smau

Algoritmi per costruire la Intelligenza Artificiale

Ph.D. Roberto Marmo

Data Scientist, Consulenza e Formazione
in analisi dati e soluzioni con Intelligenza Artificiale

http://www.robertomarmo.net info@robertomarmo.net

https://www.linkedin.com/in/robertomarmo/it

Alcune mie realizzazioni

Riconoscere oggetti in immagine

Analisi e ricerca documenti

Motori di raccomandazione

Analisi dati da Social Media, OSINT

Automotive

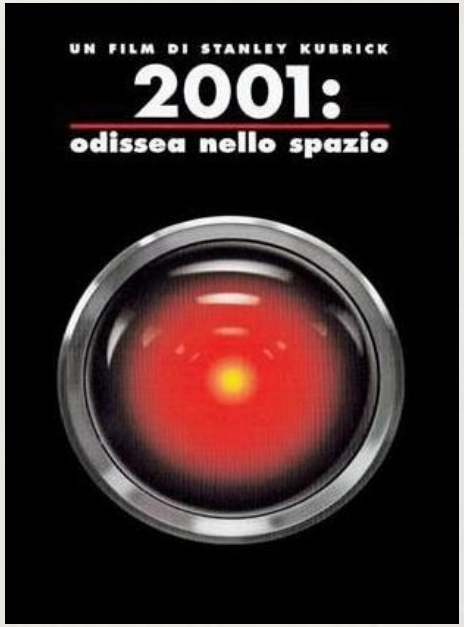
Videosorveglianza

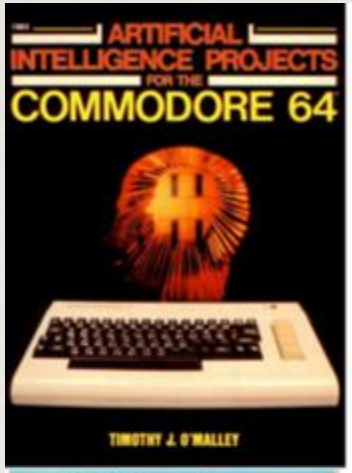
Database

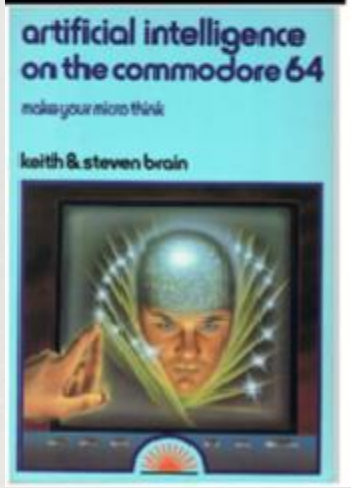
Formazione

Analisi Anti frode

Miei primi passi con IA anni '80







Roberto Marmo con editore

www.algoritmiia.it

https://robertomarmo.net/Libri.html#top

Agenda

1. Definizioni di IA
2. Logica Fuzzy
3. Algoritmo Genetico
4. Machine Learning
5. Apprendimento
6. Albero Decisione
7. Rete Neurale
8. Trovare dati per apprendere
9. Gestione del progetto
10. Prompt per ChatGPT

crea immagine di un essere umano curioso che guarda dentro al funzionamento della intelligenza artificiale per capire come usarla in modo efficiente

1. Intelligenza Umana

Una persona intelligente:

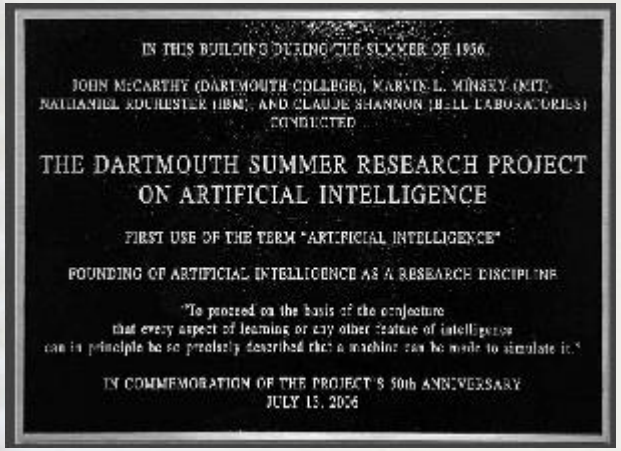
- comprende il contesto in cui si trova per adattarsi
- risolve problemi di varia tipologia
- impara dall'esperienza e non ripete errori
- sa spiegare la soluzione
- interagisce con altre persone
- agisce per realizzare un obiettivo che migliora la vita
- è flessibile e creativa

1. Intelligenza Artificiale

Intelligenza Artificiale (IA) è l'abilità del computer nel mostrare capacità umane quali il ragionamento, l'apprendimento, la pianificazione e la creatività.

Non si dovrebbe capire la differenza tra opera creata da Intelligenza Umana e quella da IA.

The Birthplace of AI, 1956 Dartmouth Workshop.



https://www.cantorsparadise.com/the-birthplace-of-ai-9ab7d4e5fb00

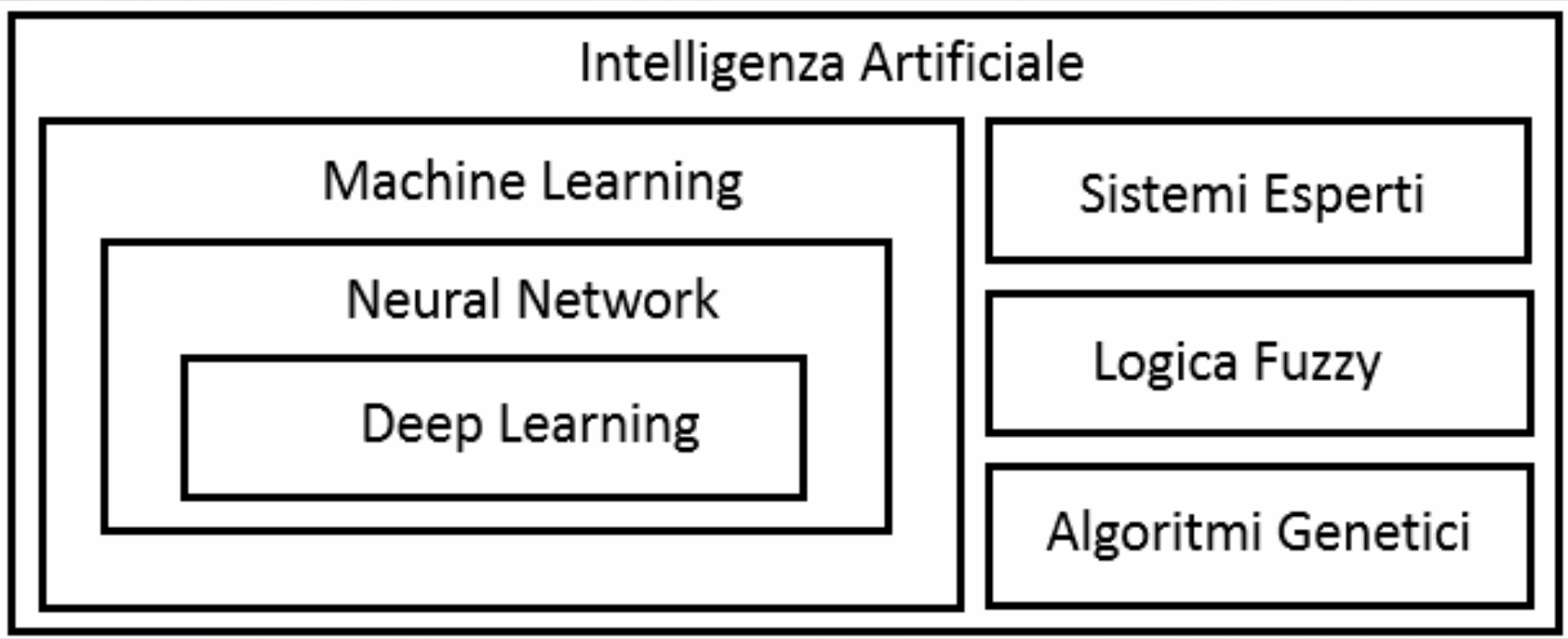
IA Super: supera l’essere umano

IA Forte o Generale: risolve qualsiasi problema

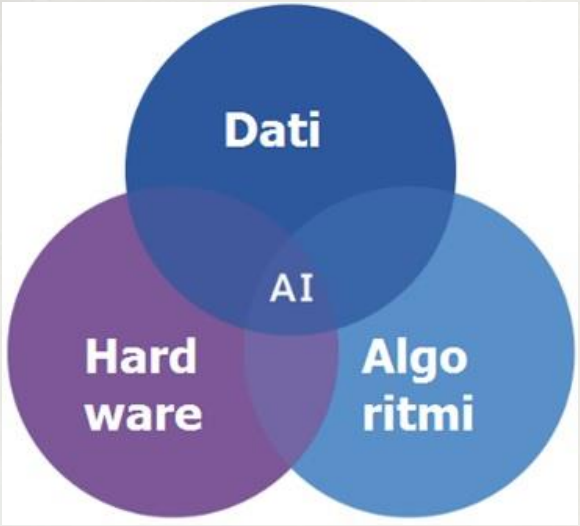
IA Debole: risolve problemi molto specifici

IA Generativa: crea testo, video, immagine

1. Intelligenza Artificiale



1. Intelligenza Artificiale



- formule di matematica, statistica, probabilità
- concetti da filosofia, psicologia, sociologia
- linguaggi di programmazione
- grande quantità di dati da varie fonti
- esperienza degli esperti umani

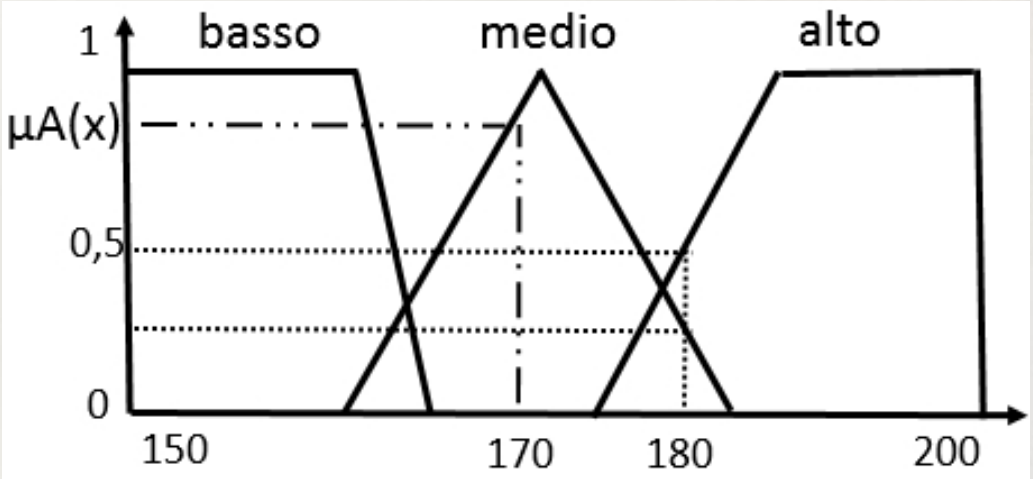
è intelligente!

Numeri

Numeri

2. Logica Fuzzy

Analogia con ragionamento umano che usa vero, falso e tutte le sfumature intermedie.



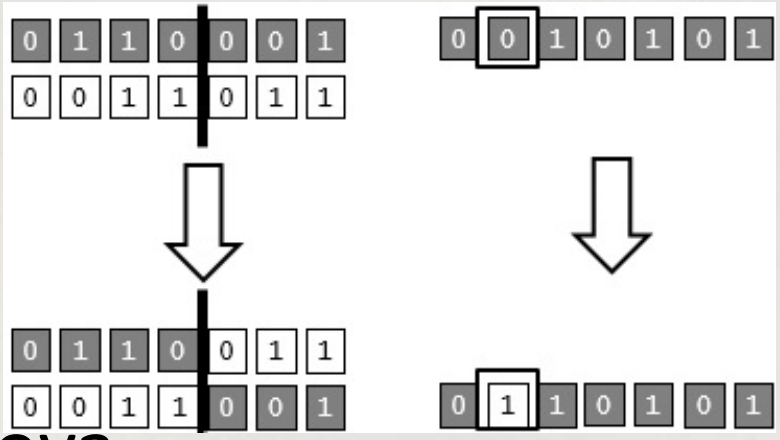
Esempi:

- robot prende un oggetto con più o meno forza
- lavatrice che gira il cestello più o meno forte
- dove fermare il treno e con quale forza

3.Algoritmo genetico

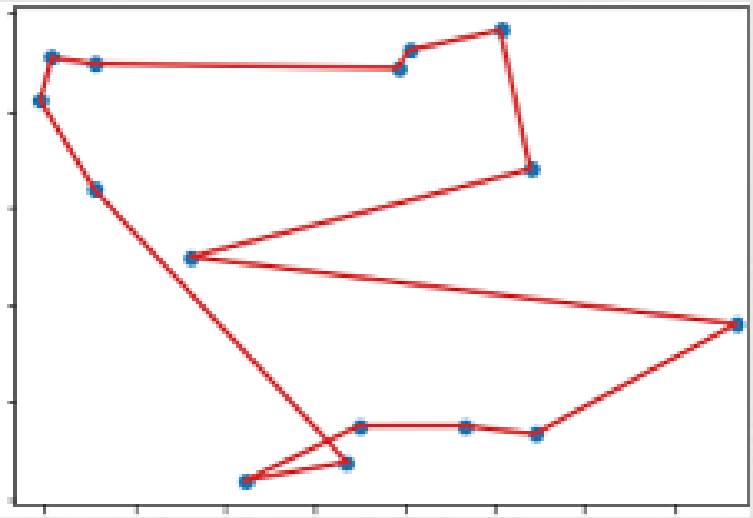
Analogia con selezione naturale in campo genetico:

1. creare due soluzioni al problema
2. Incrociare due soluzioni o cambiare un numero nella soluzione per creare una nuova soluzione



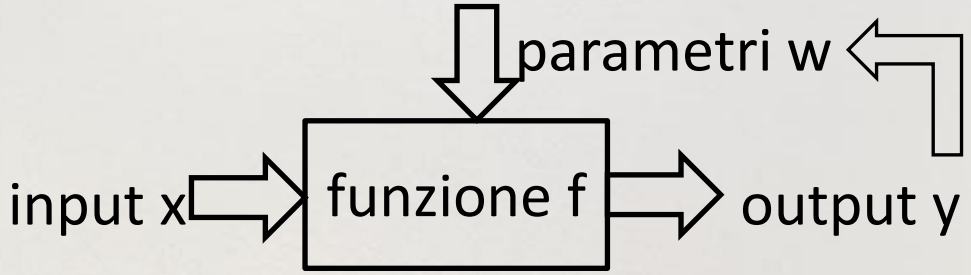
Esempi:

- percorso in N città con minor spesa possibile
- collocare il personale sui compiti
- diminuire il costo di un impianto
- riempire al massimo il container



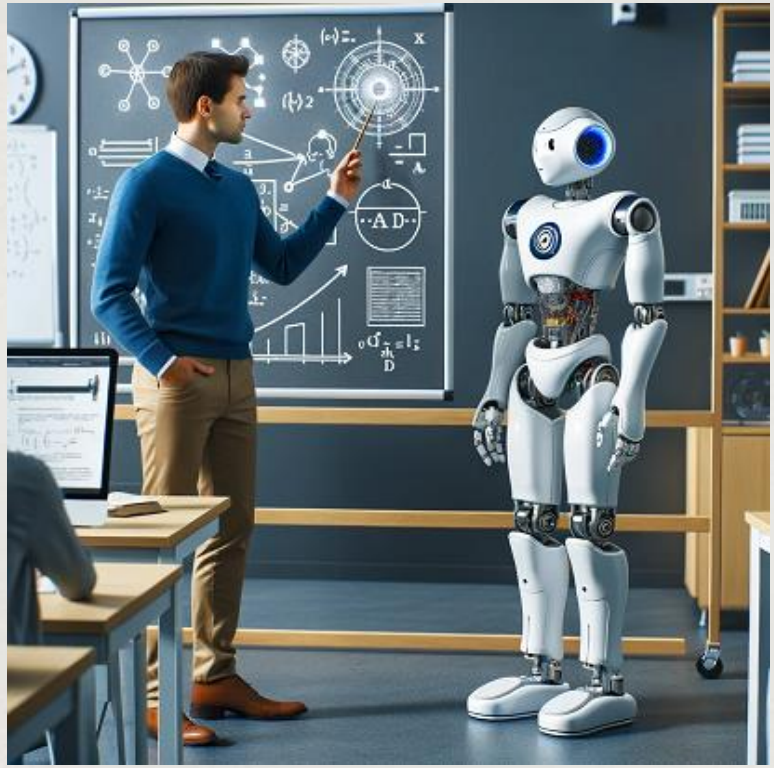
4. Machine Learning

- Machine Learning è una funzione matematica f che, applicata al dato ingresso x, predice il dato uscita y = f(x,w) dove w è un insieme di parametri con valori calcolati da algoritmo di ottimizzazione che ne prova tanti valori fino a trovare quella che crea errore =|y-f(x)| come differenza accettabile tra risultato desiderato e risultato calcolato
- ruolo dell'operatore umano:
  - scelta dei dati (x,y)
  - scelta del modello f
  - scelta iniziale dei parametri w
  - scelta criterio per giudicare un errore e cambiare i parametri w
- fare machine learning: cambiare parametri w in modo dinamico



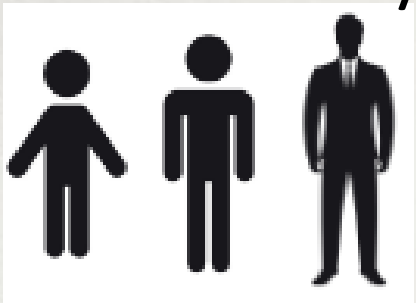
5. Apprendimento supervisionato

- Machine Learning si crea con Apprendimento Supervisionato, come un bambino che:
  - quando nasce non sa niente
  - apprende i dati (learning) tramite un insegnante che gli fornisce migliaia di esempi di input e output
  - deve fare un esame di fine anno creato dall'insegnante, se il voto è alto va a lavorare altrimenti deve ripetere l'anno
- insegnante crea esempi per coprire tutti i casi



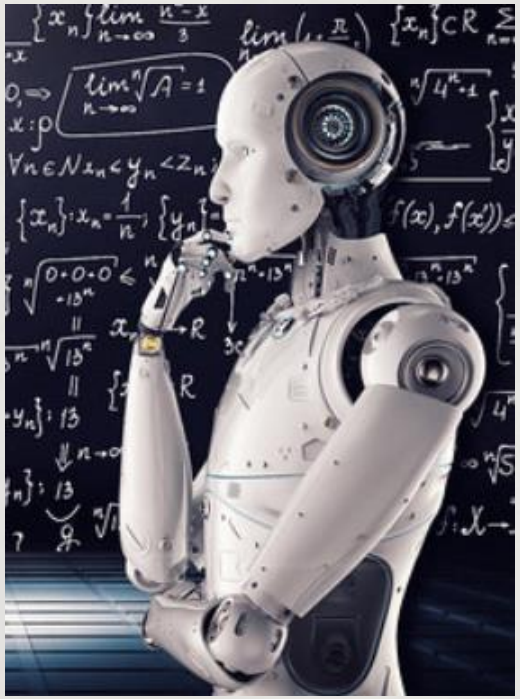
5.Apprendimento supervisionato

Riconoscere oggetti da esempi: IA apprende con disegni e classe uomo/donna creati da un esperto umano



uomo

donna



IA con disegno un po' diverso crea probabilità classe

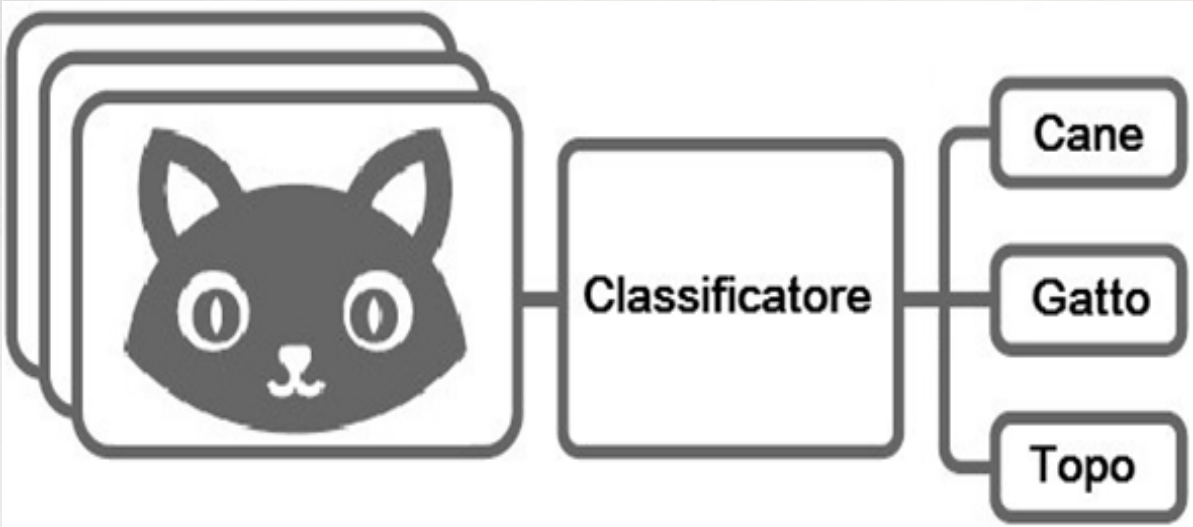
uomo 95%

Calcolare somma di numeri: esempi (input, input, output) (2,1,3) (1,1,2) (0,1,1) (2,5,7) ecc. fino a quando impara a rispondere su numeri mai visti come (2,9) per dare 11

5.Apprendimento supervisionato

- https://teachablemachine.withgoogle.com/
- sperimentare con l'apprendimento supervisionato

5.Apprendimento supervisionato



Classificazione di input nelle classi

Esempi:

- decidere se una transazione è frode o legittima
- scartare o accettare un prodotto da controllo qualità
- scegliere un prodotto da consigliare per acquisto
- riconoscere un prodotto dalla sua immagine

5.Apprendimento supervisionato

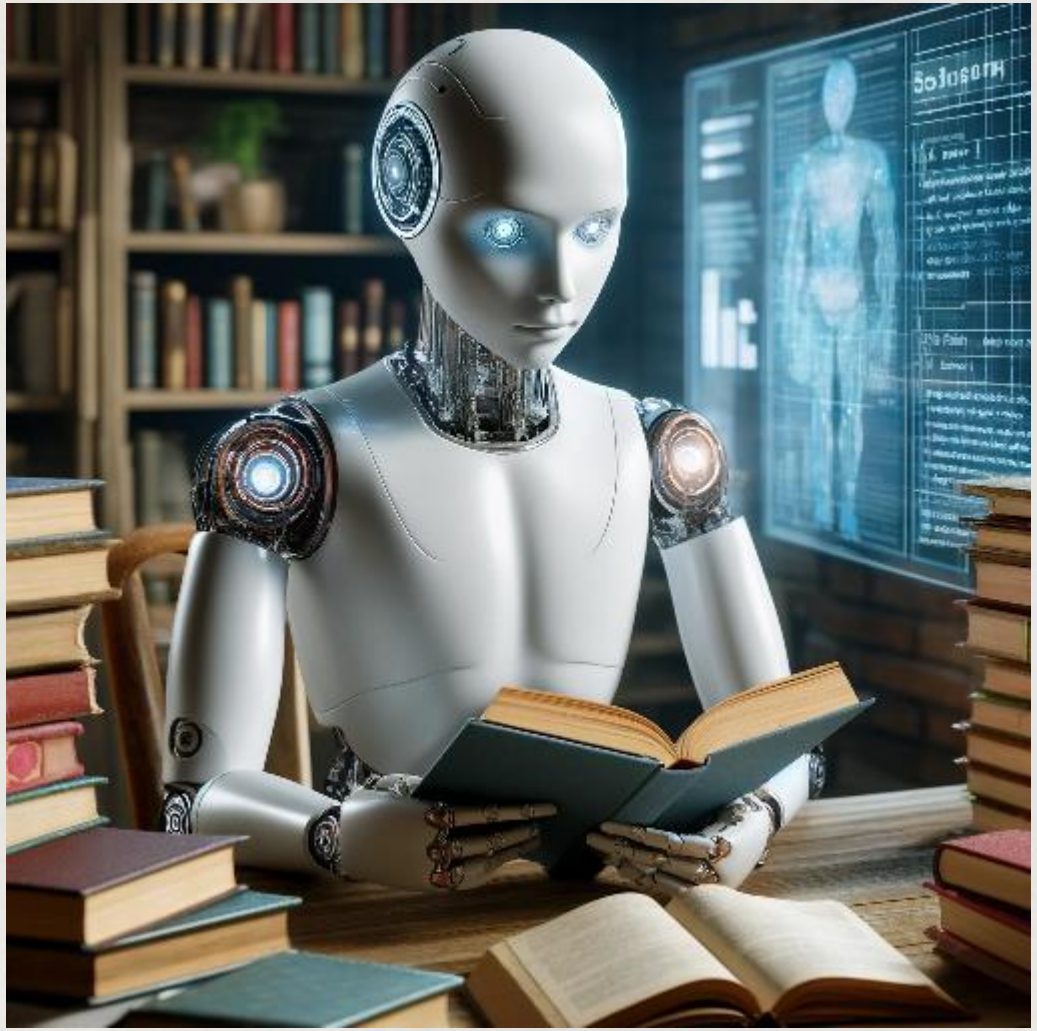
Predizione di un valore ricavato da una serie di dati

Esempi:

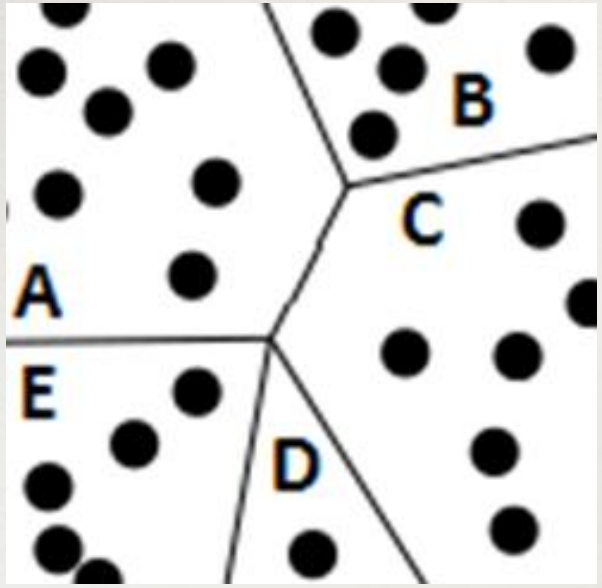
- andamento delle vendite domani secondo il mese scorso
- ricostruire dati mancanti
- predire se un macchinario si guasterà tra poco tempo
- decidere se aumentare o diminuire la prossima azione

5.Apprendimento non supervisionato

- Machine Learning può studiare da sola come un auto didatta
- trova da solo le relazioni tra i dati
- bisogna fidarsi dei suoi risultati



5.Apprendimento non supervisionato



Clustering, creare gruppi di dati omogenei

Esempi:

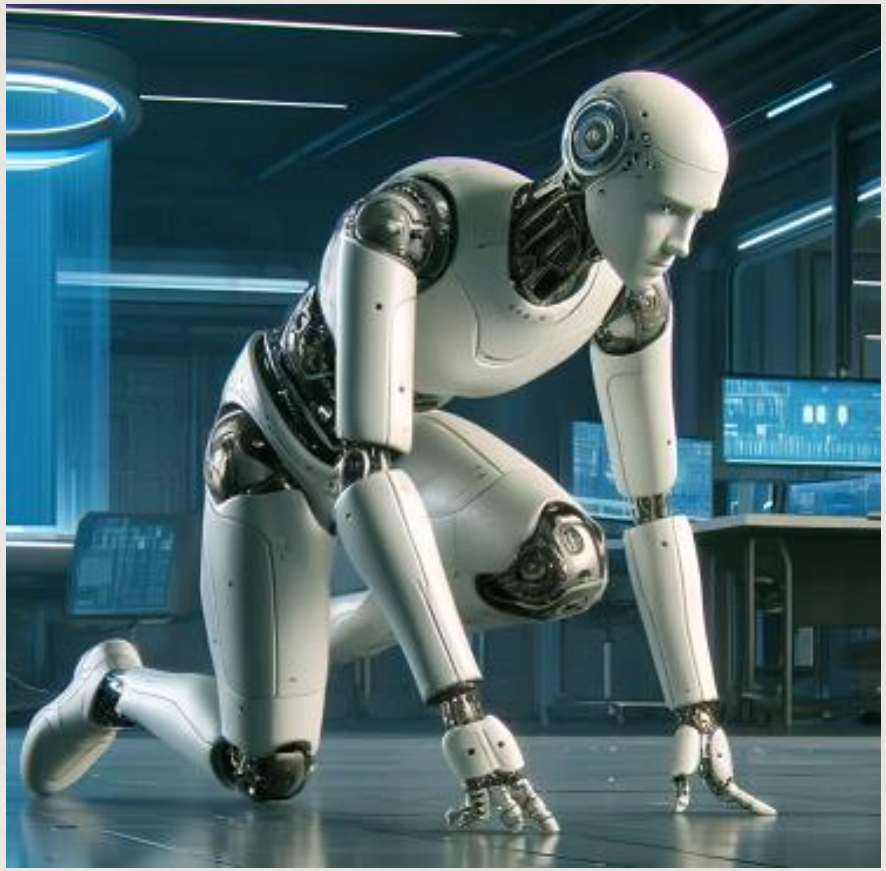
- conoscere le tipologie di acquirenti
- trovare i prodotti più comprati insieme
- distinguere un dato anomalo dagli altri
- trovare nuove relazioni tra i dati

5.Apprendimento per trasferimento

- si crea machine learning per risolvere un problema
- la soluzione viene usata anche per risolvere un problema diverso ma simile con pochi cambiamenti
- esempio: sistema riconosce immagini di pizza può riconoscere anche la pasta

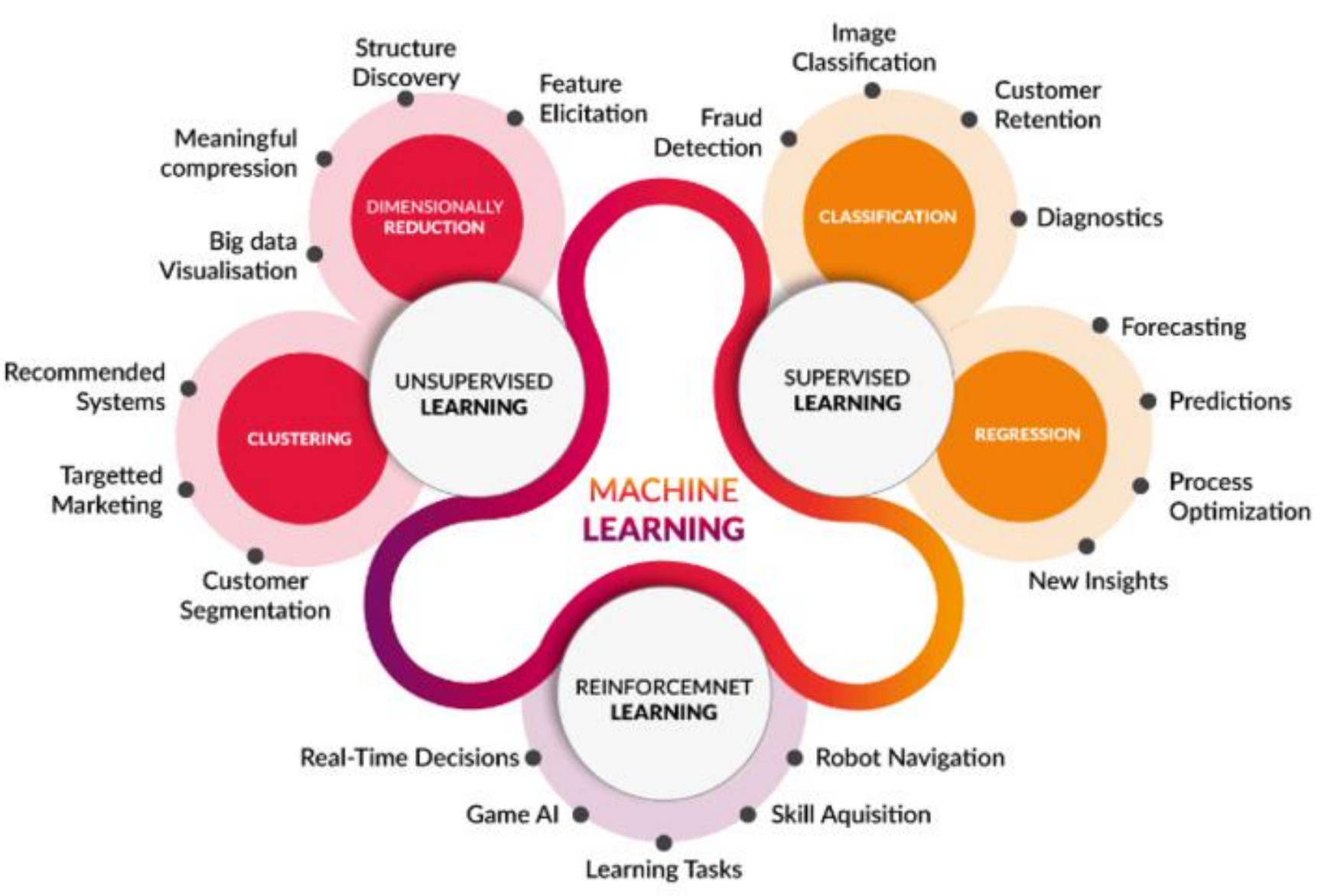
5.Apprendimento per rinforzo

- se il Machine Learning sbaglia viene punito e non ripete l'errore
- se il Machine Learning fa bene viene premiato e va avanti
- esempi:
  - robot che cammina
  - scelte nei videogiochi

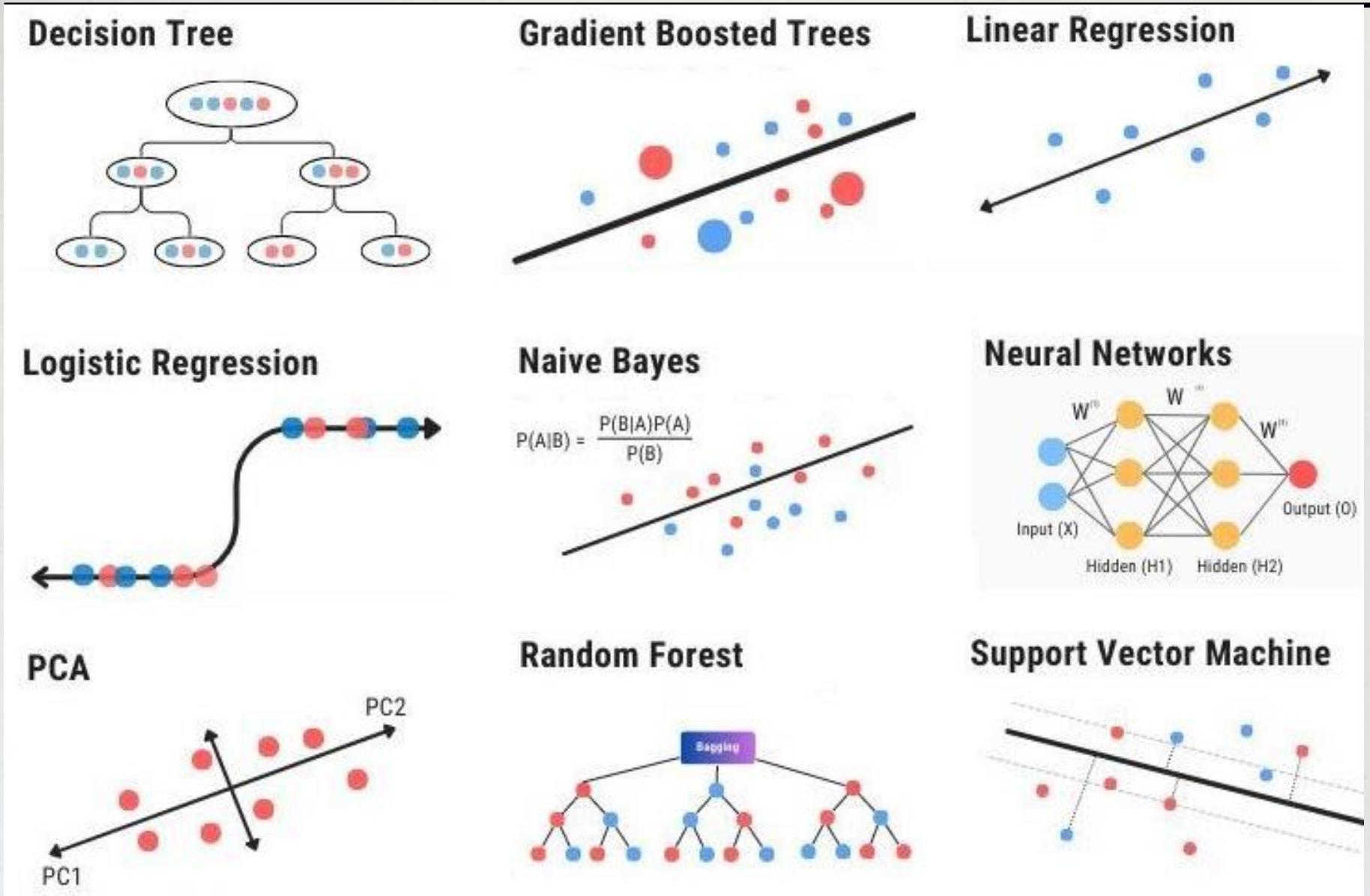


5.Riepilogo Machine Learning

https://www.datasciencecentral.com/profiles/blogs/machine-learning-in-one-picture



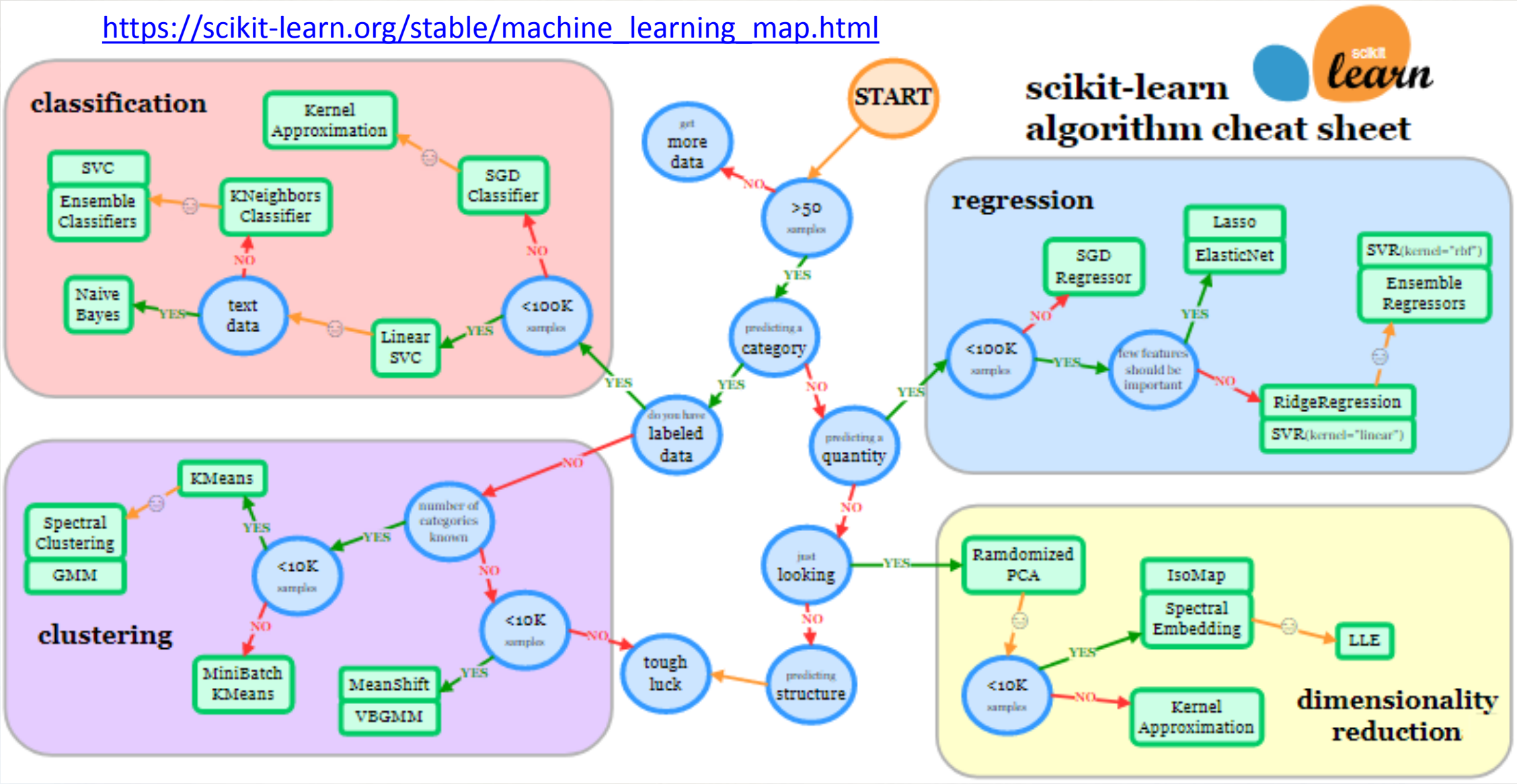
5.Riepilogo Machine Learning



ampia varietà di modelli matematici

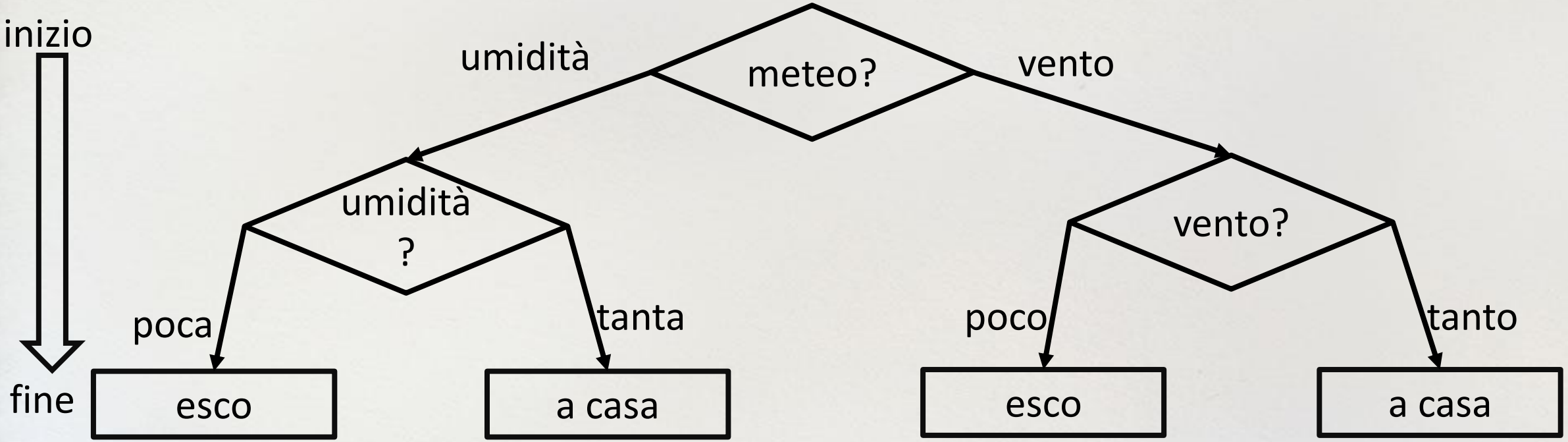
5. Python scelta libreria

https://scikit-learn.org/stable/machine_learning_map.html



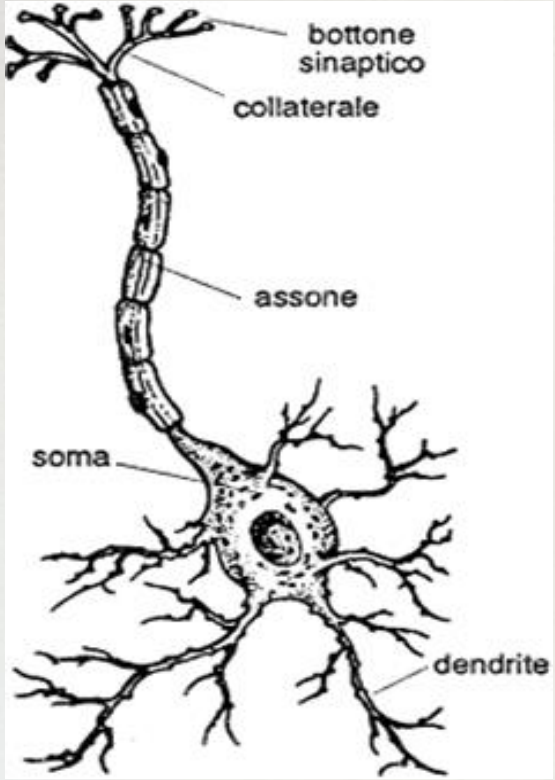
6. Albero Decisione

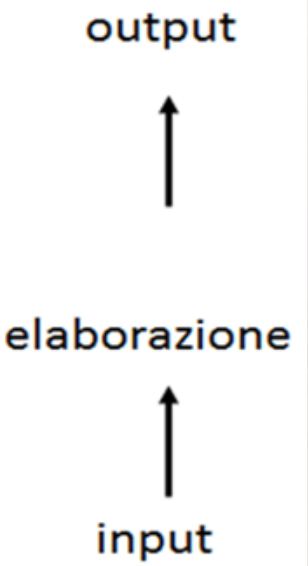
- creare tanti percorsi di domande e risposte fino a trovare quello per arrivare in poco tempo alla soluzione
- si può capire il percorso di elaborazione svolto
- fornisce una sola risposta senza margine di errore

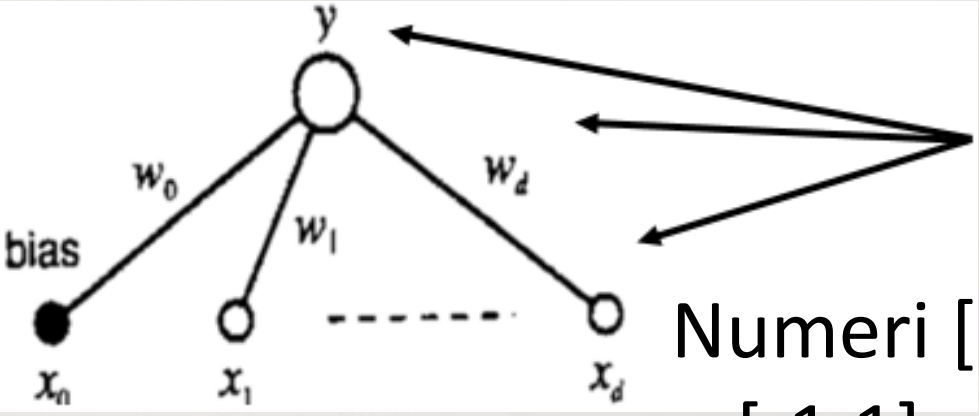


7.Neural Networks – Reti Neurali

Analogia con neurone nel cervello umano







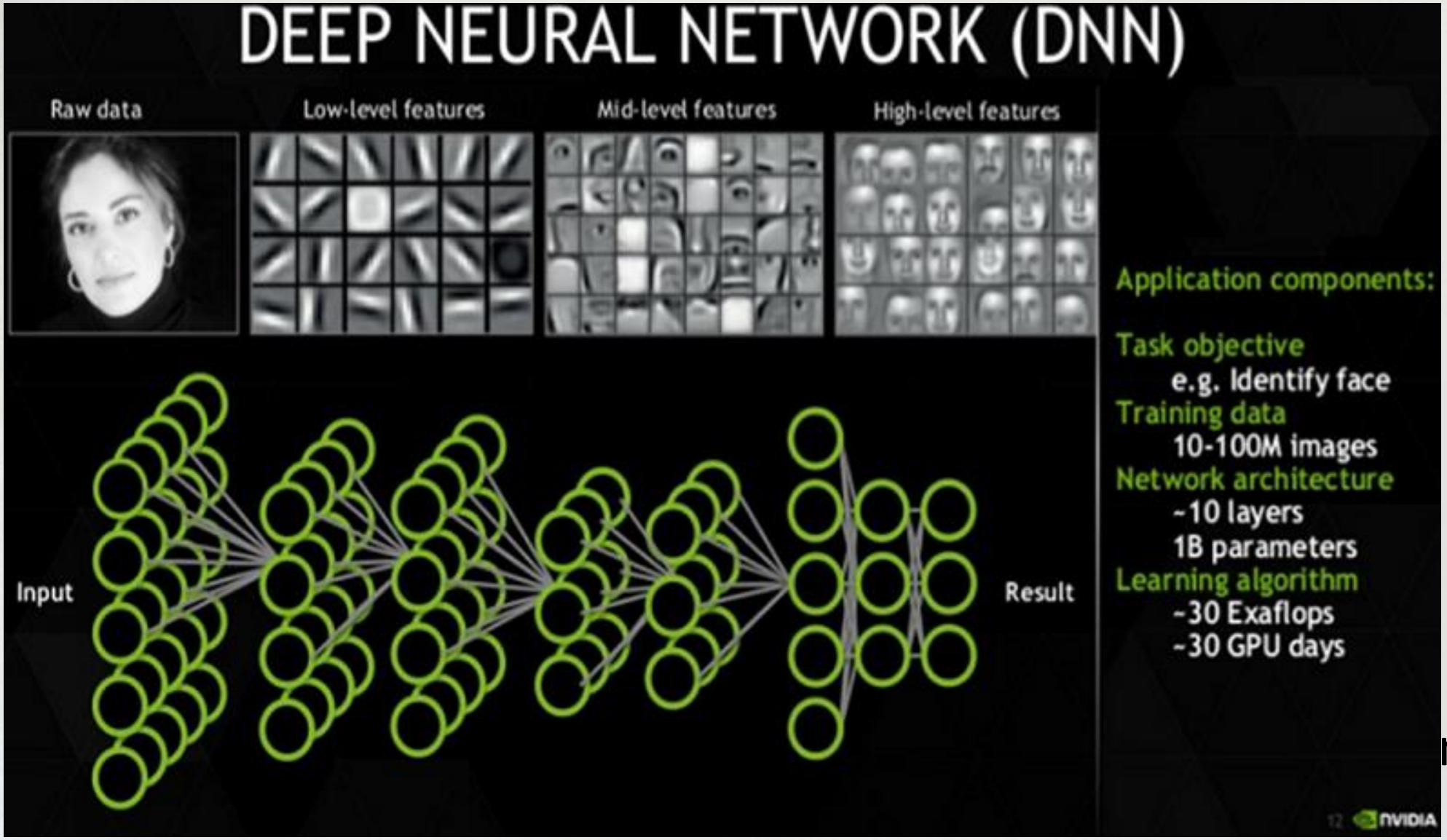
Numeri [0,1] o [-1,1]

$$y(x) = g\left(\sum_{i=1}^{d} w_i x_i + w_0 x_0\right)$$

Cervello umano con 85 miliardi di neuroni biologici

Neurone calcolato con la matematica per fare IA

7.Neural Networks, Deep Learning

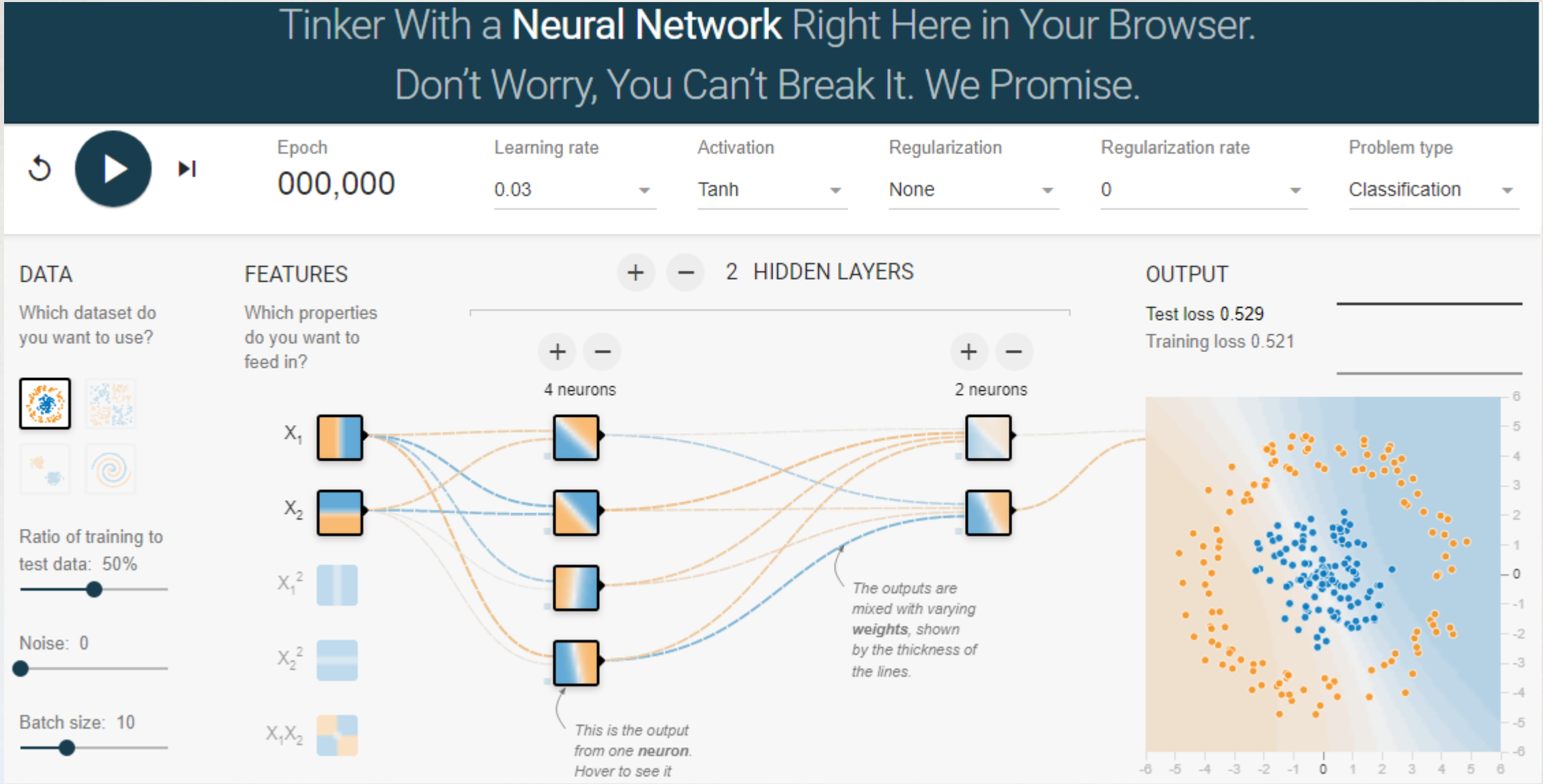


Deep Learning rete neurale formato da tantissimi neuroni matematici

esempio riconoscimento del volto umano

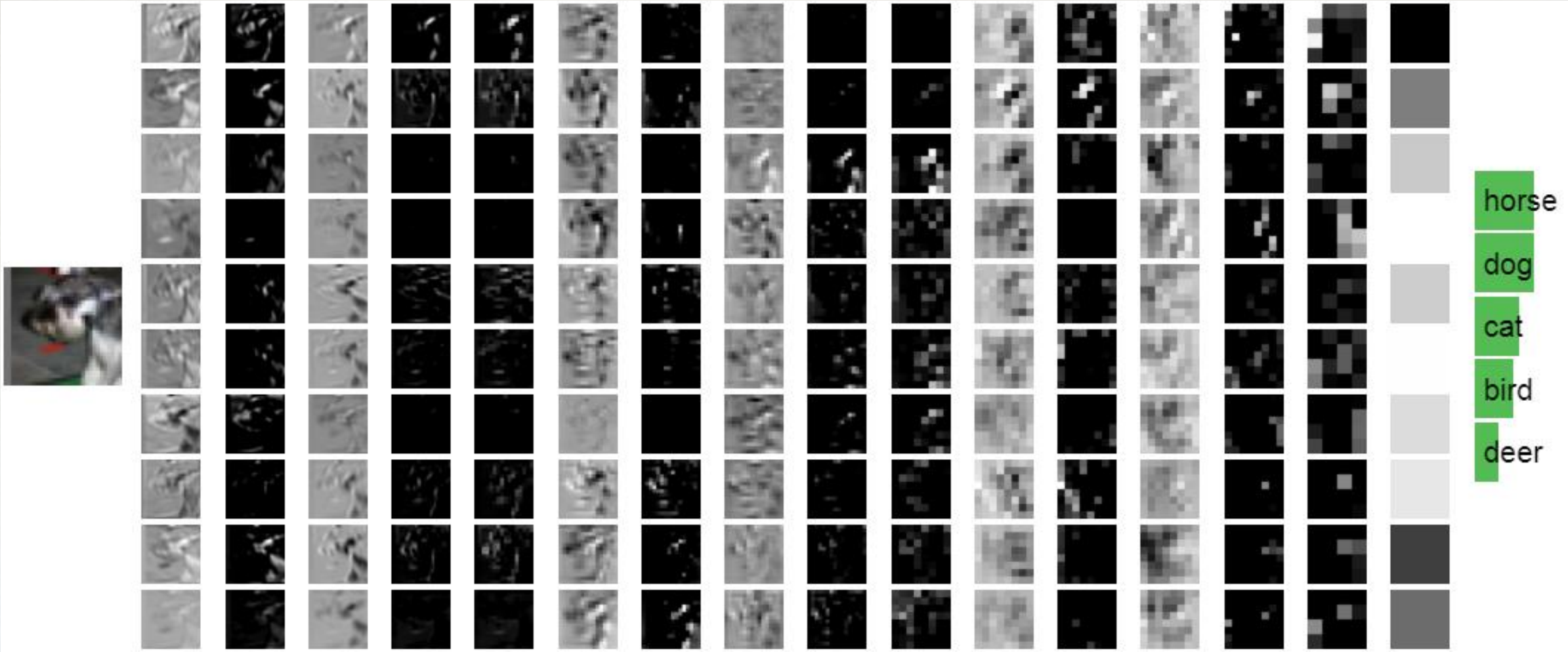
7. Prova Neural Newtork con Google

https://playground.tensorflow.org/



7. Demo visuale Deep Learning

http://cs231n.stanford.edu/

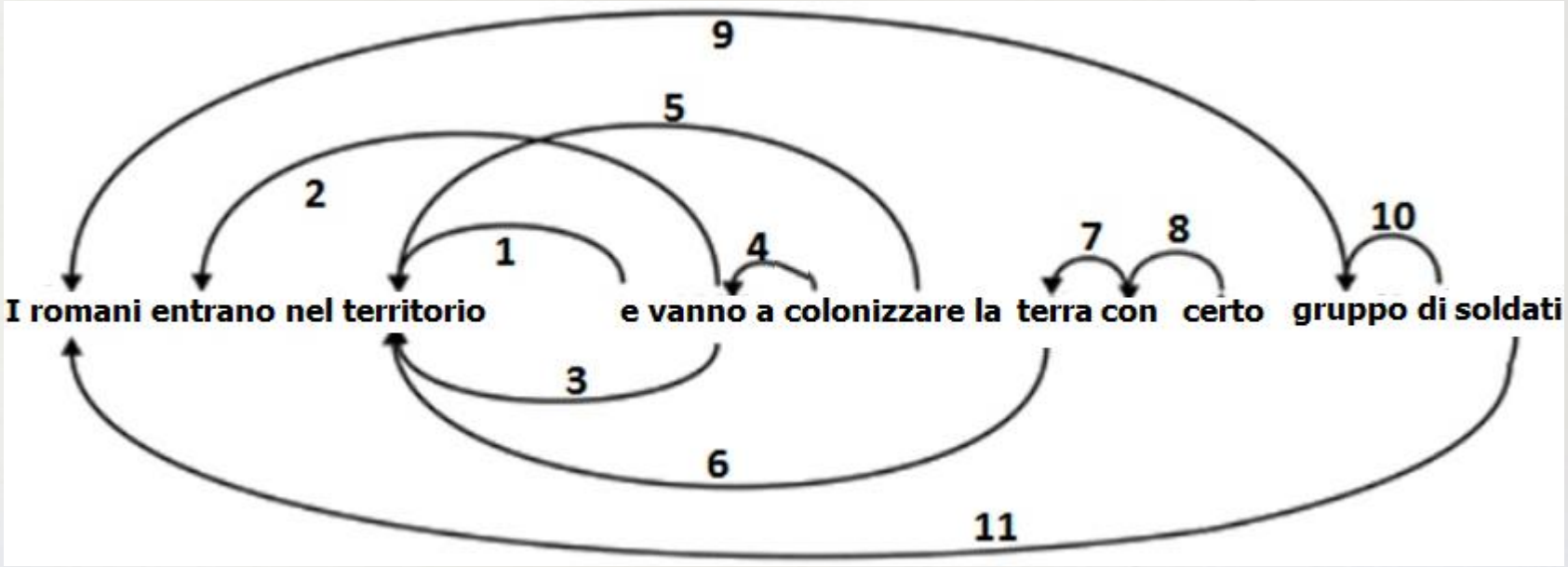


*This network is running live in your browser

7.Intelligenza Artificiale Generativa

ChatGPT ha letto gratis miliardi di parole, per esempio ha visto milioni di volte «il gatto litiga con il cane» e non ha mai visto «la gatto salta con treno» scrive «il gatto» e propone «litiga […]»

Esempio: creare parole in risposta al prompt a sinistra, ordine progressivo delle parole create su cui pone attenzione per dare la prossima parola



Controllare, non fidarsi, ragionarci sopra!

7.Neural networks linguaggio Python

```
#importazione librerie necessarie
import numpy as np
from sklearn.model_selection import train_test_split #organizza dati learning
from sklearn.linear_model import Perceptron #modello di predizione scelto
from sklearn.metrics import accuracy_score #calcolo dell'accuratezza

#carica dati di fiori IRIS da dividere su 3 classi
from numpy import genfromtxt
dati =np.genfromtxt('fioriIRIS.csv', delimiter=',')
X = dati[:,0:4] #input dati per descrivere il fiore
y = dati[:,4] #output con classe attribuita al fiore

#divide in insiemi di training 80% e test 20% senza scelta casuale
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)

#impostazione parametri del modello neural network Perceptron
ppn = Perceptron(max_iter=40, tol=0.001, eta0=0.01, random_state=0)
#supervised machine learning del modello scelto per apprendere dai dati di training set
ppn.fit(X_train, y_train)

y_pred = ppn.predict(X_test) #predizione sul test set per creare esame finale di controllo
accuracy_score(y_test, y_pred) #verifica errore sul test set

print("classe:",ppn.predict([[0.3, 4.8, 0.1, 2.1]]))#calcolo classe su nuovi dati non usati
```

rete neurale

formula matematica da calcolare per cambiare pesi

Classificazione fiori Iris con Python

download codice e dati https://bit.ly/3UHgZu2

spiegazioni https://tinyurl.com/nz5bvs3z https://tinyurl.com/2bt6m3c7

7.Neural networks linguaggio Python

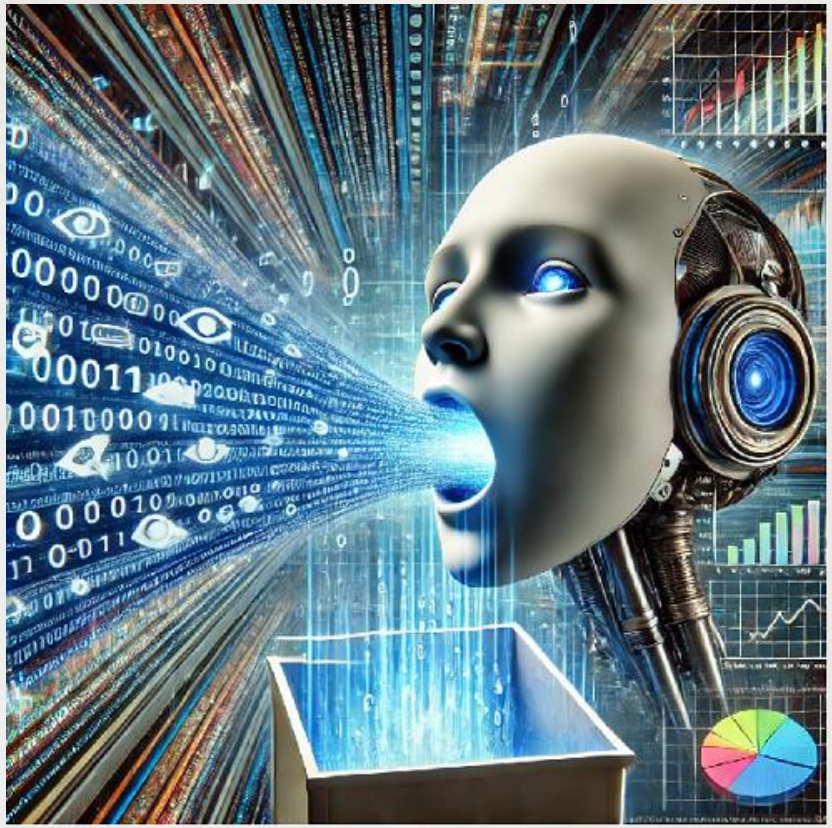
```python
import googlemaps
import itertools
import numpy as np
def tsp_google_maps(locations: list, api_key: str):
    gmaps = googlemaps.Client(key=api_key)
    distance_matrix = gmaps.distance_matrix(
        origins=locations, destinations=locations, mode="driving", units="metric")
    num_locations = len(locations)
    distances = np.zeros((num_locations, num_locations))
    for i, row in enumerate(distance_matrix['rows']):
        for j, element in enumerate(row['elements']):
            if element['status'] == 'OK':
                distances[i][j] = element['distance']['value']  # Distanza in metri
            else:
                distances[i][j] = np.inf  # Se c'è un errore, assegniamo infinito
if __name__ == "__main__":
    locations = [
        "Milano, Italia",          "Roma, Italia",          "Torino, Italia"
    ]
    route, distance = tsp_google_maps(locations, API_KEY)
    print(f"Il percorso ottimale è: {route}")
    print(f"La distanza totale è: {distance / 1000:.2f} km")  # Conversione in chilometri
```

richiamo funzioni su server Google

formule matematiche calcolate su server Google

8. Trovare dati per apprendere

- Servono tanti dati di qualità GIGO Garbage In Garbage Out: nell'informatica se entra spazzatura esce spazzatura



8.I migliori dataset

- https://365datascience.com/trending/public-datasets-machine-learning/

1. Boston House Price Dataset
2. Iris Dataset
3. MNIST dataset
4. Dog Breed Identification
5. ImageNet
6. Breast Cancer Wisconsin Diagnostic Dataset
7. Amazon Reviews Dataset
8. BBC News
9. YouTube Dataset
10. Catching Illegal Fishing

8.OpenML

- https://www.openml.org/search?type=data&sort=runs&status=active oltre 25.000 dataset 5800 verificati
- https://www.openml.org/search?type=data&status=active&id=43828

8.Kaggle Wine Dataset

- https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009
- https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009/code

Caratteristiche per descrivere un vino rosso:

1. fixed acidity
2. volatile acidity
3. citric acid
4. residual sugar
5. chlorides
6. free sulfur dioxide
7. total sulfur dioxide
8. density
9. ph
10. sulphates
11. alcohol
12. quality, classificare buono/cattivo

8.UCI Machine Learning Repository

- https://archive.ics.uci.edu/datasets 600 dataset
- https://archive.ics.uci.edu/dataset/53/iris

Iris

Donated on 6/30/1988

A small classic dataset from Fisher, 1936. One of the earliest known datasets used for evaluating classification methods.

Dataset Characteristics	Subject Area	Associated Tasks
Tabular	Biology	Classification
Feature Type	**# Instances**	**# Features**
Real	150	4

Dataset Information

What do the instances in this dataset represent?

Each instance is a plant

Additional Information

This is one of the earliest datasets used in the literature on classification methods and widely used in statistics and machine learning. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are not linearly separable from each other....

SHOW MORE

DOWNLOAD (3.7 KB)

IMPORT IN PYTHON

CITE

352 citations

670873 views

Keywords

ecology

Creators

R. A. Fisher

DOI

10.24432/C56C76

8.Dataset immagini

- https://research.google/blog/introducing-the-open-images-dataset/ 9 milioni di URL di immagini annotate con etichette che abbracciano oltre 6000 categorie

- http://vision.stanford.edu/aditya86/ImageNetDogs/ 20.580 immagini per 120 classi

8.Pizza dataset

- https://universe.roboflow.com/search?q=class%3Apizza
- Guide: How to Train a Computer Vision Model to Detect Pizzas
- https://www.kaggle.com/datasets/shilongzhuang/pizza-sales dati di vendita
- https://www.kaggle.com/datasets/carlosrunner/pizza-not-pizza 1000 immagini di pizza
- https://www.tensorflow.org/datasets/catalog/food101?hl=it 101 categorie di alimenti, con 101'000 immagini

8. Dataset Chihuahua e muffin

- Input con oggetti simili in apparenza
- Classificatore immagini con Machine Learning riconosce quanti Chihuahua e muffin?
- https://www.freecodecamp.org/news/chihuahua-or-muffin-my-search-for-the-best-computer-vision-api-cbda4d6b425d/
- https://www.semanticscholar.org/paper/Deep-Learning-Approach-for-Very-Similar-Objects-on-Togootogtokh-Amartuvshin/511a3a0a690d0e992a2502f9be4ffb1b1d4236e2
- https://www.robots.ox.ac.uk/~vgg/data/pets/ dataset Chihuahua

9. Gestione del progetto

1. Perché usare IA? E' di moda, lo usa il concorrente, risolve un problema, porta un vantaggio, lo chiede il mercato, curiosità, fa vendere
2. Voglio sostituire una persona o aiutarla a lavorare?
3. Ho i dati necessari? Tanta quantità, con qualità, facile accedere, di proprietà
4. Quale margine di errore ammetto su input e output?
5. Quanto tempo, personale, denaro, strumenti posso investire nella creazione?
6. Dopo averla creata, ho le risorse per manutenzione e aggiornamento?

9. Gestione del progetto

1. Come devo integrare nel mio processo aziendale?
2. Come ho scelto chi la crea? Ha le capacità adeguate?
3. Chi ha creato il sistema quale documentazione fornisce?
4. Chi ha creato il sistema quali responsabilità si assume?
5. In caso di errori di funzionamento ne rispondo io?
6. Il sistema di IA può spiegare coma ha deciso su certi dati? Problema black-box di funzionamento ignoto?

9. Gestione del progetto

Procedura per cercare soluzione che usa IA:

1. analisi del problema
2. specifica della soluzione da ottenere
3. scelta ambiente software di sviluppo
4. scelta di input, output e altri parametri da regolare
5. scelta approccio e modello per fare ia
6. scelta dei parametri da ottimizzare nel modello
7. test su ampie casistiche
8. verifica dei risultati con test e metriche
9. decidere cosa cambiare e ripetere

9. Gestione del progetto

- linguaggio di programmazione e librerie coinvolte
- quali servizi cloud usati, riservatezza dei dati
- quali tipologie di input sono ammessi e quale output fornito
- margine di errore ammesso in input e output
- come è stato addestrato il sistema
- caratteristiche di input usati nell'addestramento e nel test
- quanto tempo impiegato per la creazione

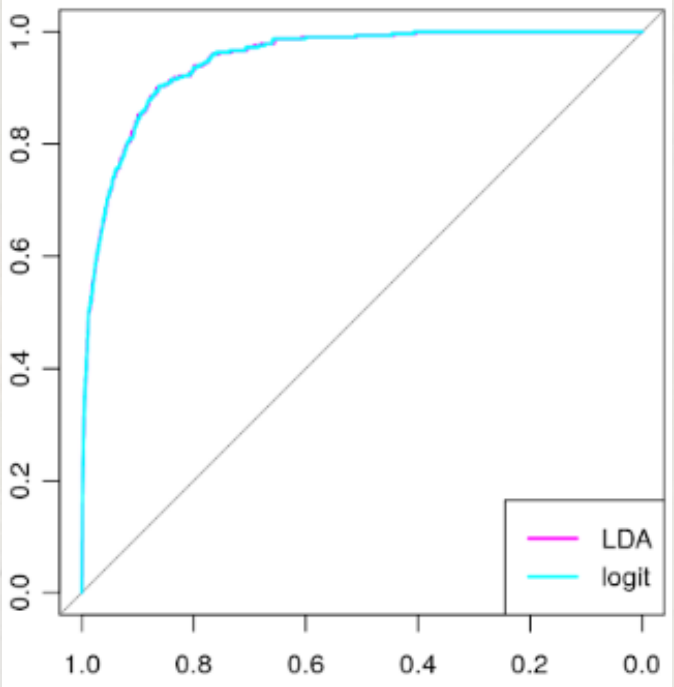
9. Gestione del progetto

Riepilogo dei costi:

- costo ricerca stato dell'arte
- costo studio di fattibilità
- costo di sviluppo
- costo di mantenimento in esercizio
- costo di eventuale comunicazione
- costo di eventuale formazione di utente e personale
- costo di licenze software e servizi cloud
- costo per hardware
- costo da affrontare in caso di errori e fallimenti

9. Gestione del progetto

- il collaudo è fondamentale
- verificare rispetto AI ACT e GDPR
- verificare copyright dei dati
- test set deve coprire tutte le casistiche di input e output
- verifica esistenza di bias e pregiudizi
- usare dati non ancora considerati
- creare matrice confusione, curva ROC e altri metodi statistici per valutazione
- decidere se l'errore commesso è ammissibile
- documentare con cura

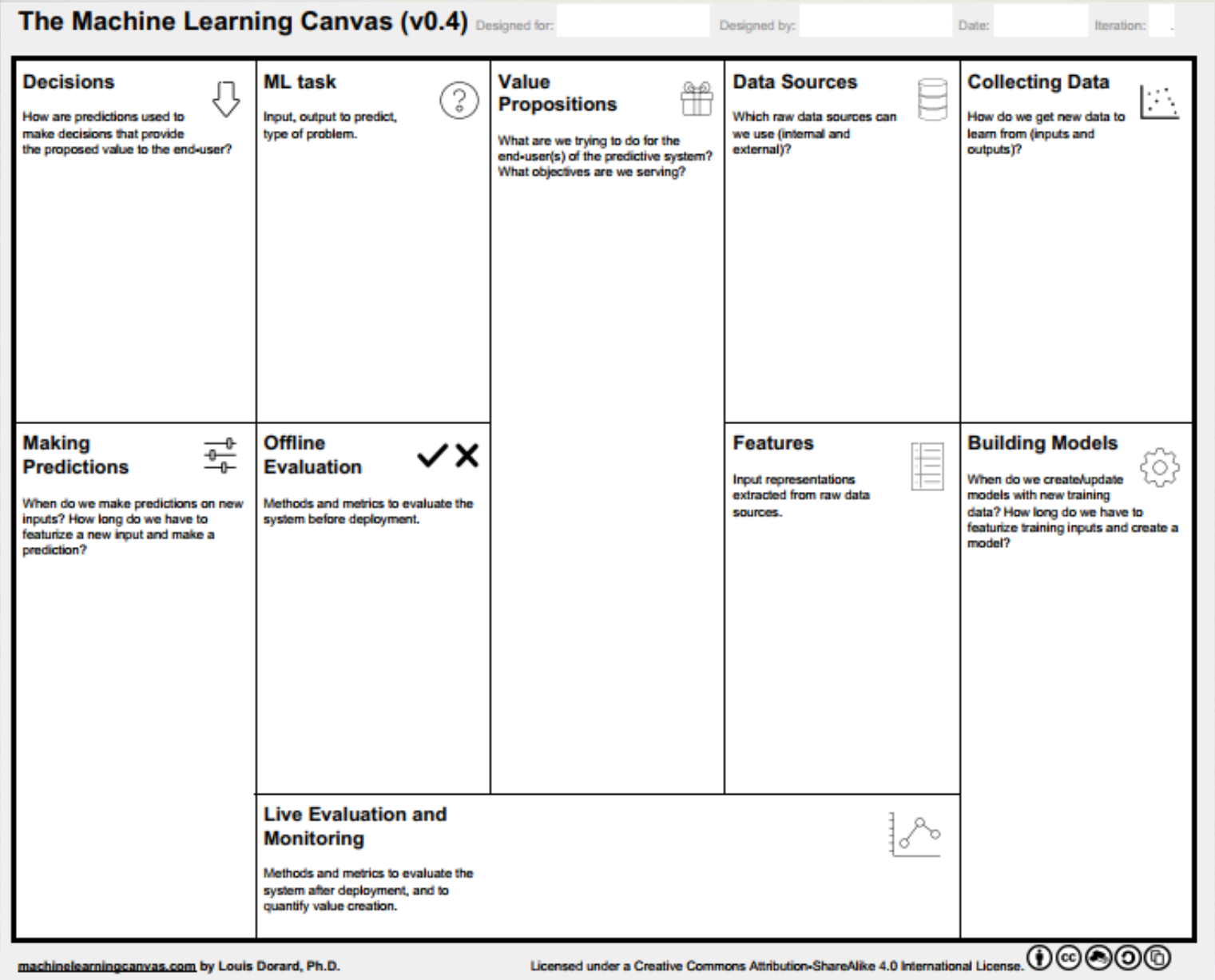


	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

9. Gestione del progetto

- IA ha bisogno di manutenzione continua
- I dati input possono cambiare nel tempo, come mi accorgo delle modifiche?
- Ogni quanto fermare IA per aggiornamento?
- Se IA è fermo chi opera al suo posto?
- Come selezionare i dati per aggiornamento?
- Come valutare le nuove prestazioni?
- Conservare vecchio modello come riserva?

9. Gestione del progetto



- https://www.ownml.co/machine-learning-canvas
- per riepilogare le idee

9. Gestione del progetto

- ISO/IEC 42001 primo standard internazionale per sistemi con IA
- certificazione da un ente terzo indipendente
- requisiti per avere un sistema di gestione dell'IA in modo responsabile
- elenco di controlli sullo sviluppo, fornitura, uso e monitoraggio del sistema IA
- vantaggi in gestione dei rischi e fiducia degli stakeholder

ATTESTATO DI IDONEITÀ

N. A4

rilasciato a

ROBERTO MARMO

C.F.

per aver superato il Corso

AUDITOR / LEAD AUDITOR

INFORMATION TECHNOLOGY - ARTIFICIAL INTELLIGENCE - MANAGEMENT SYSTEM

Secondo la norma

ISO/IEC 42001:2023

Il Corso, di 24 ore, iscritto al N° del Registro dei Corsi Riconosciuti AICQ SICEV, è stato concluso il giorno 08 Maggio

9. Gestione del progetto

IA può fornire un risultato sbagliato per colpa di:

- malfunzionamento di hardware
- dati pochi, cattivi, non rappresentano tutti i casi
- errore del programmatore nel software
- errore del sistema operativo
- cattiva analisi del problema
- modello matematico non adeguato
- scelta sbagliata dell'apprendimento
- mancanza di tempo e denaro per creare bene

10. Prompt per ChatGPT

- Agisci come esperto in machine learning. Aiutami a scegliere un modello per il mio problema di classificazione immagini di pizze. Fammi una domanda, attendi la risposta, poi fai un'altra domanda.
- Agisci come esperto in machine learning. Aiutami a scegliere un modello per il mio problema di classificazione immagini di pizze. 10 variabili input, 3 variabili output. Usare un modello spiegabile.
- Sono un principiante che vuole studiare il machine learning, consigliami un modello semplice per capire il funzionamento del machine learning.
- Scrivi il codice Python commentato e scegli un dataset per fare le prove.
- Crea la tabella per spiegare funzionamento di algoritmi di machine learning, vantaggi, svantaggi, spiegabile o non spiegabile, quantità minima di input richiesti, indicazione sul tempo medio di esecuzione.

10. Prompt per ChatGPT

- Ho un problema di classificazione. Quale algoritmo di machine learning mi consigli di usare per ottenere un buon bilanciamento tra accuratezza e spiegabilità?
- Quali algoritmi di machine learning sono consigliati se ho bisogno di un modello interpretabile per prendere decisioni in un contesto di business?
- Quale algoritmo di machine learning funziona bene con un dataset numericamente grande, ma di bassa complessità, e che sia spiegabile?
- Vorrei un algoritmo di classificazione che funzioni bene con dati limitati e che sia semplice da interpretare. Mi suggerisci un’opzione?
- Agisci come esperto in machine learning. Aiutami a capire se devo risolvere un problema di classificazione o di regressione. Fammi una domanda, attendi la risposta, poi fai un'altra domanda.

Conclusioni

- intelligenza artificiale ci offre tante nuove soluzioni a problemi complicati con cui:
  - ridurre costi
  - aumentare ricavi
  - dare senso di modernità
- non farsi prendere da facili entusiasmi o prime delusioni
- preparare obiettivi, domande, risorse e valutazioni adeguate
- servono consapevolezza e preparazione

Roberto Marmo con editore

www.algoritmiia.it

https://robertomarmo.net/Libri.html#top

Domande e risposte?
Serve una mano per costruire la Intelligenza Artificiale?

www.robertomarmo.net
info@robertomarmo.net