

Apprendimento Automatico Non Supervisionato

Un'introduzione a uno degli approcci più potenti nel moderno machine learning



Cosa Esploreremo Oggi

Comprendere l'Apprendimento Non Supervisionato

01

Cos'è l'Apprendimento Non Supervisionato?

Concetti chiave e come si differenzia dagli approcci supervisionati

02

Algoritmi di Clustering

Raggruppare i punti dati per scoprire schemi nascosti

03

Applicazioni nel Mondo Reale

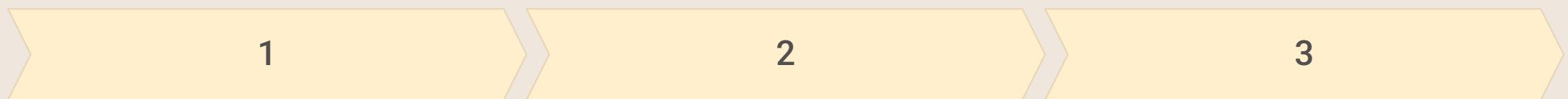
Dall'aggregazione di notizie alla ricerca genetica e alla segmentazione del mercato

04

Altri Metodi Non Supervisionati

Rilevamento delle anomalie e tecniche di riduzione della dimensionalità

Definizione Formale: Apprendimento Non Supervisionato



Apprendimento Supervisionato

I dati includono sia gli input X che le etichette di output Y

Apprendimento Non Supervisionato

I dati includono solo gli input X ma non le etichette di output Y

3

Il Compito dell'Algoritmo

Trovare struttura, modelli o qualcosa di interessante nei dati

Questa distinzione definisce la sfida fondamentale e l'opportunità dell'apprendimento non supervisionato: estrarre significato senza una guida esplicita su quale dovrebbe essere l'interpretazione "corretta".

Supervisionato vs Non Supervisionato: La Differenza Chiave

Apprendimento Supervisionato

I dati includono sia gli input (X) che le etichette di output (Y).

L'algoritmo impara a mappare gli input agli output corretti.

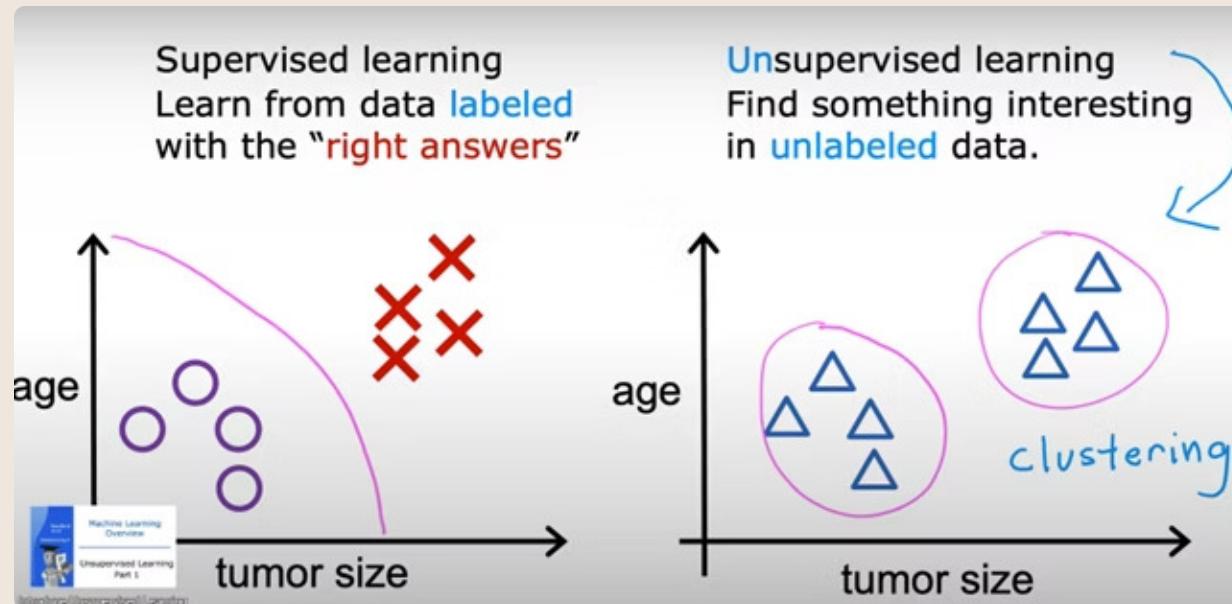
Esempio: Classificare i tumori come benigni (0) o maligni (X) in base alle dimensioni e ai dati del paziente.

Apprendimento Non Supervisionato

I dati includono solo gli input (X) ma nessuna etichetta di output.

L'algoritmo deve scoprire la struttura da solo.

Esempio: Date le dimensioni del tumore e l'età del paziente senza diagnosi, trovare schemi interessanti nei dati.



La Libertà dell'Apprendimento Non Supervisionato

"Non stiamo cercando di supervisionare l'algoritmo per dare una 'risposta giusta' per ogni input. Invece, chiediamo all'algoritmo di scoprire da solo cosa sia interessante, quali schemi o strutture possano esserci in questi dati."

Questo cambiamento fondamentale nell'approccio conferisce all'apprendimento non supervisionato una notevole flessibilità. Senza etichette predefinite che vincolano l'analisi, gli algoritmi possono scoprire relazioni inaspettate, raggruppamenti nuovi e strutture nascoste che gli osservatori umani potrebbero non anticipare mai.

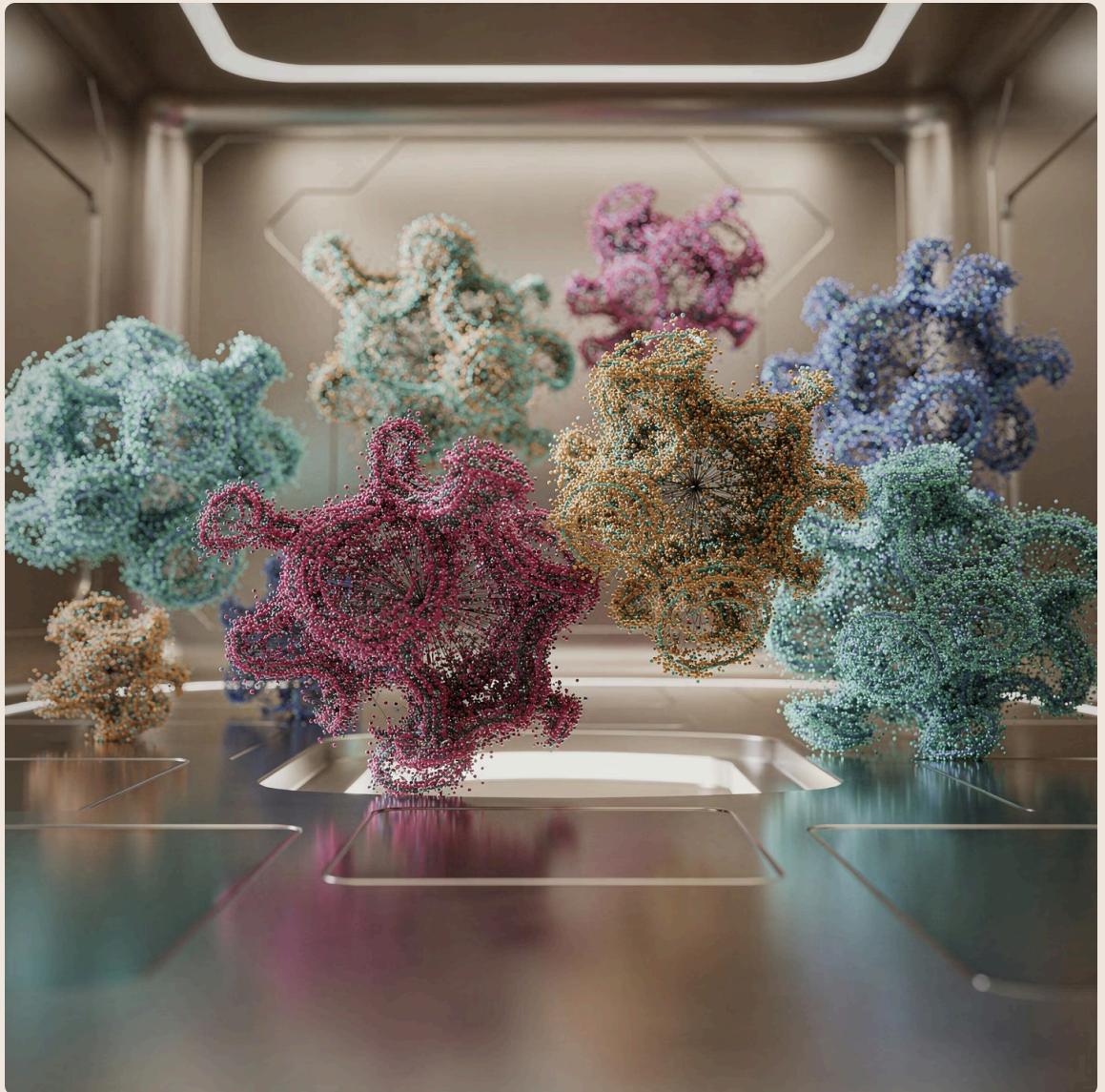
L'algoritmo diventa un esploratore piuttosto che uno studente, trovando il proprio percorso attraverso il panorama dei dati.

Introduzione al Clustering

Scoperta Automatica di Modelli

Quando un algoritmo di apprendimento non supervisionato esamina dati non etichettati, potrebbe scoprire che i dati si raggruppano naturalmente in gruppi o cluster distinti. Questo particolare tipo di apprendimento non supervisionato è chiamato **un algoritmo di clustering**.

Ad esempio, con dati di pazienti che mostrano la dimensione del tumore e l'età, l'algoritmo potrebbe identificare due cluster distinti, senza che gli venga detto cosa rappresentano. Riconosce semplicemente che certi punti dati appartengono insieme in base alle loro caratteristiche.



Applicazione nel Mondo Reale

Google News: Il Clustering in Azione

Ogni giorno, Google News analizza centinaia di migliaia di articoli di notizie su Internet. Un algoritmo di clustering raggruppa automaticamente le storie correlate, creando raccolte coerenti di articoli sullo stesso argomento.

Consideriamo un articolo sul prezzo dei biglietti per i giochi olimpici invernali. L'algoritmo rileva che molti altri articoli contengono le parole "Olimpici" e "biglietto", suggerendo che trattano la stessa notizia. Senza intervento umano, raggruppa questi articoli in un unico cluster.

The screenshot shows a Google News feed with several news items related to the 2026 Winter Olympics in Milan-Cortina d'Ampezzo. The items are clustered together, indicating they are part of the same topic. Red boxes highlight specific text segments in the headlines of the first two articles, likely demonstrating how Google's clustering algorithm identifies related content based on keywords like "Olimpiadi", "biglietti", and "prezzi".

- 10 ore fa • Di Beatrice Curti
Cerimonia di apertura Olimpiadi Milano – Cortina 2026: date, biglietti e ospiti
- 17 giorni fa • Di Massimiliano Melley
Caro biglietto alle Olimpiadi, secondo Malagò la metà ha prezzi economici
- 30 lug
Il Giorno
Olimpiadi, la speranza di esserci. Oggi riapre la biglietteria online
- 4 giorni fa
Il Post
Quanto costa vedere dal vivo le Olimpiadi di Milano Cortina?

L'Intelligenza Dietro il Raggruppamento delle Notizie

Nessuna Supervisione Umana Richiesta

Non c'è un dipendente di Google News che istruisce manualmente l'algoritmo su quali parole indicano articoli correlati. Il sistema lo determina autonomamente.

Si Adatta ai Cambiamenti Quotidiani

Gli argomenti delle notizie cambiano ogni singolo giorno. Con innumerevoli storie pubblicate a livello globale, la curatela umana non è semplicemente fattibile su questa scala.

Scopre Modelli Indipendentemente

L'algoritmo scopre da solo quali parole e frasi suggeriscono che gli articoli appartengono allo stesso gruppo, rendendolo un vero sistema di apprendimento non supervisionato.

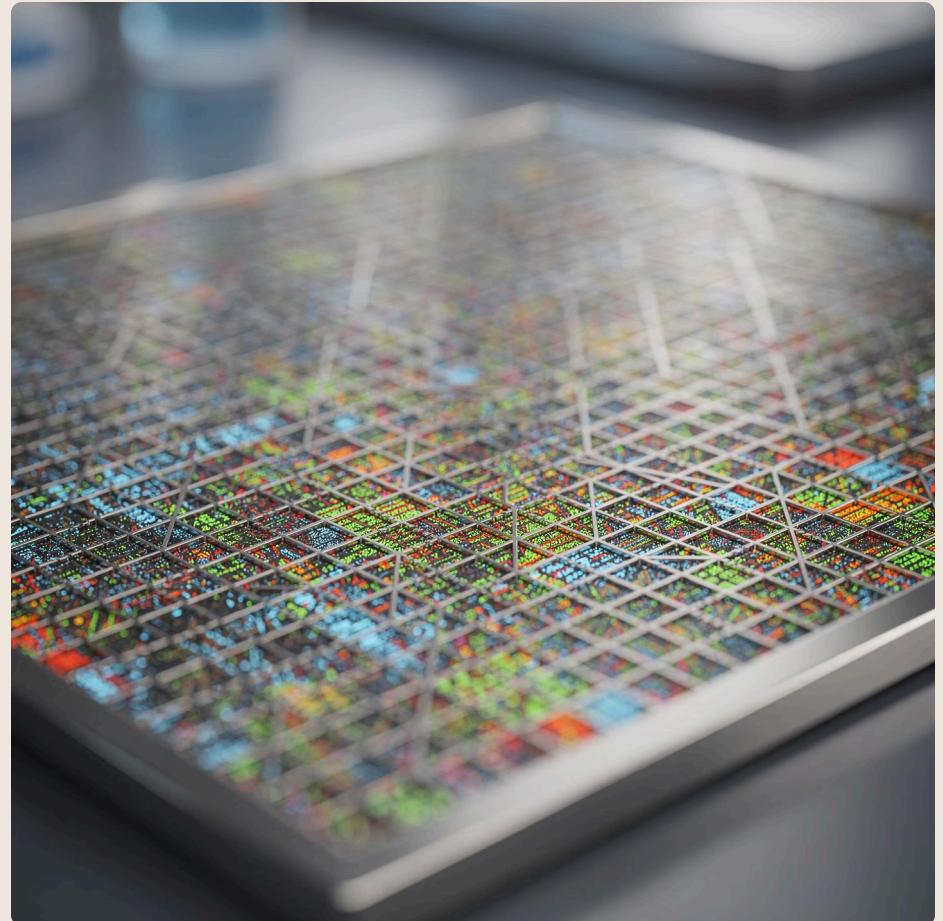
Applicazione di Clustering #2

Analisi dei Dati Genetici

L'apprendimento non supervisionato si rivela inestimabile nella ricerca genetica. I dati dei microarray di DNA appaiono come piccole griglie simili a fogli di calcolo, dove ogni colonna rappresenta l'attività genetica di una persona e ogni riga rappresenta un gene specifico.

Ad esempio, una riga potrebbe rappresentare un gene che influenza il colore degli occhi, un'altra l'altezza, e i ricercatori hanno persino identificato collegamenti genetici alle preferenze vegetali (come l'avversione per broccoli, cavolini di Bruxelles o asparagi).

I colori—rosso, verde, grigio—indicano il grado in cui i diversi geni sono espressi in ogni individuo.

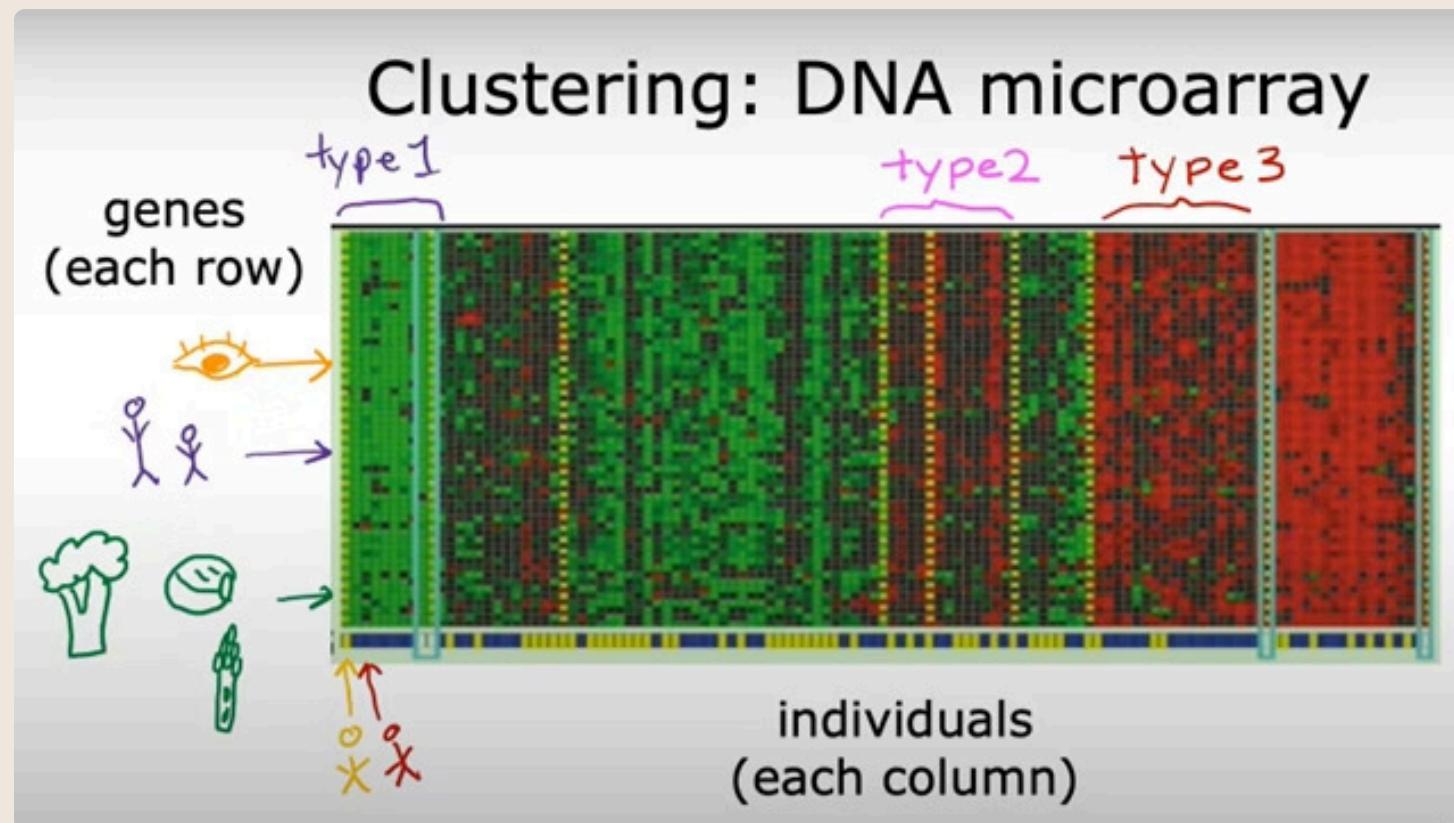


Scoprire i Tipi Genetici

Un algoritmo di clustering può analizzare i dati dei microarray di DNA e raggruppare automaticamente gli individui in diverse categorie o tipi basati sui loro profili genetici. L'algoritmo potrebbe identificare individui di Tipo 1, Tipo 2 e Tipo 3, ognuno con caratteristiche genetiche distinte.

Fondamentalmente, si tratta di apprendimento non supervisionato perché non stiamo dicendo all'algoritmo in anticipo cosa costituisce ciascun tipo. Invece, forniamo i dati e chiediamo: *"Puoi trovare automaticamente una struttura in questi dati e capire quali sono i principali tipi di individui?"*

Poiché non forniamo la "risposta giusta" per gli esempi in anticipo, questo rimane saldamente nel dominio non supervisionato.



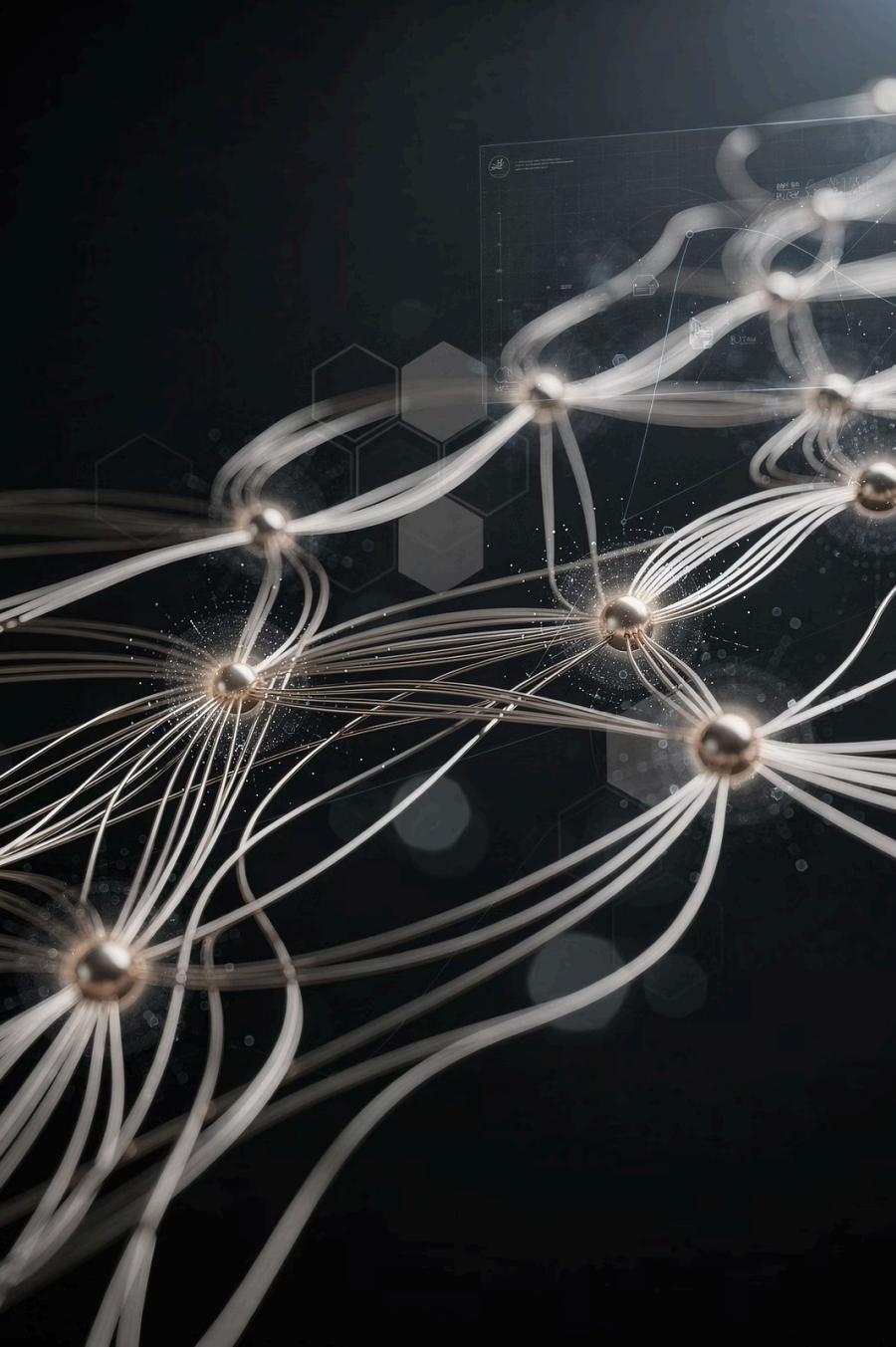


Applicazione di Clustering #3

Segmentazione del Mercato

Molte aziende gestiscono vasti database di informazioni sui clienti. Un approccio di apprendimento non supervisionato può raggruppare automaticamente i clienti in diversi segmenti di mercato, consentendo un servizio più efficiente e mirato.

Anziché definire manualmente le categorie di clienti basandosi su supposizioni, gli algoritmi di clustering scoprono raggruppamenti naturali all'interno dei dati – segmenti che riflettono genuinamente il comportamento e le motivazioni dei clienti.



Oltre il Clustering: Altri Metodi Non Supervisionati

Sebbene il clustering rappresenti una delle tecniche di apprendimento non supervisionato più comuni, è tutt'altro che l'unico approccio. L'apprendimento non supervisionato comprende una vasta famiglia di algoritmi, ciascuno progettato per estrarre diversi tipi di intuizioni da dati non etichettati.

Esploriamo altri due importanti tipi di apprendimento non supervisionato che risolvono problemi distinti nelle applicazioni di machine learning.

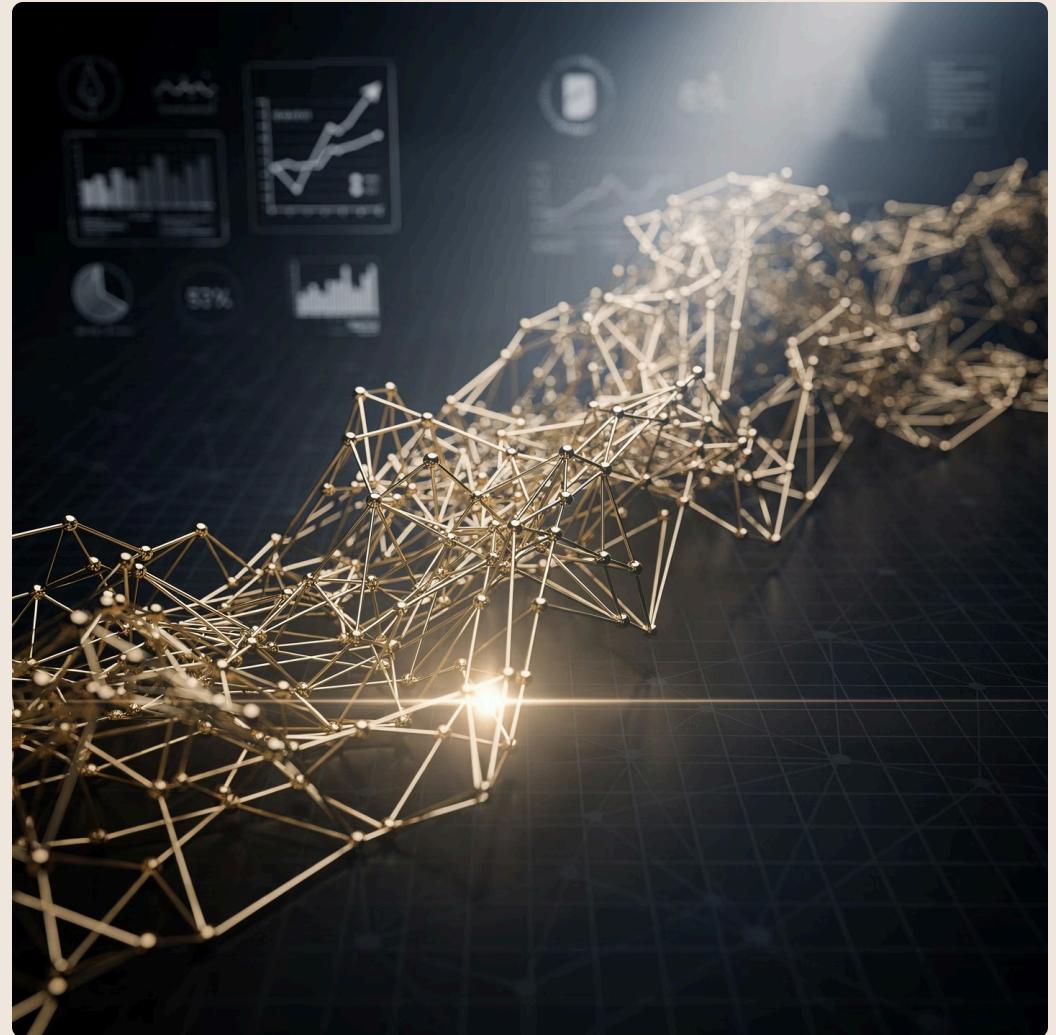
Tipo di Apprendimento Non Supervisionato #2

Rilevamento delle Anomalie

Identificare Eventi Insoliti

Gli algoritmi di rilevamento delle anomalie imparano come appare la "normalità" nei tuoi dati, quindi segnalano tutto ciò che devia significativamente da questo modello. Questi eventi insoliti spesso segnalano situazioni importanti che richiedono attenzione.

Questa tecnica si rivela essenziale per il rilevamento delle frodi nei sistemi finanziari, dove transazioni insolite potrebbero indicare attività fraudolente. È anche utilizzata per la sicurezza della rete, il controllo qualità della produzione e il monitoraggio sanitario.





Tipo di Apprendimento Non Supervisionato #3

Riduzione della Dimensionalità

Gli algoritmi di riduzione della dimensionalità compiono qualcosa che sembra quasi magico: prendono grandi e complessi set di dati e li comprimono in set di dati molto più piccoli, perdendo il minor numero di informazioni possibile.

Questo si rivela inestimabile quando si lavora con dati ad alta dimensionalità. Riducendo il numero di variabili pur preservando la struttura essenziale, questi algoritmi rendono i dati più facili da visualizzare, velocizzano altri algoritmi di machine learning e aiutano a identificare le caratteristiche più importanti nel tuo set di dati.

Le tecniche comuni includono l'Analisi delle Componenti Principali (PCA) e t-SNE, ciascuna con i propri punti di forza per diversi tipi di dati e applicazioni.

Tre Tipi di Apprendimento Non Supervisionato



Clustering

Raggruppa punti dati simili per scoprire categorie naturali e modelli all'interno di dati non etichettati



Rilevamento delle Anomalie

Identifica eventi insoliti o outlier che deviano dai modelli normali—critico per il rilevamento delle frodi e la sicurezza



Riduzione della Dimensionalità

Comprime grandi set di dati in rappresentazioni più piccole preservando le informazioni e la struttura essenziali

Verifica la Tua Comprensione

Quali dei seguenti sono esempi di apprendimento non supervisionato? Considera ciò che hai imparato sulla distinzione chiave: l'algoritmo ha accesso a dati di output etichettati (Y) o solo a dati di input (X)?

1

Filtro Antispam

Email etichettate come spam o non-spam, utilizzate per addestrare un sistema di classificazione

2

Raggruppamento di Notizie

Raggruppamento automatico di articoli di notizie correlati senza categorie predefinite

3

Segmentazione del Mercato

Scoprire automaticamente segmenti di clienti dai dati senza gruppi predeterminati

4

Diagnosi di Diabete

Classificare i pazienti come diabetici o meno basandosi su cartelle cliniche etichettate

Punti Chiave



L'apprendimento non supervisionato funziona senza etichette di output

L'algoritmo scopre la struttura e i modelli nei dati contenenti solo input (X), rendendolo ideale per l'analisi esplorativa e la scoperta di pattern.



Il clustering raggruppa automaticamente dati simili

Dall'aggregazione di notizie alla ricerca genetica alla segmentazione dei clienti, gli algoritmi di clustering trovano raggruppamenti naturali senza supervisione.



Esistono molteplici tecniche non supervisionate

Oltre al clustering, il rilevamento delle anomalie identifica eventi insoliti mentre la riduzione della dimensionalità comprime i dati in modo efficiente, ognuno soddisfacendo distinte esigenze analitiche.

L'apprendimento non supervisionato rimane una pietra angolare del machine learning moderno, permettendoci di estrarre intuizioni dalle vaste quantità di dati non etichettati disponibili nel mondo reale.