

Data Science efficiente

Linee guida per rendere l'innovazione digitale un viaggio di successo

Introduzione

Racconti da un mondo guidato dai dati

Da tempo le aziende private e pubbliche, così come i governi, parlano di dati e dell'opportunità di **investire sui dati**.

Eppure, il più delle volte, si trattava solo di parole, idee, ipotesi.

Il problema principale era ed è che di solito nessuno aveva capito che per investire sui dati è **assolutamente necessario un cambiamento profondo e dirompente della mentalità**.

Inoltre, investire sui dati richiede la **matematica**. Come vedremo, non c'è modo di investire sui dati senza la matematica. Un altro ingrediente obbligatorio è una dose adeguata di **tecnologia informatica moderna**, che è piuttosto rara.

Infine, la ricetta è completa se e solo se c'è la capacità di creare un **dialogo** costruttivo tra imprenditori e tecnici.

Poiché questi ingredienti mancavano, i risultati non potevano soddisfare le aspettative.

Vedendo che i risultati erano solo parziali, o a volte nulli, il comportamento tipico era quello di cambiare nome e approccio ideale, invece di affrontare le sfide più difficili rappresentate dalla matematica, dal cambio di mentalità e da una nuova tecnologia informatica.

Tutte queste dinamiche hanno portato ai tanti nomi che tutti conosciamo: data science, industria 4.0, intelligenza artificiale, innovazione digitale, advanced analytics, big data, ...

Una rivelazione fondamentale

Qualunque sia il nome dell'approccio, c'è un'affermazione specifica che descrive ciò che accade con i dati:

> Ogni moderna rivoluzione guidata dai dati è in realtà una rivoluzione *culturale* resa possibile dalla tecnologia.

Con culturale non si intende che basta studiare per ottenere questo risultato. Sì, è vero che anche voi dovete studiare, ma il punto principale è che ***la cultura*** si riferisce al fatto che dovete cambiare la vostra ***mentalità***.

Un altro punto importante riguarda la tecnologia: essa rende possibile la rivoluzione. Significa che non ci può essere alcuna rivoluzione senza un certo livello di investimenti in tecnologia.

I dati e la matematica erano già lì, in attesa di essere utilizzati e utilizzabili. La tecnologia è il nuovo attore che chiude il cerchio.

Un errore comune è quello di rifiutare di abbracciare una nuova mentalità. L'altro terribile errore è quello di non capire che una **tecnologia** adeguata **è obbligatoria**.

Una foresta di nomi

Negli ultimi decenni sono stati utilizzati molti nomi diversi, a seconda della moda del momento, per indicare una sorta di approccio guidato dai dati che prometteva di cambiare il mondo:

- Statistica industriale e quality statistics
- Six Sigma e Lean Six Sigma
- Large language models
- Dati sintetici
- Analisi dei grandi dati
- Intelligenza artificiale
- Deep learning
- Industria 4.0 / 5.0
- IA generativa

...E altri ne verranno in futuro.

In che modo questi termini sono collegati l'uno all'altro?

Ognuno di essi ha a che fare con un qualche tipo di **investimento sui dati**.

E tutte richiedono una **tecnologia** adeguata e, come scopriremo, la **matematica**. La matematica è lo strumento indispensabile per lavorare sui dati.

Fondamentalmente, qualunque sia il modo in cui ci si approccia ai dati, se si vuole *investire* sui dati è necessario disporre di tecnologia e matematica.

Ora si tratta di capire le differenze dettagliate tra i molti nomi e metodi che sono per lo più conosciuti e utilizzati.

La colonna più forte del tempio: Statistica

La statistica è una branca applicata della matematica. In realtà, tra i molti strumenti e discipline disponibili per lavorare sui dati, la Statistica è quella con la storia più lunga e la struttura migliore.

Esistono molti nomi per la statistica applicata, come ad esempio statistica industriale o statistica della qualità, ma la vera differenza è solo la denominazione, nient'altro.

La Statistica che ci interessa è stata sviluppata principalmente all'interno di alcune aziende, quindi è nata direttamente guidata da esigenze aziendali.

Le prime opere sono state realizzate da Guinness, Ford, AT&T, General Electrics e Bell Laboratories e risalgono agli anni tra il 1900 e il 1925.

Nel 1987 Motorola ha introdotto una metodologia denominata Six Sigma, concepita per utilizzare tutti i metodi statistici creati in precedenza attraverso un protocollo solido, con regole, linee guida e KPI.

Le statistiche hanno 3 aree principali:

1. Statistica descrittiva
2. Statistica inferenziale
3. Statistica matematica

Tutti questi elementi sono **estremamente importanti** per investire sui dati.

Cercare di utilizzare il machine learning o il deep learning senza conoscere la Statistica è come giocare a golf senza la pallina.

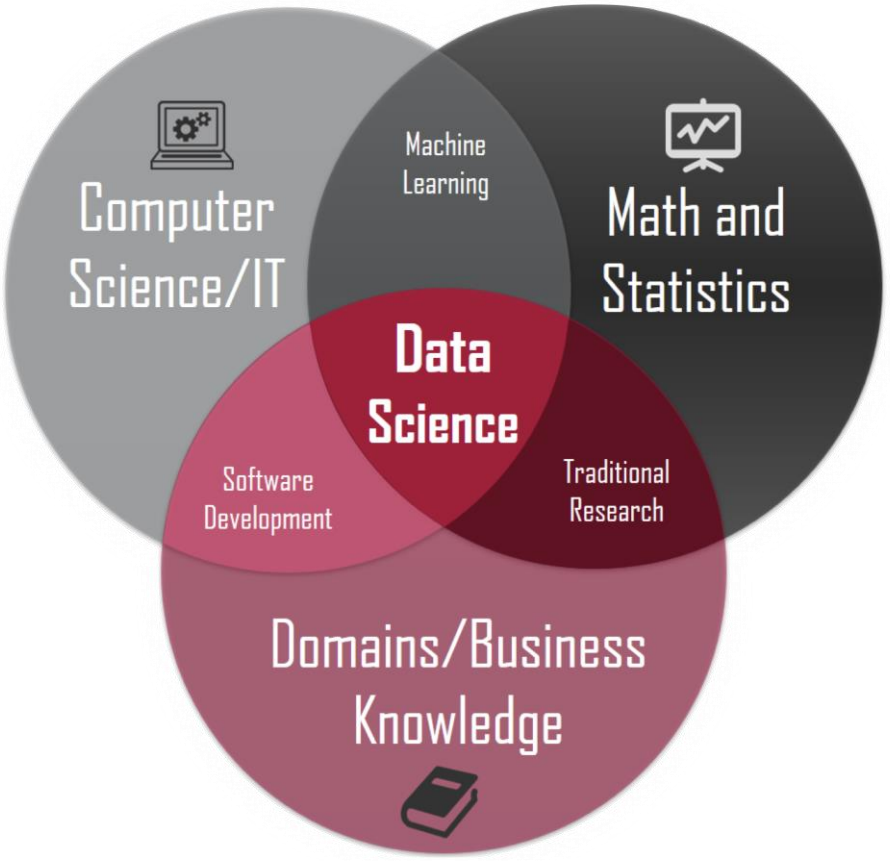
Data Science

La Data Science è la scienza e l'arte di trasformare i dati in ricavi.

La scienza dei dati è sempre orientata al business.

Le **implicazioni economiche** di un progetto di Data Science sono sempre di primaria importanza e un risultato è utile solo se può essere venduto con successo all'azienda. Ciò significa che la pura ricerca o i risultati ambigui *non* fanno parte della scienza dei dati.

Diciamo che è una scienza perché c'è molta matematica e tecnologia e diciamo anche che è una specie di arte, perché ogni progetto ha le sue sfide e richiede al data scientist creatività, immaginazione e capacità di collegare cose e concetti.



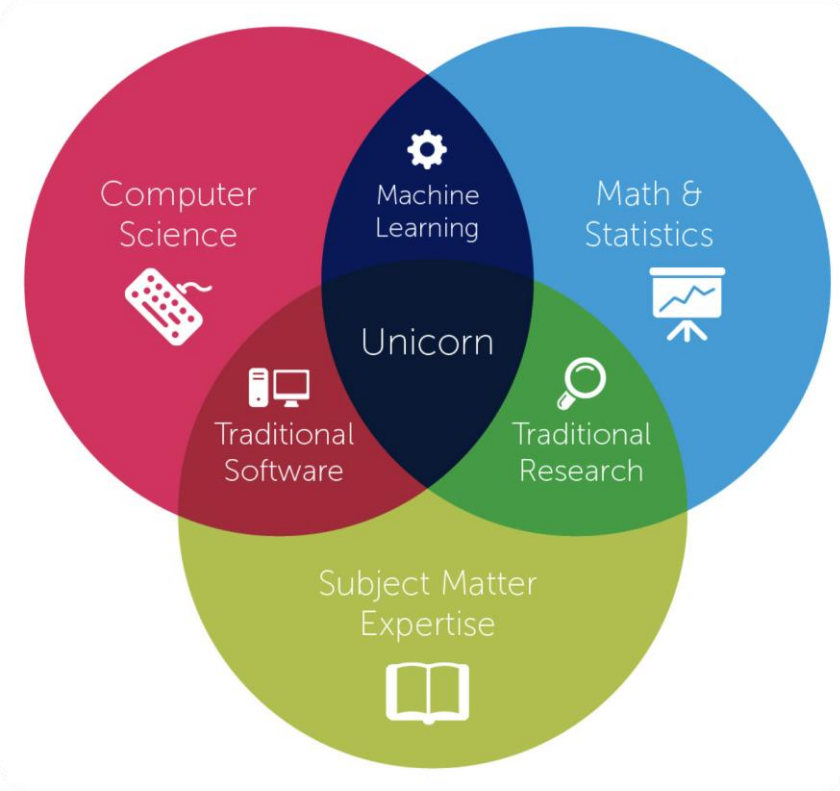
Data Science

Capita che i manager credano che un data scientist sia sufficiente e che possa fare magie incredibili con i dati.

In realtà, un solo data scientist è un non senso, soprattutto se d i giovane età.

Un team di Data Science deve essere composto da diversi profili:

- Esperti di matematica e statistica
- Ingegneri dei dati
- Ingegneri AI
- Sviluppatori
- Esperti del settore
- Responsabile di progetto

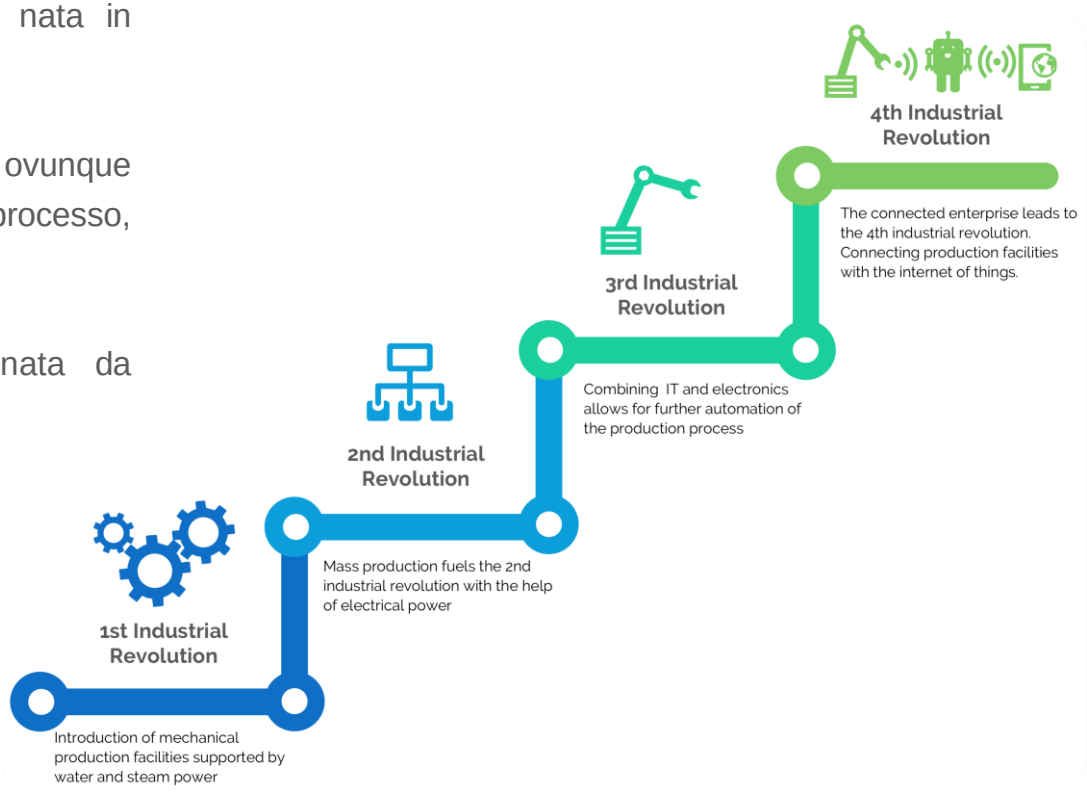


Industry 4.0

La quarta rivoluzione industriale risale al 2011, quando è nata in Germania.

Si basa sull'idea che tutto è connesso: i dati vengono creati ovunque e sono condivisi tra ogni possibile componente del processo, compresi gli esseri umani.

L'Industria 4.0 è del tutto inutile se non è accompagnata da un'adeguata tecnologia e da una corretta analisi dei dati.



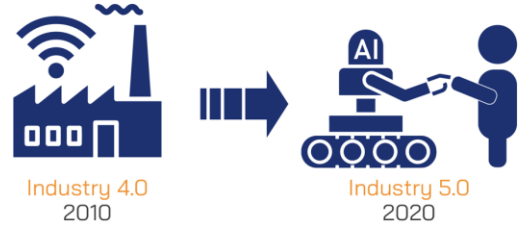
Industry 5.0

L'approccio proposto da Industria 5.0 fornisce una visione dell'industria che va oltre l'efficienza e la produttività come unici obiettivi e rafforza il ruolo e il contributo dell'industria alla società.

Pone il **benessere del lavoratore** al centro del processo produttivo e utilizza le nuove tecnologie per fornire prosperità oltre a posti di lavoro e crescita, **rispettando** i limiti di produzione del **pianeta**.

Il progetto integra l'attuale approccio "Industria 4.0" mettendo specificamente la ricerca e l'innovazione al servizio della transizione verso un'industria **sostenibile, incentrata sull'uomo** e **resiliente**.





AI, ML, DL, ...

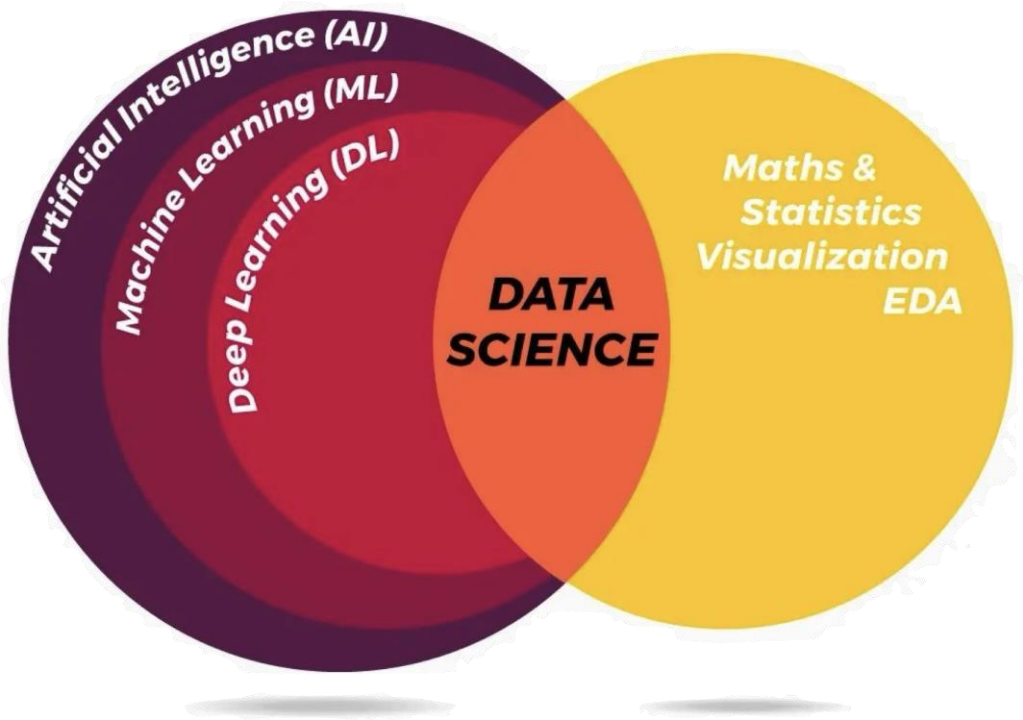
Intelligenza artificiale è un termine generico che indica un'intelligenza creata dall'uomo. È solo un concetto, non indica alcuna tecnologia o algoritmo specifico.

L'apprendimento automatico è un'applicazione dell'IA. Di solito ML indica un insieme di algoritmi creati per consentire a una macchina di imparare dai dati.

L'apprendimento profondo è un sottoinsieme del ML che si concentra sulle reti neurali. In particolare, si occupa di reti neurali profonde (cioè reti neurali con molti strati nascosti).

ML e DL si basano sulla matematica e richiedono una conoscenza rilevante della statistica.

La scienza dei dati fa un uso intensivo di tutte queste competenze.



Big Data

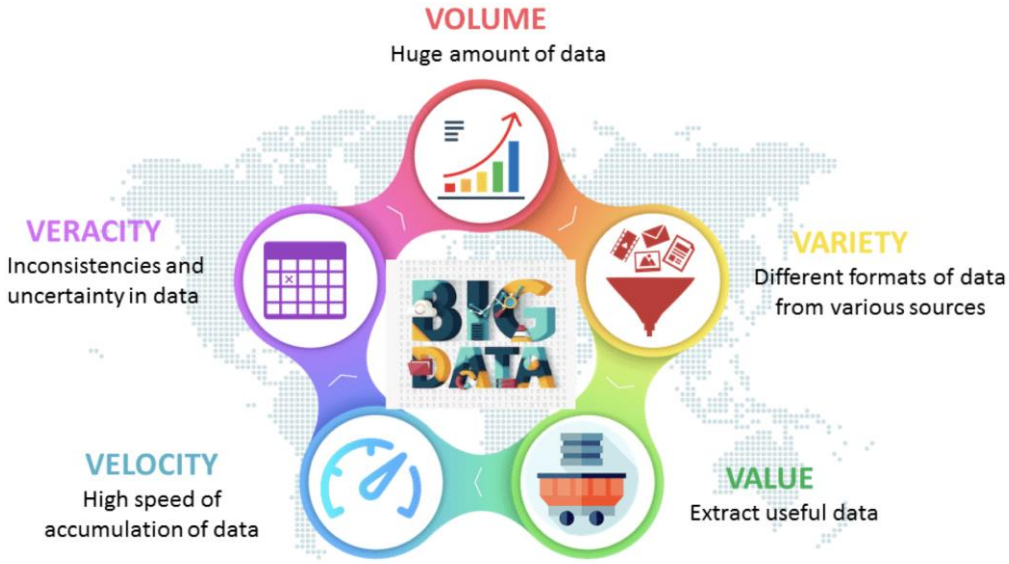


Che cosa significa in realtà *»big«*?

Forse i dati sono grandi quando non possono essere gestiti da una sola macchina? La risposta è che si tratta di big data quando è necessaria una *tecnologia* big data.

Per tecnologia dei big data si intendono soluzioni basate su Hadoop e/o Spark.

La Data Science è l'unica soluzione per capire se i dati sono davvero grandi e se richiedono una tecnologia specifica. Spesso vengono raccolti molti più dati del necessario.

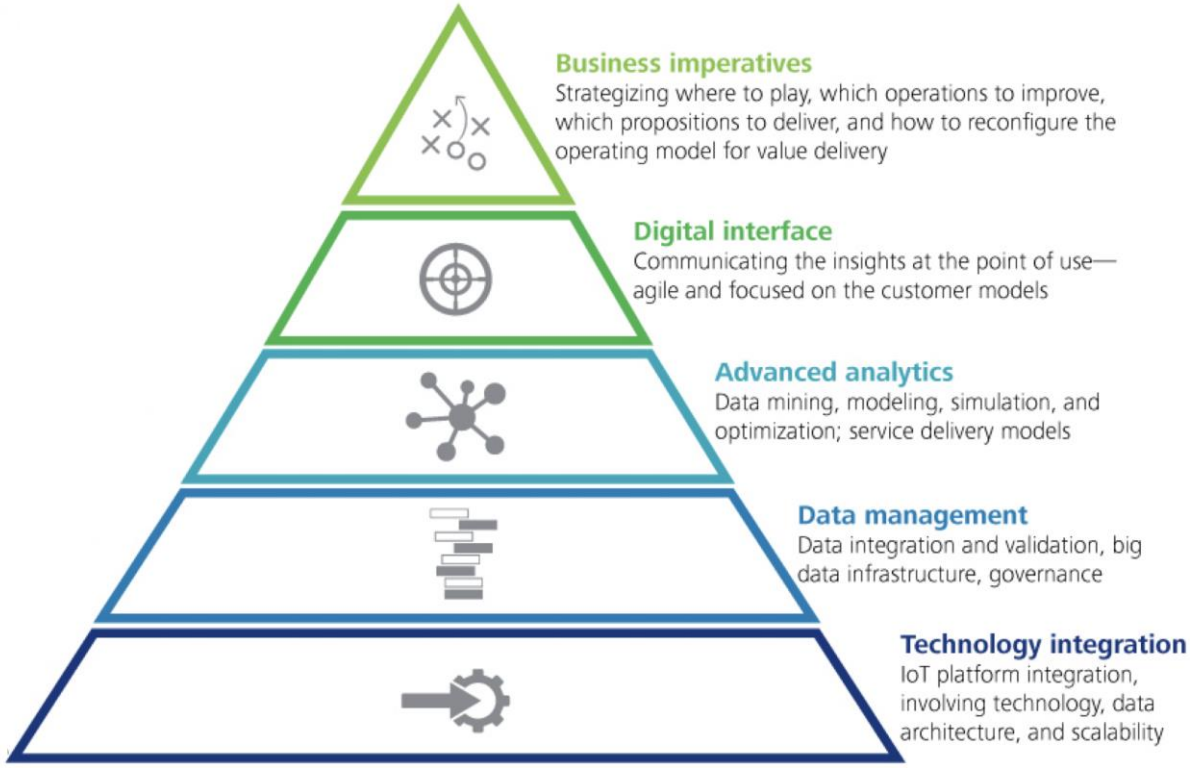


Innovazione digitale

L'innovazione digitale, o trasformazione digitale, consiste nell'applicare soluzioni digitali alla vita umana, al fine di apportare un miglioramento.

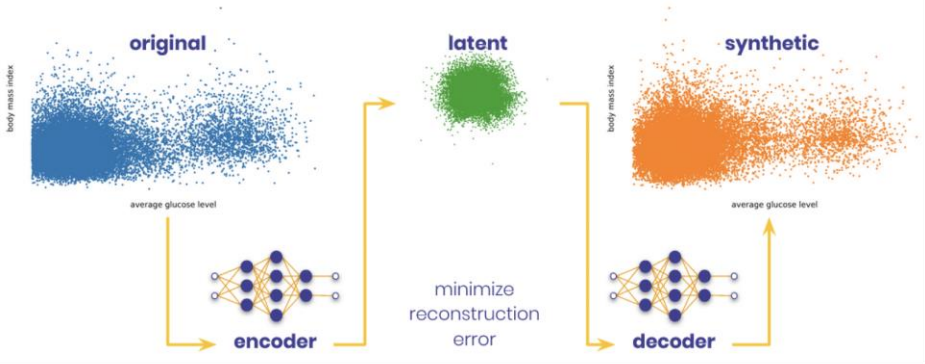
Il punto fondamentale è che i dati fanno sempre parte del gioco e quindi la Data Science è un ingrediente obbligatorio dell'innovazione/trasformazione digitale.

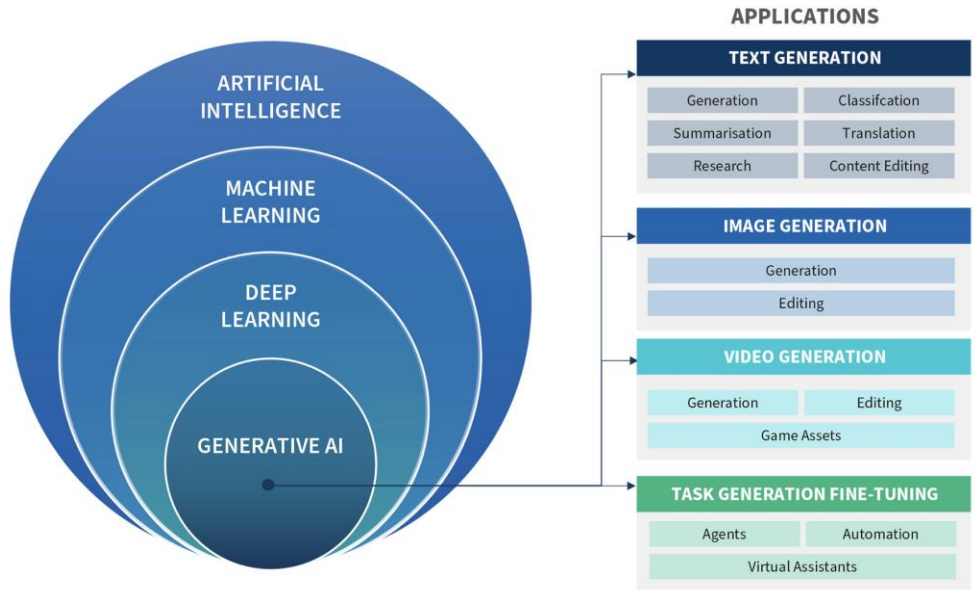
Al giorno d'oggi, l'innovazione digitale è la nuova parola d'ordine che racchiude tutto: data science, AI, big data, ...



IA generativa e dati sintetici

Tipicamente basata sull'apprendimento profondo, l'**IA generativa** apprende "l'essenza" di una data entità reale e poi genera nuove istanze dello stesso tipo di entità.





L'IA generativa è strettamente legata, dal punto di vista concettuale, ai **dati sintetici** che trovano sempre più applicazione nell'apprendimento automatico quando i dati

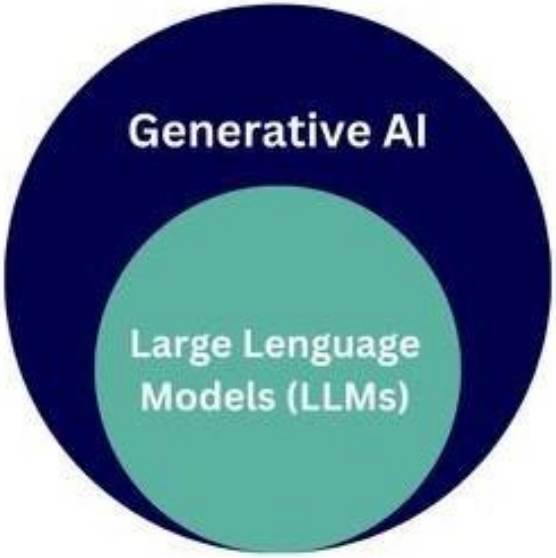
Large Language Models

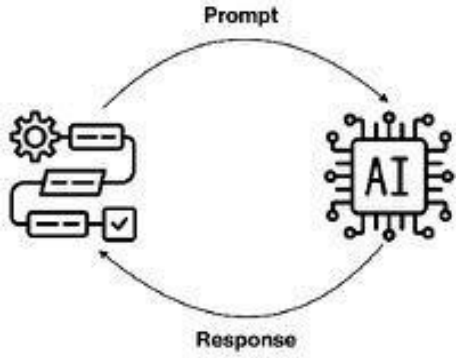
I LLM sono un sottoinsieme dell'IA generativa. Sono specializzati nel linguaggio umano. Il loro scopo è comprendere il linguaggio umano ed essere in grado di interagire con esso in modo dinamico e "intelligente".

Ogni LLM si basa su grandi reti neurali profonde addestrate su enormi set di dati.

È facile intuire che gli LLM avranno un ruolo rilevante nell'evoluzione dirompente dell'apprendimento automatico e del modo in cui lo utilizziamo.

L'interfaccia di base si chiama **prompt** e non è altro che un campo di immissione di testo in cui l'utente umano può inserire un testo per fornire una descrizione del compito e parametri dettagliati all'intelligenza artificiale.





Leggi fondamentali

Legge 1: i 3 livelli della Data Science

Ogni strato ha un'importanza fondamentale.

Esiste una gerarchia tra i livelli, che pretende di essere osservata.

Ogni livello richiede competenze e tecnologie specifiche.

Se manca uno di questi livelli, non si tratta di Data Science.

Dal punto di vista della direzione aziendale, è importante capire che sono necessarie competenze distinte, ma complementari.

In particolare, nei tempi attuali, il data lake è quasi sempre assente e incompreso. Poiché è la base di tutto, un'enfasi sul data lake è
da raccomandare.



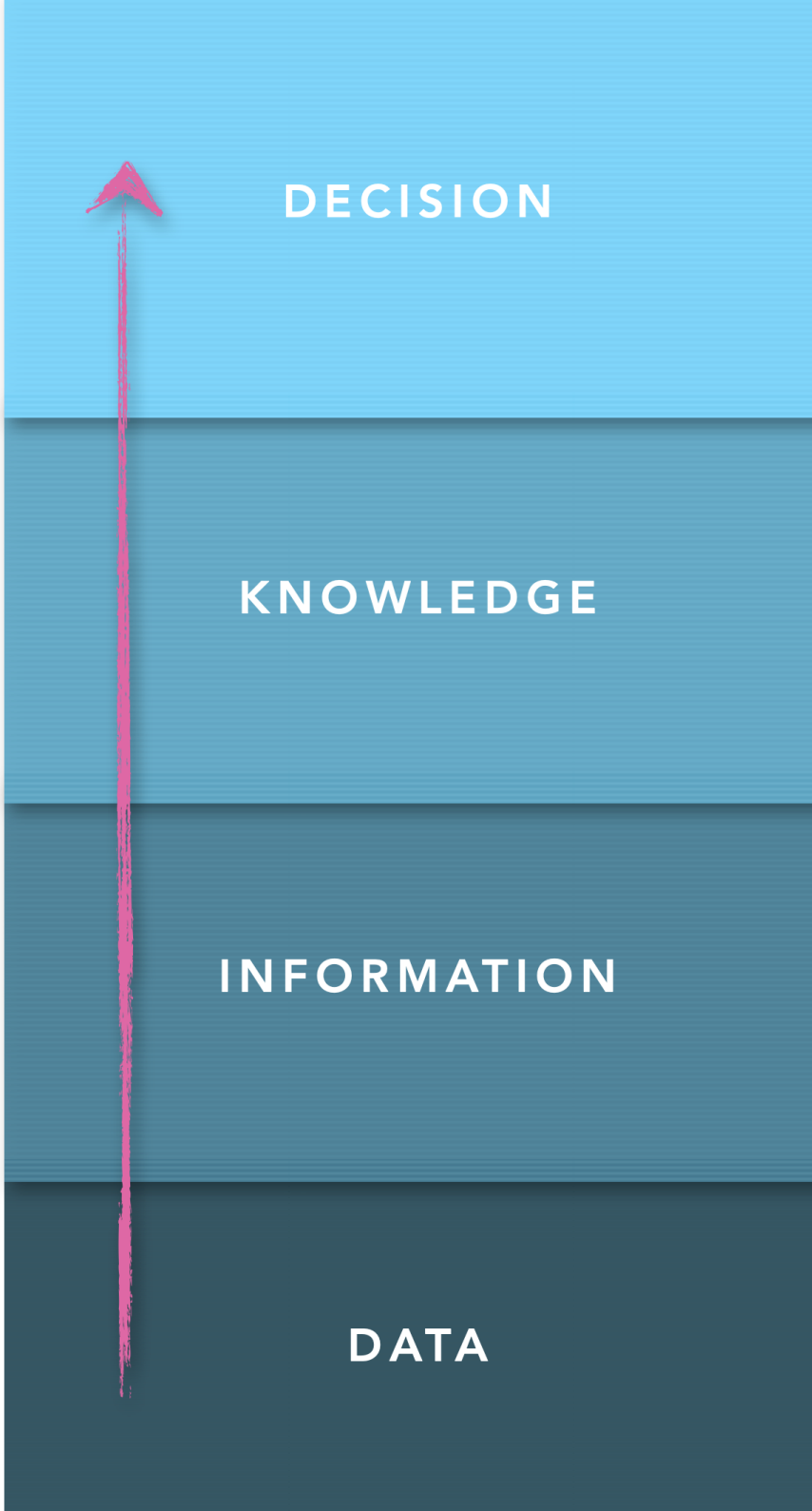
Legge 2: Alchimia dei dati

I dati sono solo dati, semplicemente energia potenziale. È l'equivalente del piombo nell'antica alchimia. Se cerchiamo l'oro, come dovremmo lavorare sui dati?

Questa è l'alchimia dei dati: trasformare i dati in decisioni guidate dai dati (oro).

Per passare dai dati all'informazione, la matematica è obbligatoria: non c'è informazione dai dati senza matematica.

L'alchimia dei dati ci porta dalla dura tecnologia dell'informazione al business e al top management. In mezzo c'è la Data Science con i suoi 3 livelli.



Cultura dei dati

La triste storia del maltrattamento dei dati

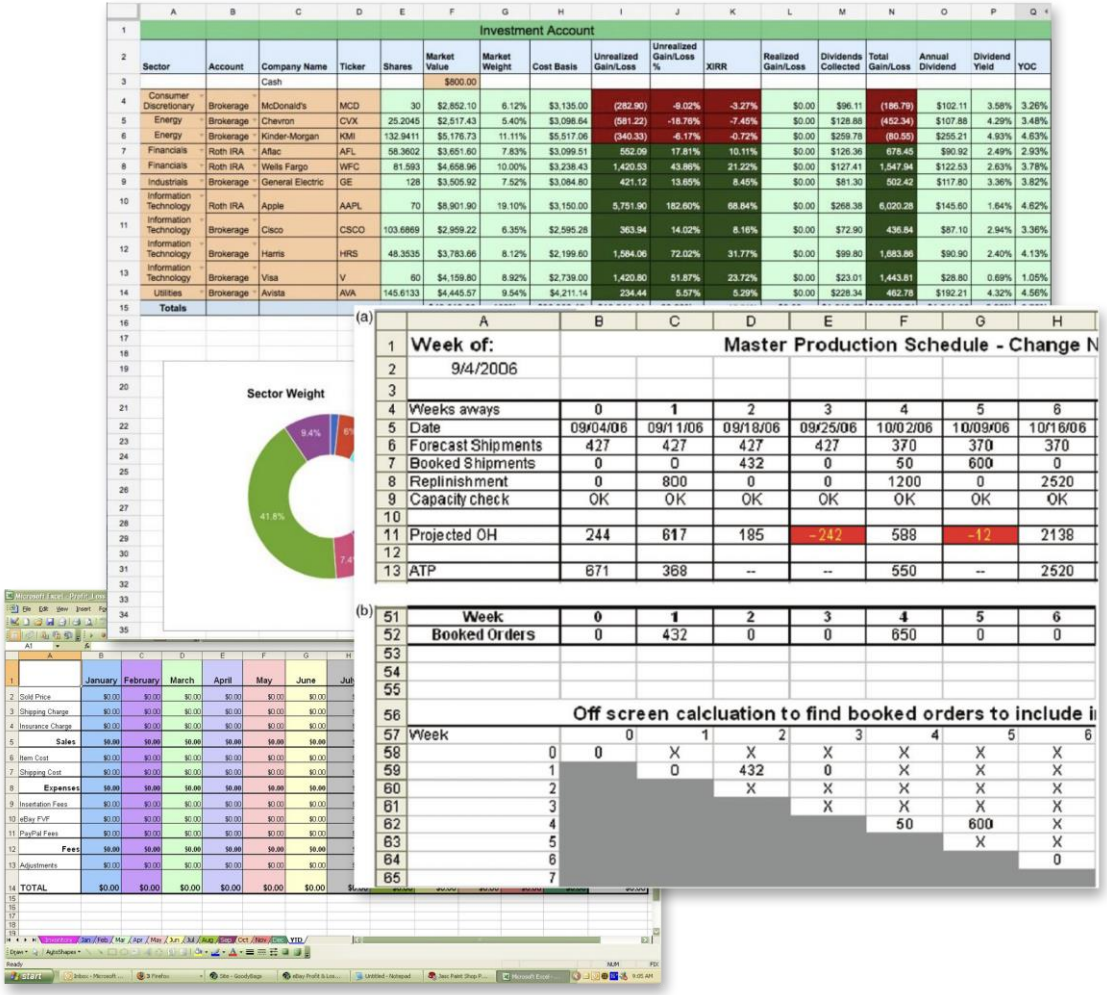
È estremamente comune vedere i dati raccolti in file Excel in modi simili alle immagini a destra.

Excel e qualsiasi altro software simile, dal punto di vista della Data Science, sono solo il male puro. E peccato.

I dati non devono mai essere mescolati con l'analisi. I colori e la formattazione non hanno senso se collegati ai dati.

I dati sono dati.
I dati non sono una presentazione. I dati non sono analisi.

I dati devono essere tenuti separati dall'analisi e dalla



Alcune regole salva vita sui dati

- Più i dati sono leggibili dall'uomo, meno saranno utilizzabili da computer e algoritmi e viceversa.
- Gli esseri umani non leggono i dati, i computer e gli algoritmi sì.
- Il modo migliore per organizzare i dati è per colonne, poiché è estremamente compatibile con gli strumenti analitici e con i linguaggi di programmazione.
- Le date sono difficili: troppi formati. Il formato più sicuro è aaaa-mm-gg.
- I separatori decimali sono pericolosi: è meglio seguire la notazione inglese con "." piuttosto che ",".
- I dati mancanti devono essere gestiti con un metodo coerente, senza modificare il tipo di dati della colonna.
- Lo zero è un numero, non un segnaposto per i dati mancanti.
- I dati non hanno bisogno di colori: ciò che è descritto da un colore deve essere elencato in una colonna dedicata.
- I dati non hanno bisogno di essere formattati, ma di essere normalizzati e deduplicati.
- I dati non vivono in Excel, ma nel data lake.
- I nomi delle colonne devono essere privi di caratteri speciali e possibilmente chiari e semplici.
- Le variabili categoriali necessitano di un dizionario
- I campi liberi tendono a diventare un incubo per le analisi
- Anche se il modello di dati relazionale non è l'unico, né il migliore, è un ottimo esercizio per ottenere un punto di vista analitico.

Verso un modello

Come razionalizzare un sistema e identificare i dati

Per passare dai dati alle informazioni è necessaria la matematica. Pertanto, la prima cosa da fare è razionalizzare il sistema che si intende studiare, al fine di identificare gli elementi che saranno utilizzati per creare un modello matematico.

In altre parole, dobbiamo pensare in termini di dominio e codominio: quali sono le variabili indipendenti (x)? Quali sono le variabili dipendenti (y)?

Dobbiamo identificare gli ingressi (x) e le uscite (y).

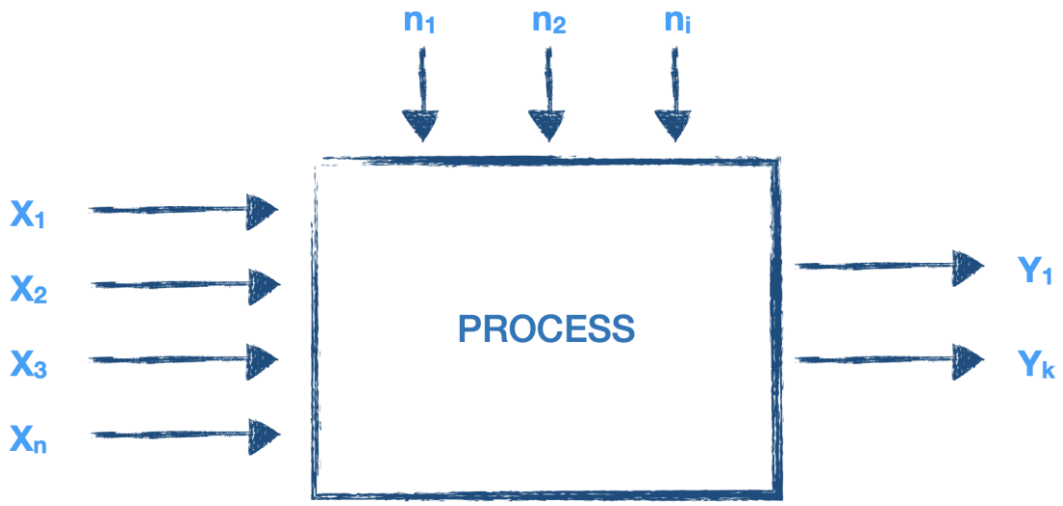
Occorre creare una funzione che descriva y sulla base di x.

Buone domande da cui partire sono:

- Quali variabili sono in ingresso al mio processo?
- Quali degli input possono essere misurati e controllati e quali possono essere solo misurati?
- Quali sono gli output del mio processo?
- Siamo in grado di misurare gli input misurabili e tutti gli output?
- Cosa vogliamo fare con un dato output? Previsione? Classificazione?
- I dati sono già disponibili o disponibili in qualche modo?
- Quanti dati saranno disponibili?
- Quante dimensioni (cioè colonne) avremo?

Come razionalizzare un sistema e identificare i dati

La seguente rappresentazione è un ottimo aiuto per definire il sistema che si vuole studiare:



In questo schema:

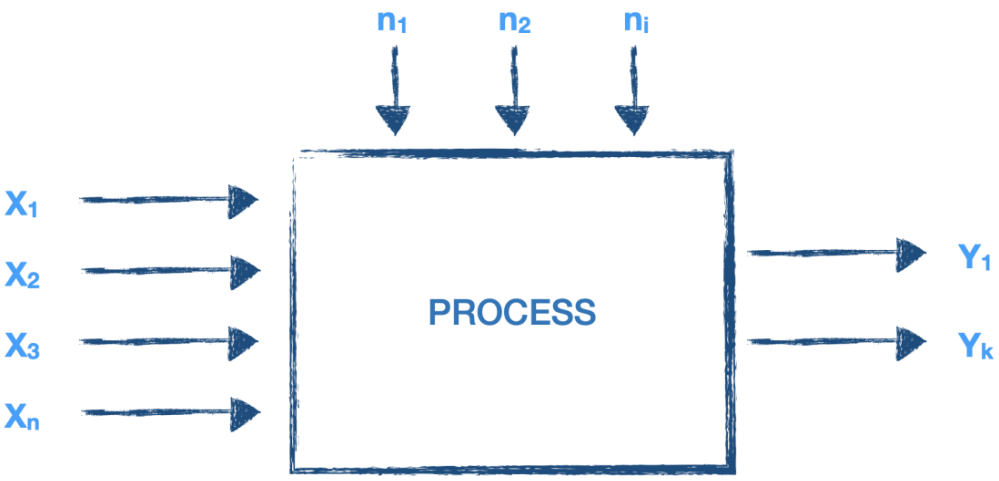
- X sono gli input che si possono misurare e controllare
- N sono gli input che si possono solo misurare. Storicamente sono chiamati fattori di rumore e in tempi moderni sono chiamati covariati.
- Y sono gli output che si vogliono prevedere, classificare, ecc.
  in funzione di X e N

Alla fine, ciò che cerchiamo per ogni singolo Y è un modello matematico che restituisce Y in funzione di X e N:

$$Y = f(\vec{X}, \vec{N})$$

Un errore comune...

Questo approccio, nato decenni fa nella statistica industriale, è generalmente sconosciuto a chi non lavora con i modelli matematici.



Ciò che accade è che un cliente che si rivolge al team di Data Science si concentra principalmente e solo sul processo.

Sbagliato!!!

Il processo stesso è secondario all'inizio. Quello che dobbiamo sapere è quali dati entrano e quali escono dal processo.

All'interno della scatola del processo possono accadere un sacco di cose tecniche, che sono inutili per definire il problema in ordine
per affrontare la questione in modo guidato dai dati.

Abbiamo bisogno dell'elenco di X, N, Y, dei loro tipi di dati, del modo in cui vengono misurati, della variabilità,

Intelligenza artificiale

L’intelligenza artificiale non è intelligente

L’intelligenza artificiale è stupida.

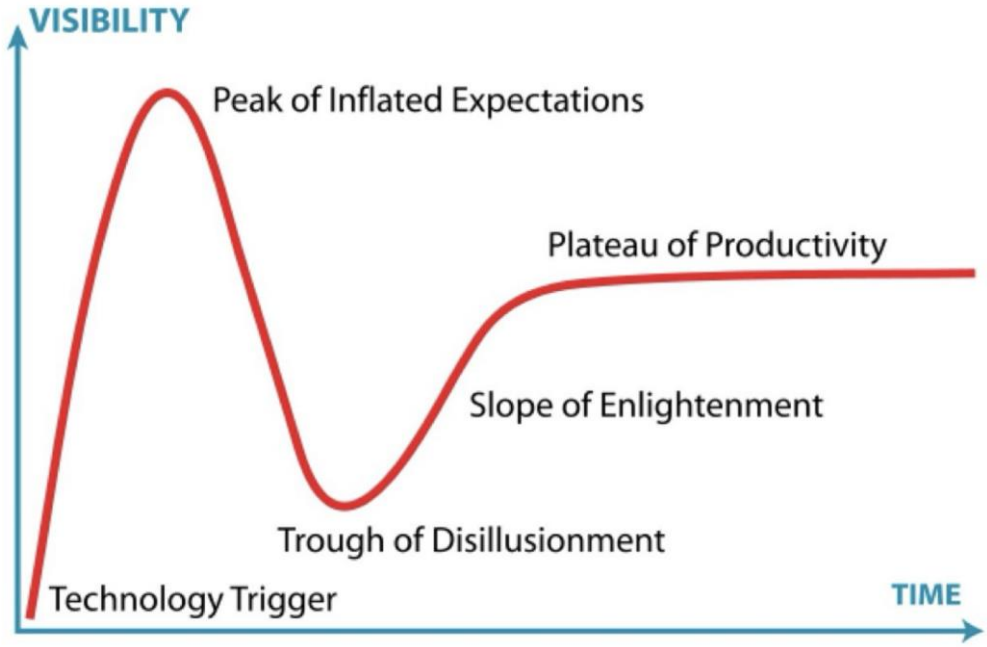
Sì, è così.

Il ciclo di hype generato dall'IA e dalla scienza dei dati è sempre attuale: le persone pensano che l'IA possa fare magie, ma non è così.

Non esiste una vera intelligenza nell'IA.

Un bambino di 3 anni è molto più intelligente di un modello avanzato di deep learning.

Il problema principale è che l'IA non ha coscienza di ciò che sta facendo, è come una persona stupida che svolge un lavoro secondo istruzioni specifiche e dettagliate che non ha capito.



Non abbiamo alcuna capacità di codificare la coscienza, quindi i nostri modelli sono necessariamente inconsapevoli

Tuttavia l'intelligenza artificiale è utile

Ci sono molti compiti per i quali l'intelligenza biologica non è un requisito essenziale. Per molti compiti reali è molto più importante evitare la soggettività e utilizzare un'esperienza ampia e selezionata su cui addestrare alcune regole statistiche.

L'intelligenza artificiale è ottima con le statistiche e

non ha sentimenti. Queste caratteristiche sono

ottime per molte applicazioni.

In alcuni casi l'intelligenza artificiale sarà persino migliore (o molto migliore) degli esseri umani nei problemi di classificazione.

L'intelligenza artificiale sarà in genere più veloce degli esseri umani e non avrà mai bisogno di riposare.

Inoltre, gli esseri umani hanno una bassa capacità di rilevare anche le interazioni semplici e non sono in grado di rilevare le interazioni complesse.

L'intelligenza artificiale è perfetta per questo tipo di compiti.

Velocità, automazione e obiettività sono le caratteristiche più rilevanti.

Automazione significa che l'IA è in grado di fare autonomamente ciò che le è stato insegnato a fare.

Ciò non significa che l'IA possa imparare autonomamente nuove cose arbitrarie.

Possiamo codificare l'IA in modo che impari dagli errori, ma gli esseri umani devono verificare che l'IA abbia imparato correttamente le cose nuove.

Non possiamo fidarci ciecamente dell'intelligenza artificiale.

Set di addestramento e set di test

Poiché dobbiamo insegnare qualcosa all'intelligenza artificiale, abbiamo bisogno di un set di addestramento.

Il set di addestramento è un insieme di dati selezionati che verrà preparato per addestrare l'algoritmo. Qualunque sia l'algoritmo, è necessario fornire un set di addestramento. Senza dati, niente modello.

La preparazione di un set di allenamento è sempre un compito lungo e cruciale.

Si tenga presente che il modello apprenderà solo ciò che si trova nell'insieme di addestramento. Pertanto, l'insieme di addestramento deve essere rappresentativo della realtà che si desidera descrivere con il modello.

Inoltre, l'insieme di allenamento deve essere generalmente bilanciato ed evitare valori vuoti.

Il set di allenamento non è l'unico set necessario.

Un altro set obbligatorio è il set di test: si tratta di un insieme di dati, dello stesso tipo del set di addestramento, che non viene mai utilizzato per addestrare il modello.

Quindi, ci si allena sul set di addestramento e poi si esegue il test sul set di test.

Le prestazioni differiscono troppo tra l'allenamento e il test? Non dovrebbe. Ovviamente una differenza apparirà sempre. Il punto è tenere sotto controllo la differenza.

Minore è la differenza, maggiore è la capacità di generalizzazione. Si tenga presente che la generalizzazione avviene su dati dello stesso tipo e della stessa distribuzione di quelli utilizzati per l'addestramento/test.

Set di validazione e validazione incrociata

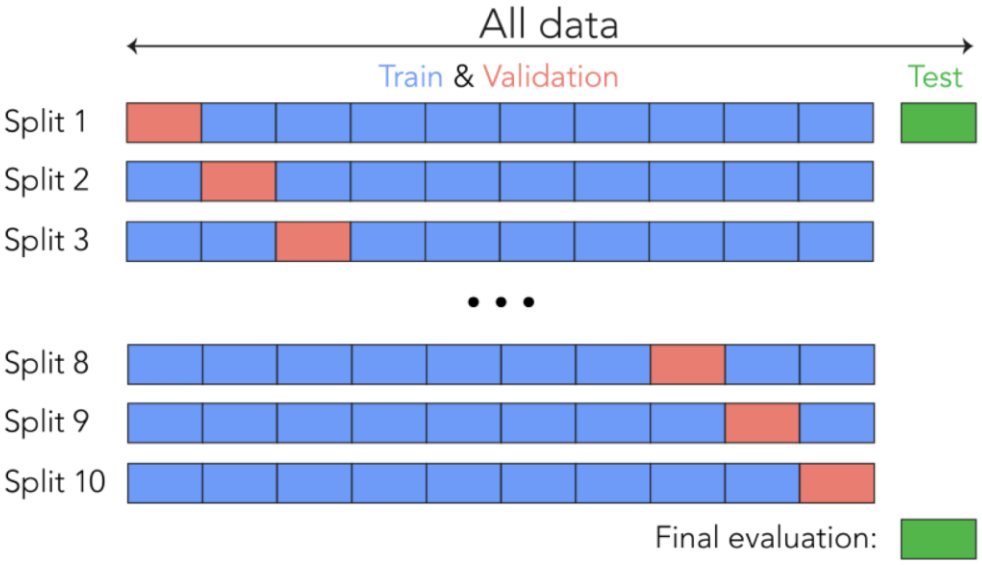
Un altro possibile insieme è l'insieme di

convalida. La logica è:

1. Allenarsi sul set di allenamento
2. Test sul set di validazione per regolare i parametri del modello e migliorarne le prestazioni.
3. Una volta ottenuta la convalida, testare il risultato finale sul set di test.

Quando la numerosità dei dati non è elevata, può essere utile un altro approccio chiamato validazione incrociata K-folds.





L'intelligenza artificiale in pratica

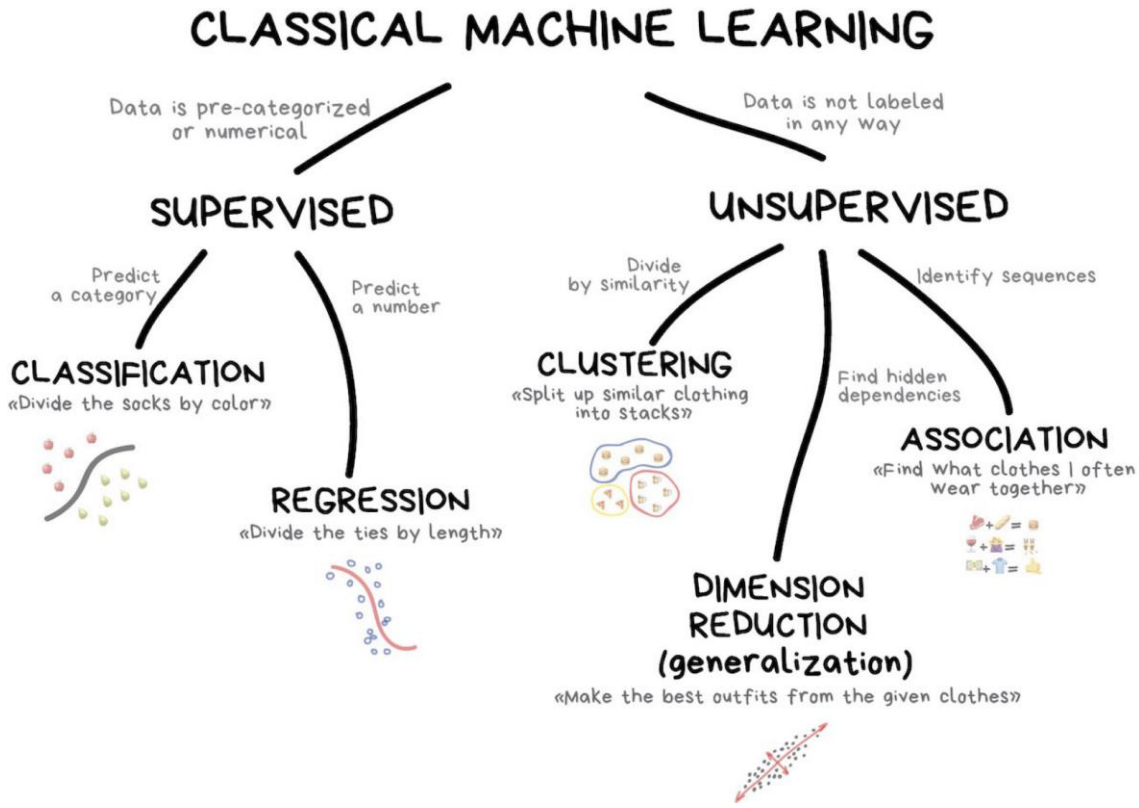
Con e senza supervisione

Da un punto di vista metodologico puramente classico, ogni algoritmo appartiene a una delle due famiglie seguenti:

- SUPERVISO
- NON SUPERVISIONATO

L'apprendimento supervisionato avviene quando l'algoritmo apprende da un set di addestramento che contiene indicazioni esplicite sul comportamento delle cose. Si dice che l'insieme di dati di addestramento è etichettato.

L'apprendimento non supervisionato avviene quando il set di dati non è etichettato. Non forniamo alcuna informazione a priori e lasciamo che l'algoritmo identifichi alcune possibili regole.



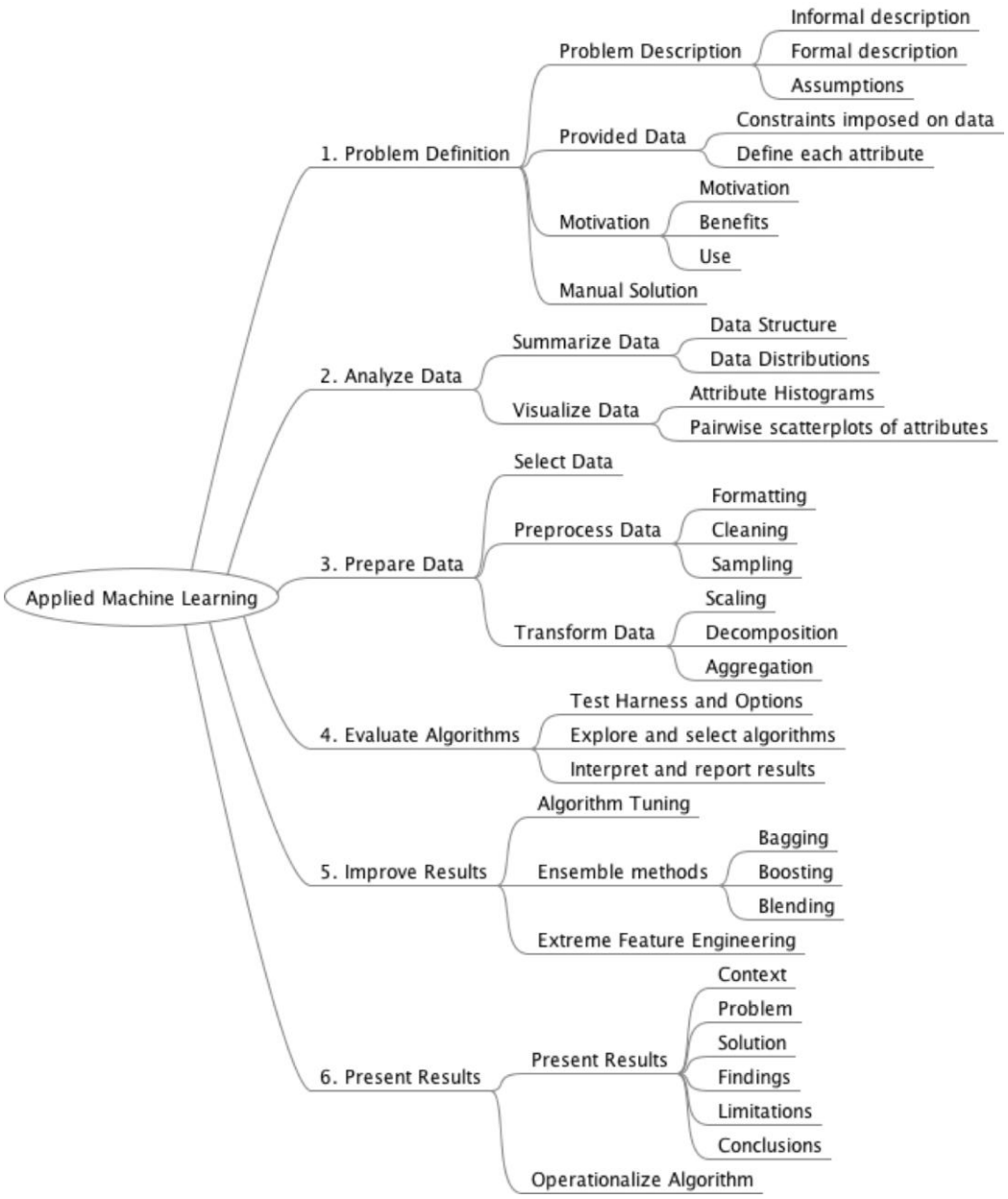
Andare sul concreto

Tutti i concetti precedenti potrebbero confondere, a meno che non si ricordi che la linea guida migliore è sempre quella di partire dalla definizione del problema.

Allora e solo allora, dato ciò che vogliamo fare, quali dati abbiamo, quanti dati abbiamo, identificheremo i possibili algoritmi da utilizzare.

Linea guida:

1. Definire ciò che l'azienda vuole
2. Osservare i dati per capire cosa si può fare
3. Selezionare le armi possibili (algoritmi)
4. Combattere...!



Quale algoritmo per cosa

Esistono fondamentalmente 4 tipi *principali* di algoritmi:

- Analitico: trova un'espressione analitica che modelli i dati. Utilizzabile per regressione e classificazione
- Basato su albero: modella i dati attraverso un approccio ad albero. Utilizzabile per regressione e classificazione
- Bayesiano: valuta le probabilità condizionali secondo la regola di Bayes. Utilizzabile per la classificazione
- Reti neurali: formano neuroni artificiali e le loro connessioni. Utilizzabili per regressione, classificazione, generazione e rappresentazione.

Non perdetevi in nomi, definizioni, raggruppamenti, schemi e teorie!

Prima di tutto concentratevi su ciò che volete fare e sui

dati! Poi qualcun altro identificherà i possibili algoritmi. Il

business è all'inizio e alla fine del processo.

Nel mezzo si trovano, nell'ordine: dati, algoritmi, usabilità. Ovvero i 3 livelli della scienza dei dati.

Alcune considerazioni generali

- Mantenete la semplicità e la bellezza!
- Non innamoratevi di algoritmi, teoria e tecnologia. Rimanete innamorati del successo
- In letteratura esistono moltissimi algoritmi: la maggior parte di essi è un esercizio puramente accademico ed è difficile da spiegare e condividere con le aziende. Evitate questo tipo di algoritmi
- Il successo non è trovare un modello a tutti i costi! Il successo è trovare un modello che funzioni abbastanza bene, che sia facile e conveniente da mantenere, che possa essere compreso, che possa essere venduto alle aziende e che sia in grado di soddisfare le loro esigenze.
- Un altro possibile successo è l'identificazione di questioni rilevanti che rendono impossibile la creazione di un modello adeguato secondo il punto precedente.
- Ricordate che la matematica è solo un freddo strumento di modellazione. Un modello matematico è inutile se la fisica che descrive non ha senso.
- Se non è possibile trovare un modello, riconsiderare la definizione del problema. Imparare dalla scienza: suddividere il problema globale in problemi più piccoli e più semplici.
- Ricordate che ML e DL non sono nulla senza la statistica. Partite sempre da un'analisi statistica dei dati
- In molti casi aziendali, è molto probabile che la soluzione sia già nota da decenni nel campo della statistica e dell'apprendimento statistico.
- La fisica ha un sacco di modelli matematici che hanno senso. Forse qualche buon approccio può essere trovato in Fisica.

I modelli matematici

Spesso le persone hanno problemi a collegare il concetto di modello matematico a qualcosa di pratico.

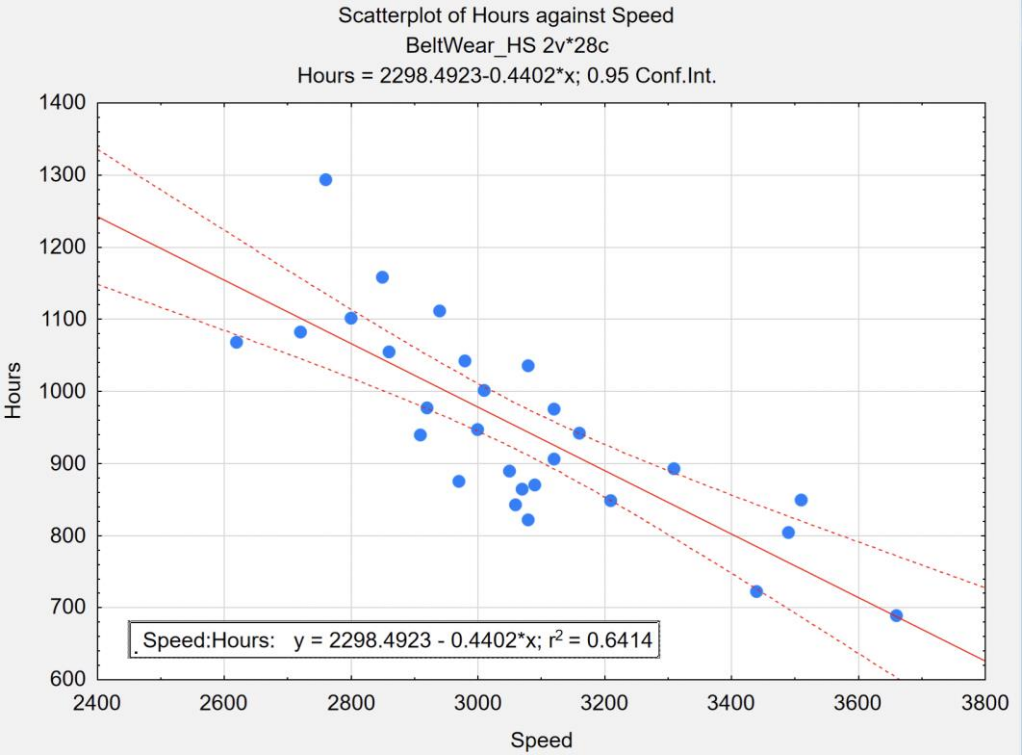
In effetti, i modelli matematici sono estremamente pratici!

Forse si tratta solo di capire cosa sia effettivamente un modello matematico.

La regressione lineare è un esempio molto semplice, ma perfetto, di modello matematico:

1. Raccogliere alcuni dati sperimentali
2. Adattare un modello lineare che descriva Y come funzione di X

Il modello può ora essere utilizzato per generalizzare il comportamento del sistema, prevederlo con un intervallo di confidenza e ottimizzare la variabile dipendente (Y) agendo sulla variabile indipendente (X).



Trovare il *vero* modello

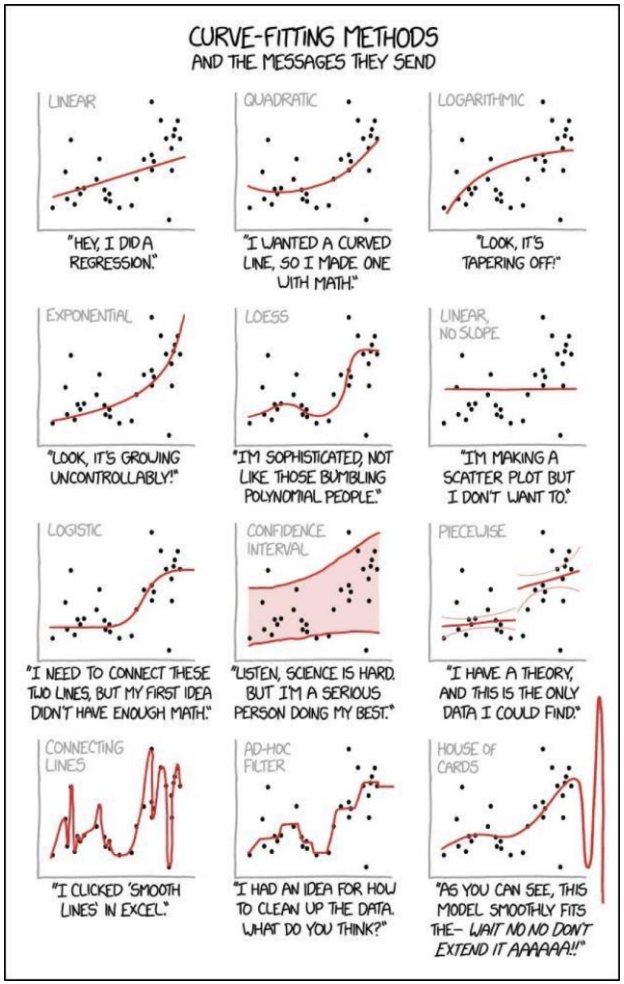
Un modello è solo un modello, non pretende di dire una verità assoluta. L'obiettivo di un modello è descrivere la realtà con una sicurezza sufficiente a soddisfare le esigenze aziendali. Nulla di più, nulla di meno.

In questo modo, un modello non potrà mai violare la fisica. Siate sempre consapevoli del significato fisico del vostro modello.

Competenza nel dominio, scienza, buon senso e logica: tutti devono contribuire alla scelta e all'accettazione di un modello.

Non lasciatevi mai guidare solo dai dati e dalla matematica!

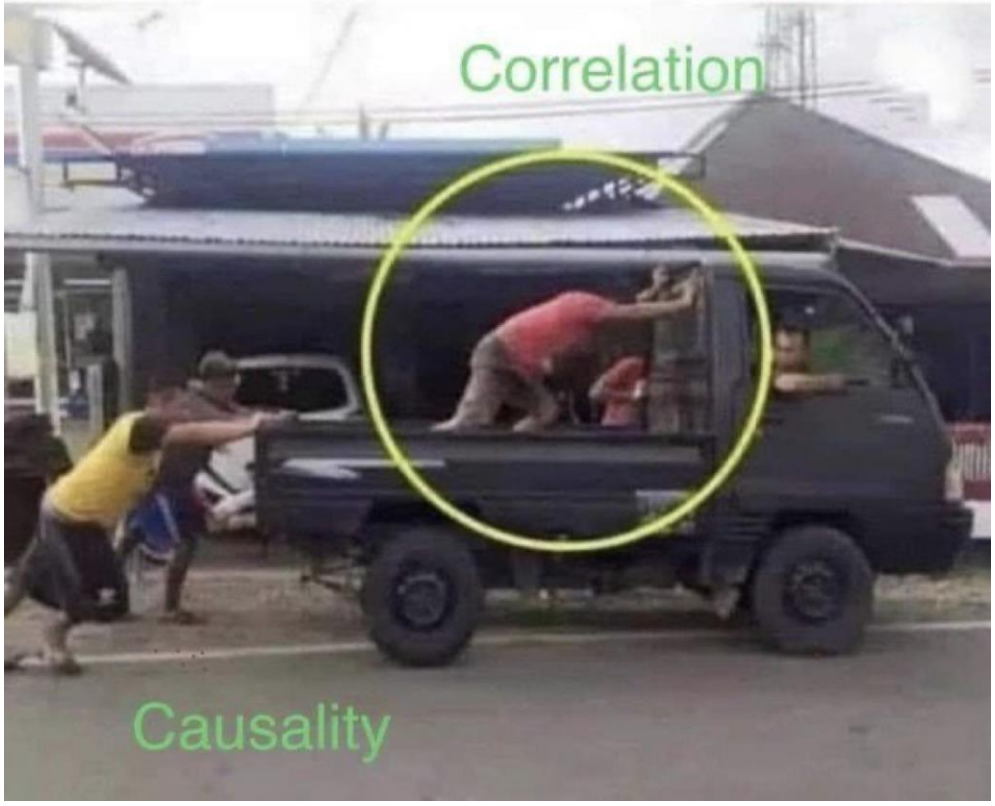
Ricordate che l'intelligenza artificiale è stupida perché non è consapevole di ciò che sta facendo. Voi sapete cosa state facendo!



Errori comuni

C'è molta confusione su alcuni concetti e termini. In particolare su media, causalità, relazione, correlazione...

- Di solito i modelli spiegano come e quanto, non perché. Il perché è lasciato agli esseri umani
- Media e deviazione standard sono falsi amici
- La correlazione è una parola chiave legata a una formula specifica.
- Il 99,9% delle volte in cui si dice "correlazione", in realtà si intende "relazione".
- Trovare relazioni tra gli input e tra gli input e gli output è molto più complicato del previsto, perché le relazioni potrebbero essere mediate da interazioni.



...interazioni?

Pensate alla cottura di una torta. Considerate l'umidità come risultato e la temperatura e il tempo come input.

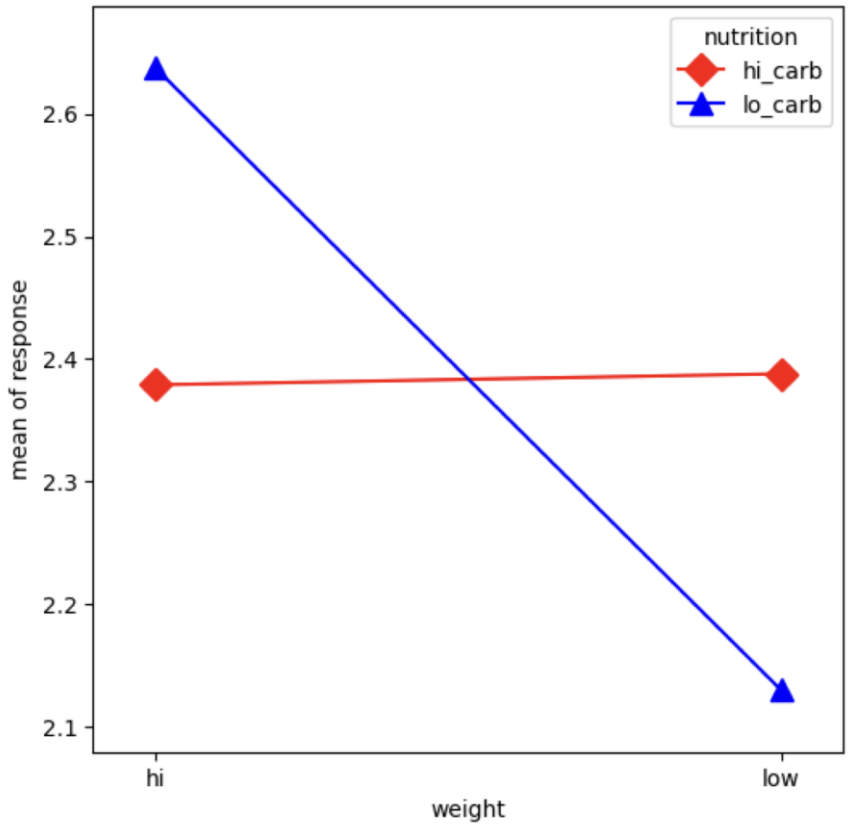
La temperatura è un descrittore significativo?

Sì. Il tempo è un descrittore significativo? Sì.

Considerate ora quanto siano diversi gli effetti del tempo e della temperatura sull'umidità, se si cambiano i loro valori reciproci. 45 minuti vanno bene, ma non se la temperatura è di 60° C.

Le interazioni non possono essere misurate direttamente. Possono essere solo frutto dell'intuizione umana OPPURE possono essere rivelate dalle statistiche e quantificate da un modello matematico.

Quando le interazioni fanno parte di un sistema, sono estremamente importanti e causano un comportamento che richiede un modello matematico per essere compreso e gestito.



Apprendimento profondo

Introduzione

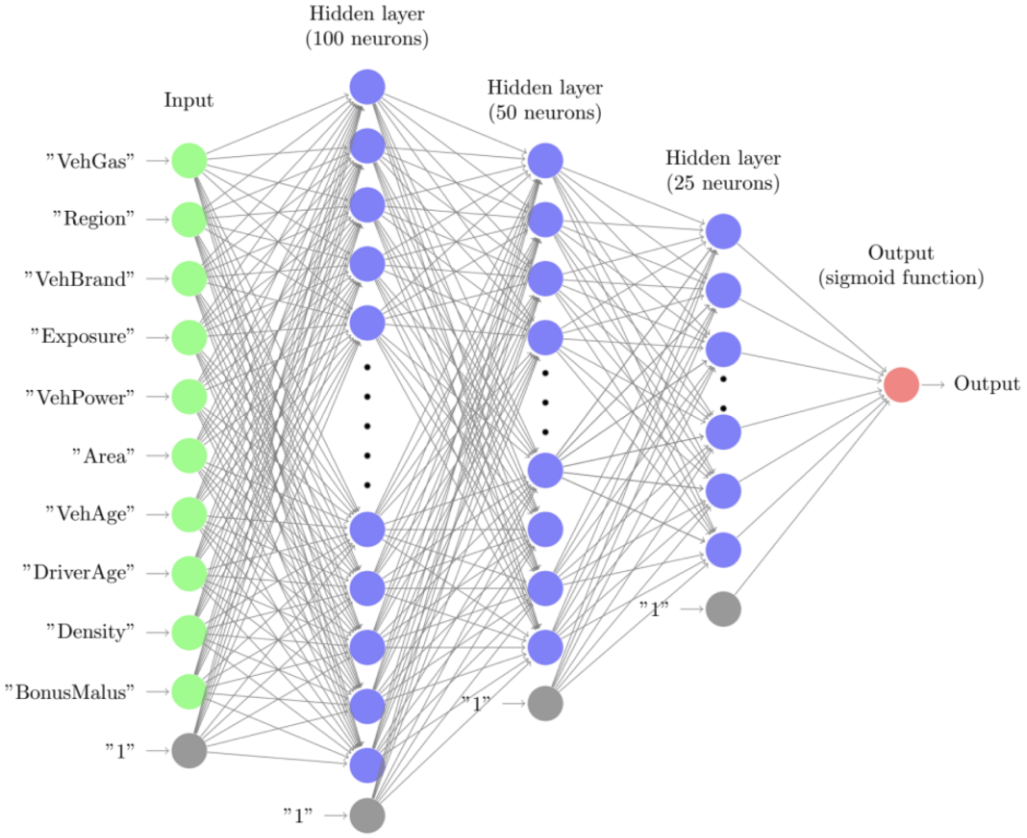
Apprendimento profondo significa reti neurali con molti strati.

Che cos'è una rete neurale?

Si tratta di un tentativo di creare un algoritmo intelligente che imiti le reti biologiche tra i neuroni del cervello.

Come nel cervello, ci sono neuroni e alcune funzioni di attivazione che attivano le sinapsi.

E come nel cervello biologico, alcune connessioni specifiche vengono definite attraverso il processo di apprendimento.



A cosa serve?

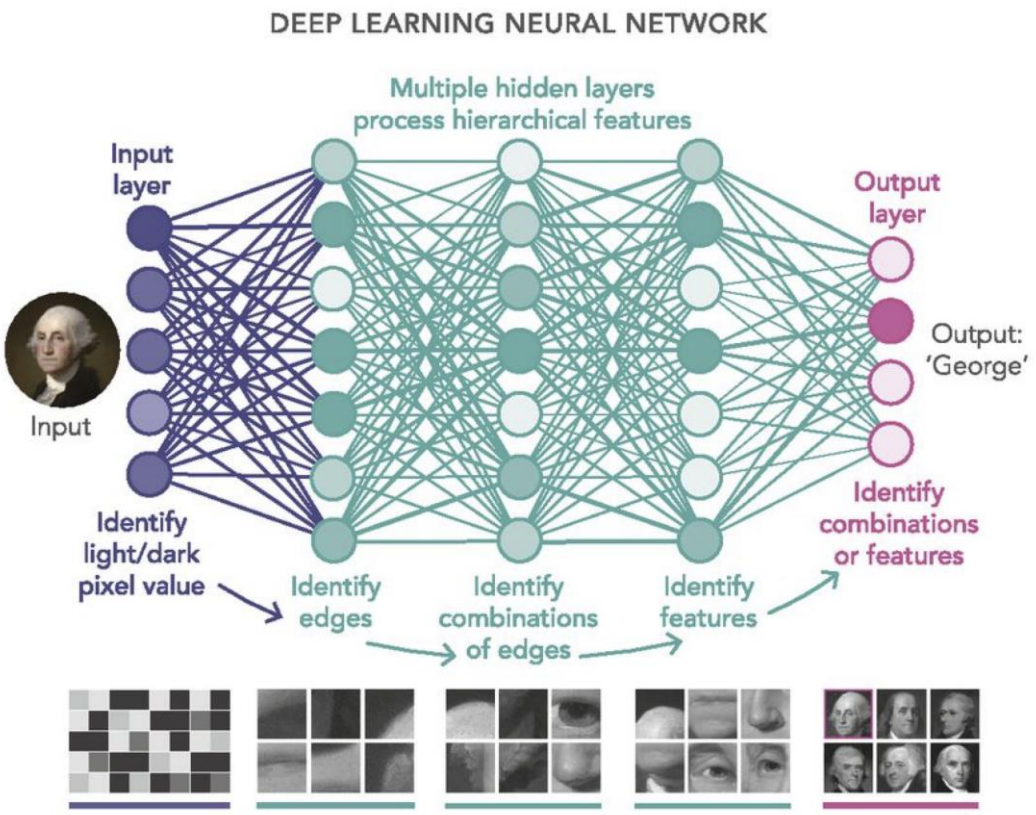
La DL è obbligatoria per alcuni compiti complessi. In particolare, dobbiamo usare la DL per la computer vision e per l'elaborazione del linguaggio naturale.

Per qualsiasi altro compito, è estremamente probabile che si possa usare il ML classico, ma è certo che potenzialmente la DL è molto, molto più potente del ML classico.

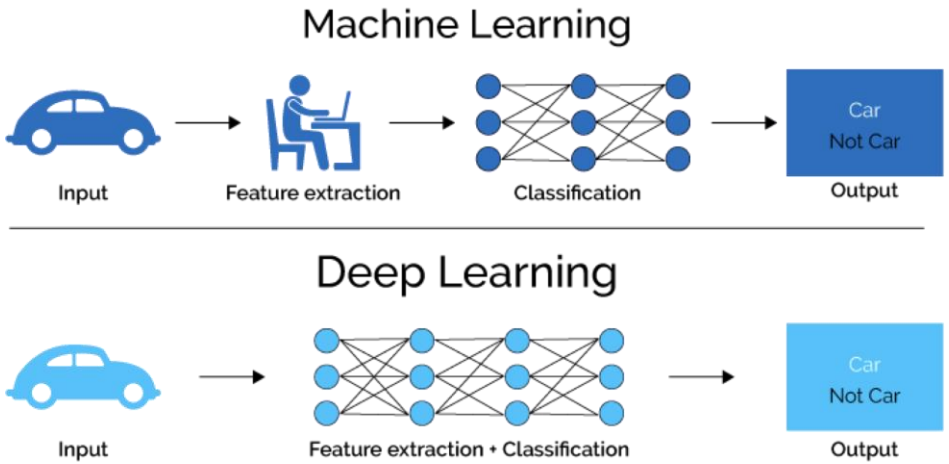
Il costo è che:

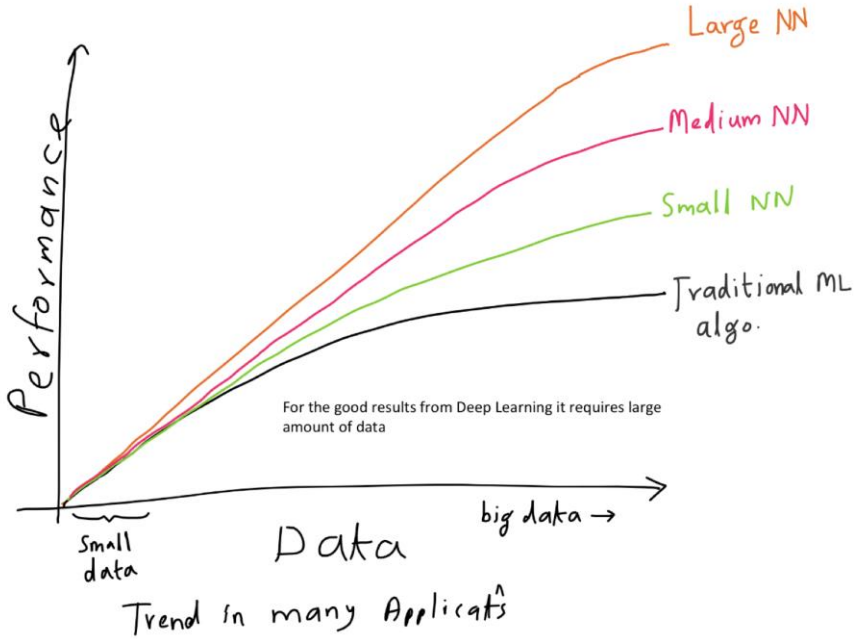
- La formazione richiede molti calcoli
- La DL richiede molti dati

Quindi, se intendete usare il DL, assicuratevi che sia necessario e che possiate farlo.



Come funziona?





Quantum computing

Introduzione

- A 1-bit classical computer can be (or store) in 1 of 2 states at a time: 0 or 1. A 1-qubit quantum computer can be (or store) in 2 states at a time. That is $2^1 = 2$.
- A 2-bit classical computer can store only 1 out of $2^2 = 4$ possible combinations. A 2-qubit quantum computer can store $2^2 = 4$ possible values simultaneously.

Table 2-5. *Qubit Simultaneous Storage Capacity*

Bits/qubits	Classic storage (bytes)	Quantum storage (bits)	Quantum storage (bytes)
4	1	16	2
8	1	256	32
32	4	4294967296	536870912
64	8	1.84467E+19	2.30584E+18

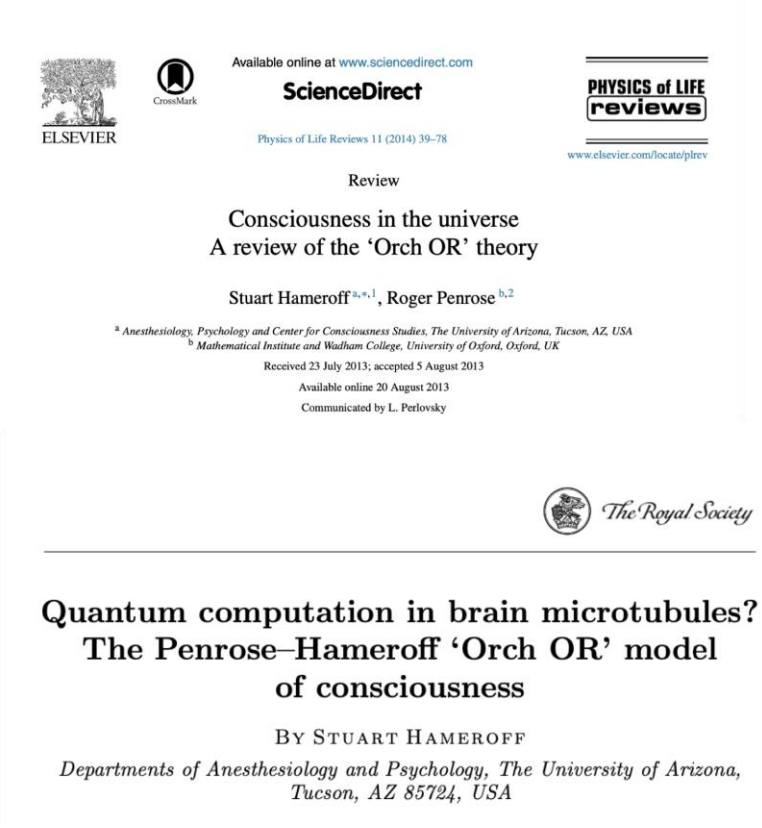
Cervello biologico quantistico?

Alcuni scienziati, tra cui uno dei più importanti fisici, pensano che il cervello biologico possa essere un computer quantistico.

In realtà, il comportamento quantistico dei neuroni biologici, una volta ipotizzato, è stato identificato sperimentalmente.

Quindi, se il dispositivo che ospita l'unica vera intelligenza conosciuta nell'universo è un computer quantistico, questo potrebbe essere il motivo per cui non siamo in grado di creare un'intelligenza artificiale su un computer digitale classico.

Questa ipotesi è assolutamente dirompente, perché implica che l'intelligenza non è una questione di (solo) un algoritmo, ma è molto più un risultato delle leggi fondamentali della Fisica.



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews 11 (2014) 39–78

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Review

Consciousness in the universe
A review of the 'Orch OR' theory

Stuart Hameroff [a,*,1], Roger Penrose [b,2]

[a] Anesthesiology, Psychology and Center for Consciousness Studies, The University of Arizona, Tucson, AZ, USA
[b] Mathematical Institute and Wadham College, University of Oxford, Oxford, UK

Received 23 July 2013; accepted 5 August 2013

Available online 20 August 2013

Communicated by L. Perlovsky

The Royal Society

**Quantum computation in brain microtubules?
The Penrose–Hameroff 'Orch OR' model
of consciousness**

BY STUART HAMEROFF

*Departments of Anesthesiology and Psychology, The University of Arizona,
Tucson, AZ 85724, USA*