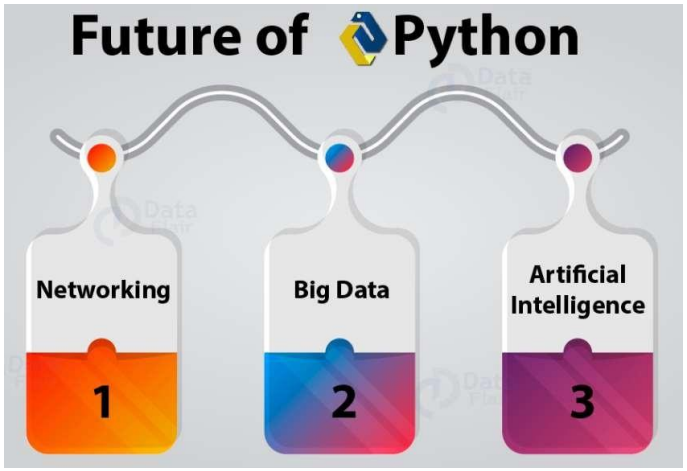

PROGRAMMING
JAVA SCRIPT
PYTHON
C# OBJECTIVE-C
GO C++ RUBY
JAVA
.COM
ARDUINO
PHP
CSS3
SWIFT
HTML5


Python 73.1%
TypeScript 73.1%
Swift 69.2%
Go 67.9%
C# 67.0%
JavaScript 66.8%
SQL 64.1%
Scala 58.3%
Java 53.4%
C++ 52.0%
PHP 45.8%
C 42.5%

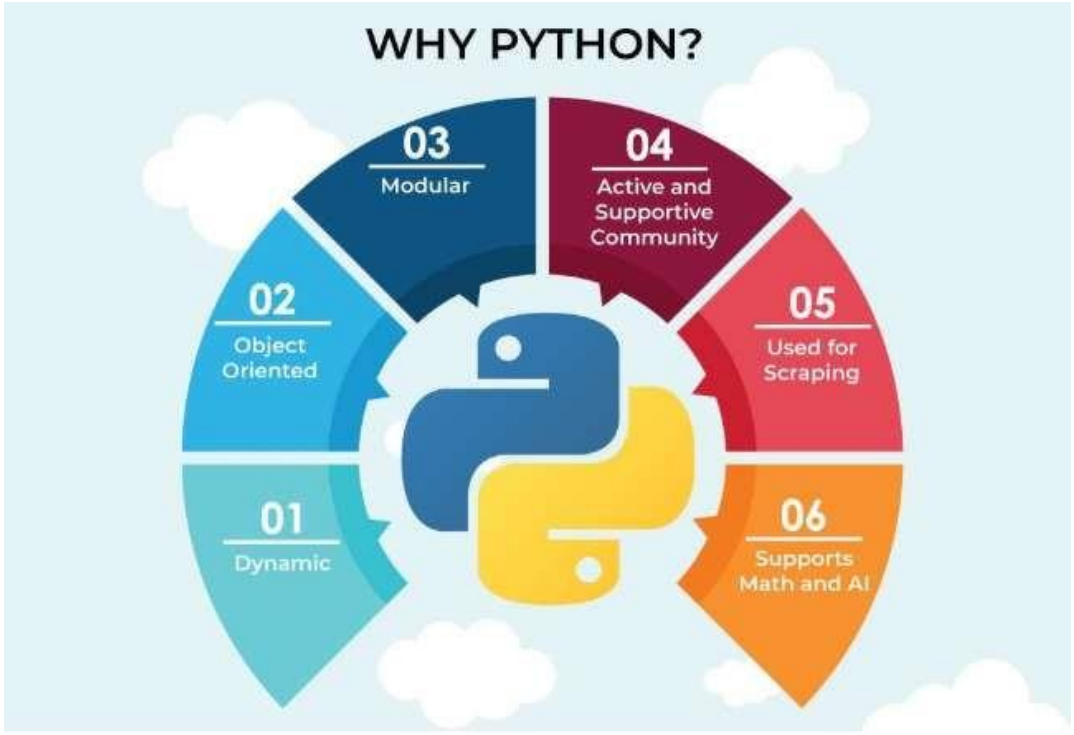

Projections of future traffic for major programming languages
Future traffic is predicted with an STL model, along with an 80% prediction interval.
% of overall question views each month
15%
10%
5%
0%
2012
2014
2016
2018
2020
Time
python
javascript
java
c#
php
c++


Data Flair
Top Companies using Python
twilio
NETFLIX
the ONION
YAHOO!
mozilla
PayPal
Expedia
JPMorganChase
redhat.
amazon
glassdoor
SurveyMonkey
SendGrid
lyft
shutterstock
UBER
9GAG
trivago
yelp
Udemy
IBM
Prezi
Vine
asana
freelancer.com
Outsourcing for Small Business
reddit
NASA
mailgun
Bitbucket
NOKIA
Massachusetts Institute of Technology
Bank of America
Spotify
redis
PANDORA
hike
DISQUS


WHY
Python
01
Simplicity
02
Large Community
03
High
Demand-Supply Ratio
04
Large
Number of Frameworks
05
Chosen Language
for AI and ML
06
Make your own DIYs


Future of Python
Networking
1
Big Data
2
Artificial
Intelligence
3


X
python
Lines of Code: 105
Wall Time: 7.89s
More complex
Lines of Code: 10
Wall Time: 0.83s
More simple



- **Python is Easy**
- **More Functions – Less Code**
- **Perfect Language for Building Prototypes**
- **Great Flexibility**
- **Perfect Language in case you're on a Budget**
- **The Internet of Things (IoT) + Python Combination**
- **A Lot of useful Frameworks (Django,Flask)**

"Python Vs Java!"

"Python Vs C++!"

Python

```
Print ("Python Vs Java!")
```

```
Print ("Python Vs C++!")
```

Java

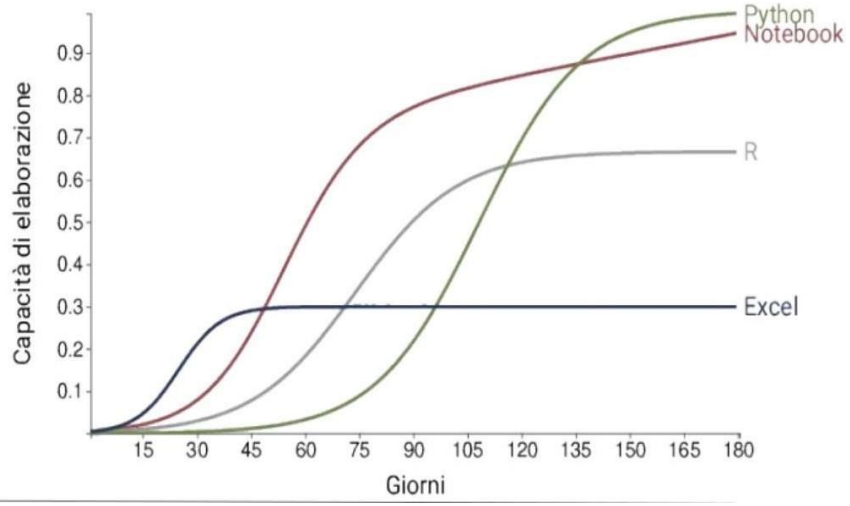
```
public class PythonVsJava {
public static void main(String[] args)
{
System.out.println("Python Vs Java!");
}
}
```

C++

```
#include <iostream>
void main()
{
cout << "Python Vs C++!";
}
```

Gratificazione e soddisfazione

Noia e frustrazione




Learn
programming
for **future work**
Learn
programming
to understand
programming jokes

COESIONE ITALIA 21-27
EMILIA-ROMAGNA

Cofinanziato dall'Unione europea

Regione Emilia-Romagna

DATA LAB

GUARDA AVANTI

Big Data, nuove competenze per nuove professioni.

www.bigdata-lab.it

FORM INDUSTRIA EMILIA ROMAGNA

ifoa Sapere utile

REISS ROMOLI la passione della conoscenza

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

Università degli Studi di Ferrara

UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

UNIVERSITÀ DI PARMA

POLITECNICO MILANO 1863 POLO TERRITORIALE DI PIACENZA

UNIVERSITÀ CATTOLICA del Sacro Cuore