

VADEMECUM

PERCORSO FORMATIVO

“Strategia
Tecniche di Intelligenza Artificiale
per entrare nel futuro”.

Big Data e Machine Learning



DataDeep

IL VALORE PROFONDO DEI DATI

www.datadeep.it



StrategIA

Tecniche di Intelligenza Artificiale per entrare nel futuro

Big Data e Machine Learning

A cura di

DataDeep

Il valore profondo dei dati

www.datadeep.it

Indice

Chi sono

Big Data

1. Volume
2. Varietà
3. Velocità
4. Veridicità o Validità
5. Valore

I dati come un terreno fertile da coltivare

- Ma che cos'è un dato?

Machine Learning

- Apprendimento supervisionato
 - Predizione del prezzo di vendita delle case
- Apprendimento non supervisionato
- Apprendimento per rinforzo

Cosa fare adesso?

Un supporto per iniziare: lo studio di fattibilità per la tua azienda senza costi

Chi sono

Mi chiamo Riccardo Manuelli e dal 2007, dall'età di 20 anni, sono un Data Scientist e un imprenditore .

Questo percorso di oltre 17 anni mi ha messo di fronte a sfide diverse, sia di natura operativa che amministrativa, che ho sempre affrontato con un approccio analitico.

La passione per i dati mi ha guidato dallo sviluppo e alla commercializzazione di software per il mondo dell'educational, fino all'approfondimento del Machine Learning nel 2016, avvicinandomi così all'analisi dati e all'intelligenza artificiale.

In risposta a queste passioni e competenze acquisite, nel 2018 ho dato vita a "DataDeep - Il valore profondo dei dati".

Attraverso DataDeep, stiamo aiutando le aziende a integrare una cultura orientata ai dati all'interno della loro struttura organizzativa, fornendo al contempo soluzioni avanzate di intelligenza artificiale. Questi algoritmi sono progettati per supportare i processi decisionali aziendali, permettendo così un avanzamento significativo verso l'efficienza e l'innovazione.

In questi anni, io e il mio team, abbiamo notato una lacuna di conoscenza all'interno delle aziende che spesso si traduce in mancate opportunità di business.

Il nostro obiettivo è colmare questa lacuna, promuovendo una maggiore consapevolezza e capacità di sfruttare queste tecnologie avanzate per il successo aziendale.

Per questo motivo il nostro impegno si è tradotto in svariate attività di divulgazione.

Come formatore, il mio obiettivo è equipaggiare il management con gli strumenti necessari per valorizzare i dati aziendali. Autore e divulgatore, ho anche sviluppato un metodo per guidare le aziende nell'analisi dei dati e nell'integrazione dell'IA, oltre a contribuire con articoli e una newsletter settimanale per diffondere la conoscenza sulla Data Science.

Ho scritto il libro **“L'azienda del futuro. Trasformare le decisioni in azioni di successo con dati e Intelligenza Artificiale”**. (<https://www.amazon.it/dp/B0CWRQLB2R>)

Ho ideato un metodo strutturato che sfrutta strumenti, check-list, modelli e canvas appositamente creati per consentire alle aziende di seguire un percorso ben definito nella gestione di tutte le fasi di progetti di analisi dati e di integrazione dell'Intelligenza Artificiale.

Ho realizzato un **videocorso** dal titolo “**Intelligenza artificiale in azienda: from ZERO to HERO**”. (www.datadeep.it/corso-ai-in-azienda-from-zero-to-hero).

Il corso rappresenta un approfondimento completo che si estende per oltre 15 lezioni, (circa 8 ore) appositamente progettato per imprenditori e manager. Fornisce una visione esaustiva su come sviluppare un approccio metodico e ben strutturato all'analisi dei dati, consentendo di ottenere risultati tangibili senza la necessità di padroneggiare algoritmi complessi.

Per qualsiasi domanda ti rispondo a riccardo.manuelli@datadeep.it

Qui trovi tutti i miei contatti: <https://linktr.ee/riccardo.manuelli>



DataDeep

DataDeep sviluppa progetti di Intelligenza Artificiale e Analisi dati su misura, grazie ad un team di esperti dedicati a fornire supporto alle imprese.

Realizza soluzioni innovative per ottimizzare la supply chain, incrementare l'efficienza e la qualità dei prodotti, affinare la manutenzione dei macchinari, migliorare i margini di profitto, gestire più efficacemente preventivi e di molte altre applicazioni strategiche.

Inoltre, DataDeep si occupa di formazione. Sviluppare una cultura data-driven è vitale per l'efficacia strategica di un'azienda. Questo servizio mira a imprimere una cultura aziendale dove la trasparenza, la decentralizzazione e la responsabilizzazione guidano l'uso dei dati per decisioni informate.

Ulteriormente le Masterclass di intelligenza artificiale equipaggiano i leader aziendali con la comprensione del valore strategico dell'IA, andando oltre le tendenze, con programmi personalizzati per ogni azienda.

Big Data

Quando il concetto di *Big Data* ha cominciato a diffondersi, l'attenzione del pubblico si è inizialmente concentrata sulla grandezza delle informazioni. L'abbondante proliferare di dati in quantità massicce ha naturalmente portato a collegare il concetto di *Big Data* principalmente a questa caratteristica.

Ma per comprendere appieno cosa sono, è essenziale riconoscere che non si tratta solo di una mole imponente di informazioni. Infatti, affinché un insieme di dati sia considerato Big Data, deve soddisfare le cosiddette “5 V”:

1. Volume (*Volume*)
2. Varietà (*Variety*)
3. Velocità (*Velocity*)
4. Veridicità o Validità (*Veracity*)
5. Valore (*Value*)

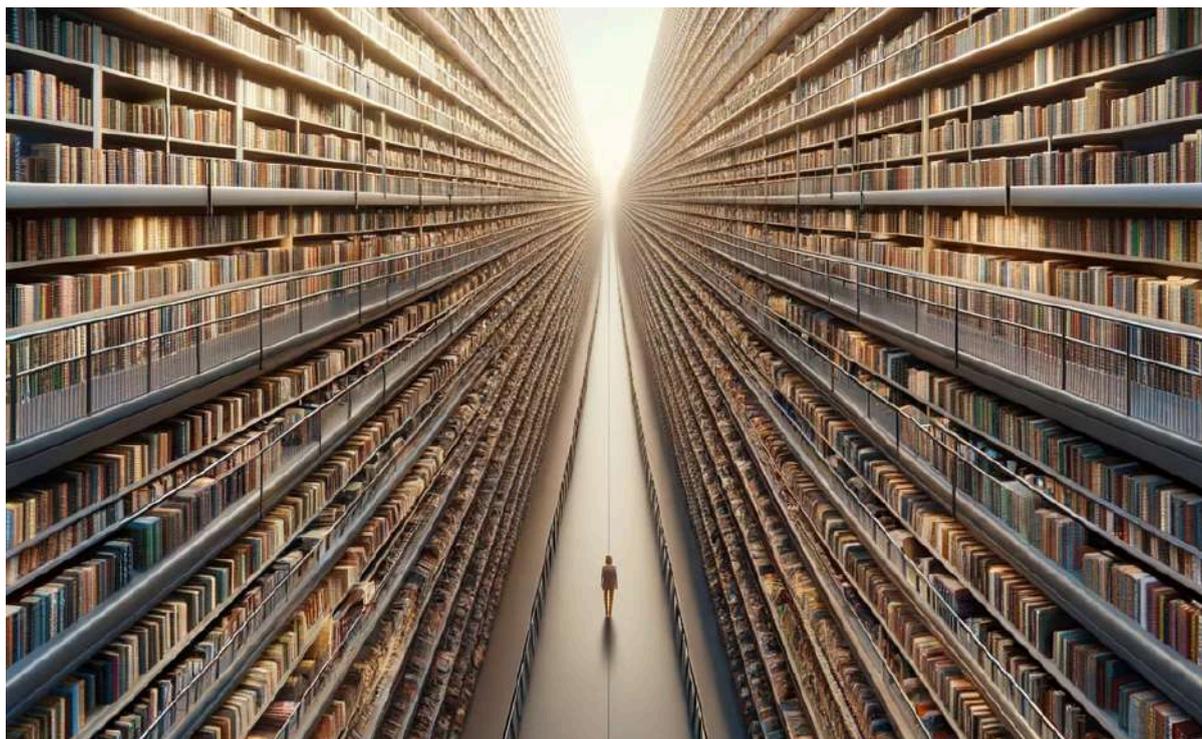
In aggiunta a queste cinque caratteristiche, spesso si considerano altre due caratteristiche:

6. Variabilità (*Variability*)
7. Complessità (*Complexity*)

Non esiste dunque una standardizzazione formale o una definizione ufficiale che stabilisca un numero specifico di caratteristiche per definire i *Big Data*.

Con l'obiettivo di offrire un quadro il più completo possibile, esaminiamo in dettaglio le 5 principali caratteristiche menzionate.

1. Volume



Il concetto di “*Volume*” riguarda la vasta quantità di dati generati e accumulati da molteplici fonti. Immagina il “*Volume*” come un fiume inarrestabile di informazioni provenienti da svariati canali, come sensori, dispositivi mobili, siti web, piattaforme di social media, transazioni finanziarie, registrazioni mediche, e molto altro ancora. Questi dati possono presentarsi in diversi formati, come testo, immagini, video, audio, e dati strutturati provenienti da database.

Quando parliamo di “*vasta quantità di dati*” ci riferiamo a una massa informativa così ampia da superare le capacità dei metodi e degli strumenti tradizionali di gestione. In pratica, la caratteristica del “*Volume*” è soddisfatta quando si accumula un numero di informazioni che supera la consueta capacità dei sistemi e delle procedure per memorizzarle, gestirle ed elaborarle in modo efficiente. Ciò implica la necessità di utilizzare tecnologie specializzate, soluzioni di storage avanzate, e metodi innovativi.

Ad esempio, un’azienda che raccoglie e immagazzina una mole di dati così consistente da richiedere l’utilizzo di infrastrutture di storage scalabili e soluzioni di *cloud computing* per la memorizzazione, la gestione e l’analisi dei dati, si trova di fronte al “*Volume*” dei Big Data.

Potresti facilmente identificare diversi contesti in cui si generano enormi quantità di dati:

- **Social media**

Milioni di persone pubblicano contenuti sui social media ogni giorno. Ogni post, foto, commento o video contribuisce a creare una massa di dati significativa.

Piattaforme come Facebook, X (ex Twitter) e Instagram generano miliardi di post e interazioni giornaliere.

- **Sensori IoT**

Dispositivi interconnessi, come sensori installati su macchinari industriali o dispositivi per la casa intelligente, generano ingenti quantità di dati in tempo reale.

- **E-commerce**

Piattaforme di e-commerce come Amazon o eBay gestiscono grandi volumi di dati relativi alle transazioni, alle preferenze dei clienti, ai comportamenti d'acquisto, e alle interazioni online. Ogni visualizzazione di prodotto, o acquisto, fornisce dati rilevanti, che vengono analizzati per migliorare l'esperienza utente e le strategie di vendita.

Questi esempi evidenziano come il “Volume” sia la prima caratteristica per definire i Big Data, sebbene non sia l'unica, come vedremo in seguito.

2. Varietà



La “*Varietà*” costituisce la seconda caratteristica chiave dei Big Data, concentrandosi sulla diversità dei tipi di informazioni presenti all’interno del loro ambito.

Questa caratteristica comprende dati strutturati, non strutturati e semistrutturati, provenienti da varie fonti e in differenti formati. Indica che i Big Data non si limitano esclusivamente a dati tabellari (*strutturati*), ma abbracciano anche dati non strutturati, come testi, immagini, audio, video, informazioni provenienti da social media, email, sensori, log di server, e altro ancora. I dati non strutturati non seguono uno schema definito, e richiedono approcci e strumenti diversi per l’analisi.

Per una migliore comprensione di queste differenze, esaminiamo più nel dettaglio le tre tipologie:

- **Dati strutturati**

Seguono un modello predefinito, e sono organizzati in tabelle o database. Ad esempio, i record di un database, come quelli relativi a vendite o inventario, rientrano nella categoria dei dati strutturati.

- **Dati non strutturati**

Non seguono una struttura predefinita. Ad esempio, tweet, post sui social media, email, immagini, video, documenti di testo, registrazioni audio, dati provenienti da sensori, rappresentano esempi di dati non strutturati.

- **Dati semistrutturati**

Non sono completamente strutturati (come quelli in un database relazionale), ma contengono un certo grado di struttura o metadati associati. File *XML*, *JSON* o dati provenienti da sensori *IoT*, che incorporano elementi strutturati insieme a informazioni non strutturate, sono esempi di dati semistrutturati.

La “*Varietà*” nei Big Data rende complessa la gestione e l’analisi, poiché richiede l’utilizzo di strumenti e tecniche capaci di trattare dati di varie forme e tipologie. Questa caratteristica pone la sfida di integrare ed elaborare efficacemente informazioni provenienti da molteplici fonti e in diversi formati, al fine di ottenere una visione completa e significativa dei dati.

3. Velocità



La terza caratteristica fondamentale dei Big Data è la “*Velocità*”, che rappresenta la rapidità con cui i dati vengono generati e raccolti, richiedendo analisi e azione in tempi brevi. Com'è noto, alcuni dati devono essere analizzati e sfruttati istantaneamente per ottenere valore. Questa caratteristica è di vitale importanza in contesti in cui le informazioni sono generate velocemente, e l'analisi in tempo reale è fondamentale per prendere decisioni tempestive.

Dispositivi *IoT* e macchine intelligenti generano dati in “*real time*”. Ad esempio, sensori installati su macchinari industriali trasmettono istantaneamente dati sulle prestazioni e condizioni degli strumenti. Una rapida analisi di tali informazioni è fondamentale per individuare anomalie e segnali di malfunzionamento.

Nei mercati finanziari, le negoziazioni avvengono in frazioni di secondo. I dati relativi alle transazioni e alle fluttuazioni dei valori delle azioni richiedono un'analisi immediata per le decisioni di *trading*.

Piattaforme come *Facebook*, *Instagram*, *TikTok*, *LinkedIn* e altre generano costantemente flussi di dati, con migliaia di nuovi post, messaggi e interazioni ogni secondo. L'analisi in

tempo reale di queste informazioni è fondamentale per le strategie di marketing e la gestione della reputazione online.

4. Veridicità o Validità



La quarta caratteristica dei Big Data è la “*Veridicità*”, nota anche come “*Validità*” dei dati, che riflette l’accuratezza, l’autenticità e l’affidabilità delle informazioni raccolte.

Questa caratteristica è cruciale, poiché la qualità dei dati ha un impatto diretto sull’efficacia delle analisi e delle decisioni basate su di essi. Dati imprecisi, incompleti o non affidabili possono portare a conclusioni errate.

Ad esempio, errori umani in fase di inserimento delle informazioni raccolte possono compromettere la veridicità dei dati, fornendo una rappresentazione distorta della realtà. È inoltre fondamentale garantire che i dati non siano stati alterati o manipolati. In contesti come la sicurezza informatica, la verifica della validità delle informazioni diventa cruciale per garantire che non siano state compromesse da attacchi informatici o errori intenzionali.

La provenienza dei dati è altrettanto importante, pertanto fonti autorevoli e affidabili (*come sensori di alta qualità, siti web ufficiali e fonti certificate*) sono indispensabili per garantire la validità delle informazioni disponibili.

5. Valore



La quinta caratteristica chiave riguarda il “*Valore*”, che evidenzia la capacità di estrarre informazioni rilevanti e di generare significativi benefici dai dati raccolti.

Il concetto di “*Valore*” sottolinea l’importanza di non raccogliere dati solamente per accumularli, ma di utilizzare le informazioni per ottenere risultati pratici, prendere decisioni informate, e generare un reale valore per l’organizzazione.

Attraverso l’utilizzo dei Big Data è possibile creare modelli predittivi che anticipano comportamenti futuri o tendenze. Ad esempio, nel settore della vendita al dettaglio l’analisi dei modelli di acquisto passati può aiutare a prevedere la domanda futura di specifici prodotti.

Le aziende che sfruttano i Big Data sono in grado di personalizzare i servizi in base ai

comportamenti e alle preferenze dei clienti. Ad esempio, le piattaforme di streaming usano i dati di visualizzazione per consigliare contenuti basati sugli interessi individuali degli utenti.

L'analisi delle informazioni può portare a un miglioramento dell'efficienza operativa. Per esempio, un'azienda di trasporto può utilizzare i dati sulla gestione delle flotte per ottimizzare le rotte e ridurre i costi operativi.

I dati come un terreno fertile da coltivare

Oggi, i dati vengono spesso descritti come il nuovo oro nero (petrolio), ovvero come una risorsa capace di apportare un valore incommensurabile alle nostre imprese.

In DataDeep, tuttavia, preferiamo vedere i dati come i semi necessari per far fiorire un campo fertile. In questa analogia, ogni particella di suolo rappresenta un singolo dato carico di potenziale. Allo stesso modo in cui un terreno ricco e ben curato può dar vita a un'ampia varietà di vegetazione, nutrendo interi ecosistemi, così i dati, se raccolti, elaborati e impiegati con sapienza, sono in grado di far emergere intuizioni profonde, innovazioni *disruptive* di notevole valore economico.

Nell'immaginario appena descritto, flussi incessanti di informazioni percorrono il paesaggio, alimentati dalle numerose fonti della nostra esistenza digitale: dai social network ai dispositivi interconnessi (Iot o IIoT), passando per i sistemi gestionali aziendali e oltre. Momento dopo momento, questi flussi depositano nuovi sedimenti di dati, arricchendo il terreno di nuove potenzialità.

Ci vediamo come i coltivatori di questi terreni, dove, seminando e curando il suolo, raccogliamo i frutti sotto forma di percezioni, soluzioni a dilemmi complessi e opportunità d'affari inedite.

Eppure, così come un terreno necessita di dedizione e attenzioni per prevenire che infestanti soffochino i raccolti o che l'erosione degradi la terra, il dominio dei dati richiede una gestione oculata. Le questioni relative alla privacy, alla sicurezza e all'integrità dei dati sono paragonabili a pratiche di coltivazione sostenibile per la nostra valle: vitali per assicurare che il suolo rimanga fertile e pronto a supportare sviluppo e innovazione nel lungo periodo.

Pensare ai dati come a un campo fertile ci stimola a considerare non solo il loro valore intrinseco, ma anche la responsabilità che ci incombe nel coltivarli eticamente e in modo sostenibile.

In questa visione, i dati non sono semplicemente una risorsa da sfruttare ma un ecosistema vivente da curare, salvaguardare e valorizzare, cosicché possa prosperare e continuare a offrire i suoi benefici alle future generazioni.

Ma che cos'è un dato?

Su *Wikipedia* possiamo trovare questa definizione: “*Un dato (dal latino “datum”, che significa dono, cosa data) è una descrizione elementare codificata di un’informazione, un’entità, di un fenomeno, di una transazione, di un avvenimento o di altro*”.

Per capire meglio che cos'è un dato e perché lo dobbiamo considerare con un valore, proviamo a metterlo sullo stesso piano altri due concetti: *i beni e i servizi*.

Come i beni e servizi, i dati hanno un valore, e possono essere scambiati, acquistati e venduti. Inoltre, come i beni e i servizi, i dati richiedono investimenti per la loro raccolta, elaborazione, archiviazione e gestione.

Ecco qui di seguito una semplice tabella comparativa:

Beni	Servizi	Dati
Immagazzinabili	Non Immagazzinabili	Immagazzinabili
Limitati	Illimitati	Limitati o Illimitati
Valore unico	Valore multiplo	Valore multiplo

Le aziende possono trarre un significativo vantaggio economico dall'uso strategico dei dati: ottimizzando i processi interni per ridurre costi e migliorare l'efficienza, innovando in

prodotti e servizi per soddisfare meglio le esigenze dei clienti e ottimizzando la produzione, e infine vendendo selettivamente dati e analisi per generare nuove entrate.

Vediamo questi 3 aspetti più in dettaglio:

- **Ottimizzazione dei processi interni**

Utilizzando i dati per analizzare e prevedere le esigenze di manutenzione delle macchine, è possibile ridurre significativamente i costi operativi e migliorare l'efficienza produttiva. L'impiego di modelli predittivi basati su IA consente di pianificare la manutenzione preventiva, minimizzare i tempi di fermo, e identificare problemi di progettazione, portando a una riduzione dei costi e un aumento dell'efficienza.

- **Innovazione in prodotti e servizi**

L'analisi dati permette di adeguare i propri prodotti e servizi alle esigenze dei clienti, migliorando la soddisfazione e la fidelizzazione, nonché di ottimizzare i processi produttivi interni e la gestione della catena di approvvigionamento. Attraverso la Data Science, è possibile identificare nuove opportunità per sviluppare o migliorare offerte, riducendo i costi e incrementando fatturato e profitti grazie a una migliore comprensione dei bisogni dei clienti e alla ottimizzazione della produzione e della qualità dei componenti.

- **Commercio dei dati**

Mentre molte aziende esitano a vendere i propri dati, la monetizzazione delle informazioni raccolte può offrire un flusso di entrate supplementare e un'opportunità per l'auto-finanziamento, purché ciò non comprometta i vantaggi competitivi. La vendita selettiva di analisi e dati può apportare benefici economici senza richiedere nuovi investimenti, contribuendo a distinguere l'azienda nel mercato e a intercettare nuove opportunità commerciali.

Machine Learning

Il Machine Learning è una branca dell'Intelligenza Artificiale. È una disciplina che si basa sull'idea che i computer possano imparare dai dati e diventare sempre più precisi nel prevedere i risultati senza essere esplicitamente programmati.

L'apprendimento automatico, o Machine Learning, si divide in tre grandi categorie:

1. Apprendimento supervisionato
2. Apprendimento non supervisionato
3. Apprendimento per rinforzo

Esistono altre tipologie di apprendimento automatico ma queste sono le principali.

Apprendimento supervisionato

L'apprendimento supervisionato è una delle tecniche più comuni di Machine Learning, ed è utilizzato in diverse situazioni.

È una tecnica in cui il modello di apprendimento viene addestrato usando un insieme di *dati etichettati*, con una corrispondenza nota tra gli *input* e gli *output* che si vogliono ottenere.

Al modello vengono perciò assegnati sia dati iniziali che informazioni sul risultato desiderato, per far sì che il sistema riesca a identificare una regola generale che colleghi i dati in ingresso con quelli in uscita, creando un nesso logico valido per altre attività simili. Proviamo a vedere un esempio di apprendimento supervisionato per capirne meglio il funzionamento.

→ Predizione del prezzo di vendita delle case

Supponiamo di trovarci di fronte alla sfida di stimare il valore di vendita delle abitazioni, utilizzando un insieme di attributi specifici di ciascuna proprietà. Abbiamo raccolto una serie di dati, tra cui la superficie in metri quadri dell'abitazione, il numero di stanze, l'anno di costruzione della proprietà, la sua ubicazione e, naturalmente, il prezzo di vendita

effettivo. Queste informazioni sono organizzate in un formato tabellare, come illustrato nell'esempio qui sotto:

Immobile	Metri quadri	N° stanze	Anno	Zona	Prezzo di vendita (€)
Immobile 1	100	4	1990	A	300.000
Immobile 2	70	3	1980	B	200.000
Immobile 3	120	5	2000	A	350.000
Immobile 4	80	3	2010	C	250.000
Immobile 5	50	2	1970	B	150.000
...

Nella nostra tabella, ciascuna riga corrisponde a una specifica proprietà immobiliare, mentre le colonne - che includono *Metri quadri*, *Numero di stanze*, *Anno di costruzione*, e *Zona* - rappresentano le *caratteristiche* o le *features* dell'immobile. La colonna *Prezzo di vendita* indica il valore in euro a cui l'immobile è stato effettivamente venduto, costituendo la nostra *etichetta* o *target*. Questi dati vengono utilizzati per addestrare un modello di apprendimento supervisionato, il quale impara a prevedere il valore di vendita di un immobile sulla base delle sue caratteristiche.

Attraverso l'analisi dei dati etichettati, il modello stabilisce una relazione tra le qualità di un immobile e il suo prezzo di vendita. Questo modello può poi essere applicato per stimare il valore di vendita di nuove proprietà, in base alle loro specifiche caratteristiche.

Sebbene l'esempio presentato qui includa solo un numero limitato di caratteristiche, nella realtà, numerosi altri fattori possono influenzare in modo significativo il valore di un immobile. Alcuni di questi includono:

- **Condizione dell'immobile.** Una valutazione dello stato generale della proprietà, considerando eventuali rinnovamenti o miglioramenti effettuati.

- **Presenza di giardino o spazio esterno.** Gli spazi esterni come giardini o terrazze possono notevolmente incrementare il valore di una proprietà.
- **Vista.** Immobili con viste particolarmente desiderabili, come quelle panoramiche sul mare, laghi, montagne o punti di interesse, tendono ad avere prezzi di vendita superiori.
- **Piano.** Per gli appartamenti, il piano di ubicazione influisce sul valore, con quelli situati ai piani superiori che spesso godono di migliori viste e minori disturbi da rumore.
- **Efficienza energetica.** Le classi energetiche elevate riflettono una maggiore efficienza nell'uso di energia per il riscaldamento e il raffreddamento, elemento di crescente importanza per gli acquirenti.
- **Numero di bagni.** Un numero maggiore di bagni può incrementare il valore dell'immobile, specialmente in confronto a proprietà con un numero simile di camere da letto.
- **Posto auto/garage.** La disponibilità di parcheggio privato è particolarmente preziosa in aree urbane dense.
- **Prossimità a servizi.** La vicinanza a infrastrutture quali scuole, parchi, negozi e trasporti pubblici può elevare il valore di un immobile.
- **Rumore e inquinamento.** Livelli elevati di rumore o inquinamento possono deprezzare immobili situati in aree trafficate o industriali.
- **Stile architettonico e finiture.** Il design estetico e la qualità delle finiture interne influenzano notevolmente il prezzo di vendita.
- **Esposizione solare e luminosità.** Proprietà luminose e ben esposte al sole sono generalmente più ricercate.
- **Densità abitativa della zona.** Aree con bassa densità offrono più spazio e privacy, attributi che possono incrementare il valore di un immobile.

Integrare una vasta gamma di variabili nel nostro modello di previsione può notevolmente affinare la sua precisione, permettendoci di cogliere con maggiore dettaglio le complessità intrinseche al mercato immobiliare.

Tuttavia, è fondamentale prestare attenzione alla qualità e all'accessibilità dei dati per ogni variabile considerata. Infatti, dati imprecisi o lacunosi possono seriamente minare l'efficacia delle nostre previsioni. Problemi quali l'*overfitting* - ovvero quando un modello, eccessivamente complesso, si adatta troppo scrupolosamente ai dati di addestramento,

imparando persino il "*rumore*" casuale anziché le sole relazioni significative - e l'*underfitting* - che si verifica quando un modello è talmente semplificato da non riuscire a interpretare adeguatamente la struttura dei dati - possono ulteriormente complicare la precisione delle stime fornite dal modello.

È altresì importante rimanere vigili riguardo ai potenziali *bias* presenti nei dati e assicurarsi che le etichette siano classificate correttamente per evitare distorsioni nei risultati.

Apprendimento non supervisionato

L'apprendimento non supervisionato rappresenta una categoria di machine learning che opera in assenza di un'*etichetta* o di un *obiettivo* specifico prefissato per la previsione da parte del modello. In questo contesto, l'obiettivo primario diventa l'individuazione autonoma di schemi, correlazioni o raggruppamenti intrinseci tra i dati analizzati.

Nel proseguire con il nostro esempio sul mercato immobiliare, ci troviamo in una situazione in cui non disponiamo di un prezzo di vendita di riferimento per le abitazioni. Di conseguenza, l'obiettivo della nostra analisi deve necessariamente mutare. Una possibile direzione potrebbe essere quella di scoprire gruppi di proprietà con caratteristiche simili attraverso l'uso di tecniche di clustering. Questi algoritmi ci permettono di aggregare le case in base a vari attributi quali la superficie, la posizione, l'anno di costruzione, e altri ancora. Il risultato di tale processo potrebbe essere l'identificazione di cluster distinti, ad esempio, abitazioni di lusso, case più accessibili economicamente, o immobili che si distinguono per particolari caratteristiche, offrendo così spunti di riflessione preziosi sia per chi vende sia per chi compra.

In questo scenario, forniamo al modello solamente le caratteristiche degli immobili che desideriamo esaminare, lasciando che sia lui stesso a organizzare i dati in gruppi in modo completamente autonomo. Sebbene questo approccio ci privi di una comprensione dettagliata e diretta delle decisioni prese dal modello — un vantaggio tipico dell'apprendimento supervisionato — esso apre le porte a un'esplorazione dati più libera e potenzialmente ricca di scoperte non intuibili a priori.

Adottando l'apprendimento non supervisionato, è importante essere consapevoli delle sfide di questa metodologia, come la determinazione del numero ottimale di cluster o

l'interpretazione dei raggruppamenti ottenuti. Queste complessità richiedono un'attenta valutazione e analisi per garantire che gli insight generati siano effettivamente significativi e utili per le applicazioni pratiche desiderate. Infatti:

- **Non è sempre chiaro il significato dei cluster identificati.** Poiché non viene fornita alcuna etichetta o categoria a priori, potremmo non capire il significato o l'interpretazione dei cluster identificati dal modello.
- **Non è sempre possibile generare cluster significativi.** A volte i dati potrebbero essere troppo complessi o eterogenei per essere divisi in cluster significativi.
- **Dipendenza dall'algoritmo di clustering scelto.** L'efficacia dell'apprendimento non supervisionato dipende molto dall'algoritmo di clustering scelto. Ci sono diverse varianti di algoritmi, e la soluzione sbagliata potrebbe portare a risultati insoddisfacenti.
- **Difficoltà nel valutare i risultati.** Poiché non viene fornita una categoria specifica per i dati, è importante valutare i risultati ottenuti dall'apprendimento non supervisionato per determinarne la validità e l'utilità. Se non comprendiamo appieno come sono stati creati i gruppi, faticiamo a capire ciò che il modello ci sta restituendo.

Apprendimento per rinforzo

La terza tecnica di cui parliamo è l'apprendimento per rinforzo. Questa tecnica è un po' più complessa nella sua forma rispetto alle due precedenti. Qui l'apprendimento automatico compie azioni e riceve *ricompense o punizioni*.

L'obiettivo è quello di trovare una strategia ottimale che massimizzi la ricompensa totale ottenuta nel tempo.

Immaginiamo che la nostra azienda produca bulloni. Possiamo usare un agente di apprendimento per rinforzo per ottimizzare il lavoro dei macchinari. L'agente riceve ricompense per le produzioni di alta qualità, e punizioni per quelle di bassa qualità.

Con il tempo, l'agente capisce cosa porta a una ricompensa e aumenta la sua probabilità di compiere azioni che generano risultati positivi. Il modello impara a partire da queste interazioni, adattando la sua strategia in base alla valutazione della ricompensa ottenuta per ogni azione.

Questo tipo di apprendimento è molto utile per i problemi per cui non esiste una soluzione definita, o per cui le conoscenze sulla relazione tra azioni e ricompense sono incomplete o incerte.

Cosa fare adesso?

Per avviare un processo di analisi dati efficace in un'azienda, è essenziale seguire alcuni passi strategici. Questi passi non solo creano le fondamenta per un ambiente *data-driven*, ma garantiscono anche che il percorso verso l'analisi dei dati sia strutturato e allineato con gli obiettivi aziendali.

1. Definizione del Data Driven Team

La creazione di un team dedicato all'analisi dei dati è il primo passo cruciale. Questo team dovrebbe includere figure con competenze diverse, come data scientists, data engineers, analisti di business e, se possibile, esperti di settore specifici dell'azienda. La definizione di questo team assicura che ci siano le competenze e le risorse necessarie per guidare l'azienda attraverso il processo di trasformazione data-driven, facilitando la raccolta, l'analisi e l'interpretazione dei dati.

2. Sviluppo della Data Governance

Stabilire un solido framework di data governance è fondamentale per garantire che i dati siano gestiti in modo sicuro, etico e conforme alle normative vigenti. Questo include la definizione di politiche per la qualità dei dati, la privacy, la sicurezza e l'accesso ai dati. Una buona governance garantisce che i dati siano affidabili, accurati e disponibili per chi ne ha bisogno, riducendo i rischi legali e di reputazione.

3. Definizione delle Milestone

In questa fase, l'obiettivo principale è quello di definire obiettivi MACRO, ovvero individuare quei processi che consumano eccessive risorse, tempo e richiedono un lavoro manuale smisurato.

È importante adottare un approccio olistico che coinvolga l'intera organizzazione, evitando di perdersi in dettagli minuziosi.

3.1 Definizione degli obiettivi specifici

Gli obiettivi specifici forniscono direzione e scopo all'analisi dei dati. Questi dovrebbero essere SMART (*Specifici, Misurabili, Raggiungibili, Rilevanti, Temporalmente*) per garantire che i risultati siano tangibili e allineati con gli obiettivi aziendali. La definizione chiara degli obiettivi guida il team nell'identificare quali dati raccogliere e analizzare.

3.2 Raccolta e comprensione dei dati

La raccolta di dati pertinenti e la loro completa comprensione sono essenziali per un'analisi efficace. Questo passo include l'identificazione delle fonti di dati interni ed esterni, la valutazione della loro qualità e l'integrazione in un formato analizzabile. Comprendere i dati aiuta a identificare le potenziali insidie e le opportunità nascoste, guidando decisioni basate sui dati più accurate.

3.3 Scelta e valutazione dei modelli

La scelta del modello giusto di machine learning o di analisi statistica è cruciale per interpretare correttamente i dati e trarne insights utili. Valutare vari modelli in base alla loro precisione, efficienza e adattabilità ai dati specifici dell'azienda consente di selezionare l'approccio analitico più efficace per raggiungere gli obiettivi definiti.

3.4 Messa in produzione

L'ultima fase consiste nell'implementare i modelli scelti nel contesto operativo dell'azienda. Questo significa integrare gli *insights* ottenuti dall'analisi dei dati nei processi decisionali quotidiani e nelle operazioni aziendali. La messa in produzione consente di trasformare i dati e gli insights in azioni concrete, misurare l'impatto e continuare a ottimizzare e raffinare i modelli e le strategie basate sui risultati.

Seguire questi passaggi con attenzione e dettaglio può trasformare l'approccio di un'azienda ai dati, da una gestione reattiva a una strategia proattiva e data-driven, sbloccando nuove opportunità e vantaggi competitivi nel lungo termine.

Un supporto per iniziare: lo studio di fattibilità per la tua azienda senza costi

Arrivati a questo punto comprendiamo che gli argomenti trattati sono complessi, per questo motivo alle aziende che vogliono approcciare a queste tecnologie proponiamo uno studio di fattibilità gratuito.

Arrivati a questo punto, siamo consapevoli della complessità degli argomenti trattati, per questo motivo offriamo alle aziende che desiderano approcciarsi a queste tecnologie uno **studio di fattibilità gratuito**.

Lo studio di fattibilità è un'analisi preliminare di una settimana durante la quale esaminiamo i dati e le informazioni fornite dalla tua azienda. Al termine di questo periodo, redigiamo e condividiamo un report dettagliato.

Questo documento offre una panoramica precisa dello stato attuale dei dati aziendali, presentando una vera e propria "fotografia" delle tue risorse informative.

Inoltre, nel report si delineano i possibili percorsi d'azione raccomandati per la tua azienda, orientati verso il raggiungimento degli obiettivi definiti, l'identificazione dei casi d'uso specifici e l'efficace implementazione dell'Intelligenza Artificiale all'interno della tua organizzazione.

Lo scopo principale di questo studio è analizzare lo stato di salute dei dati aziendali e valutare se questi possano sostenere efficacemente il progetto di IA che intendi implementare, in linea con un caso d'uso specifico.

Questa analisi preliminare è cruciale per comprendere se i dati sono sufficienti, rilevanti e di qualità ottimale per l'effettivo funzionamento dell'IA.

Perché lo studio di fattibilità è gratuito?

In DataDeep, siamo fermamente convinti che l'adozione dell'Intelligenza Artificiale (IA) debba essere un processo chiaro, misurato e soprattutto senza rischi. Ciò consente alle aziende di esplorare il potenziale dell'IA senza alcun impegno finanziario iniziale.

Ecco i motivi principali che stanno dietro questa nostra scelta strategica:

Testare senza Rischi

Offrendo uno studio di fattibilità gratuito, permettiamo alla tua azienda di valutare concretamente il nostro operato in un contesto privo di rischi. L'applicazione dell'IA va ben oltre l'utilizzo di strumenti avanzati come Chat GPT. È fondamentale per noi dimostrare in che modo possiamo adattarci alle esigenze specifiche del tuo business, garantendo una solida base per qualsiasi futuro sviluppo.

Focus sul ROI

In DataDeep, il nostro approccio si distingue per la personalizzazione e l'attenzione al ritorno sull'investimento (ROI). La nostra priorità è garantire che ogni soluzione sviluppata porti un valore aggiunto concreto alla tua azienda, assicurando che l'investimento in IA sia non solo tecnologicamente avanzato, ma anche economicamente vantaggioso.

Soluzioni Personalizzate

Attraverso lo studio di fattibilità, acquisiamo una comprensione approfondita delle tue esigenze specifiche e delle sfide che la tua azienda sta affrontando. Questa conoscenza ci permette di sviluppare soluzioni su misura, massimizzando le probabilità di successo del tuo progetto IA.

Promuovere l'Adozione dell'IA

Offrendo valutazioni iniziali senza costi, miriamo a facilitare e incentivare l'adozione dell'IA nelle aziende. Il nostro obiettivo è educare il mercato sulle potenzialità dell'IA e supportare la trasformazione digitale, rendendo questa tecnologia accessibile e comprensibile a un ampio spettro di imprese.

Dimostrazione di Competenza

Questo studio gratuito rappresenta per noi un'opportunità per mostrare le nostre competenze nel campo dell'IA. Ci impegniamo a evidenziare i benefici tangibili che la nostra tecnologia e competenza possono apportare alla tua azienda, dimostrando l'efficacia e la pertinenza delle nostre soluzioni.

Il nostro studio di fattibilità rappresenta il nostro impegno verso un'adozione dell'IA che sia efficace, personalizzata e perfettamente allineata alle esigenze del tuo business.

Questa iniziativa esclusiva ti consente di esplorare senza alcun costo iniziale il potenziale e il valore che l'intelligenza artificiale può apportare alla tua impresa, prima di decidere di intraprendere un percorso di trasformazione digitale.

Vogliamo che tu sia pienamente sicuro e convinto delle opportunità che l'IA può offrire al tuo business, fornendoti tutte le informazioni necessarie per prendere una decisione informata sulla futura integrazione dell'IA nella tua strategia aziendale.

Prenota il primo incontro conoscitivo scrivendo a

ai@datadeep.it

oppure telefonando allo

0163.035014



DataDeep

IL VALORE PROFONDO DEI DATI

Parla con noi

Contattaci per richiedere lo studio di fattibilità GRATUITO per la tua azienda

Telefono: 0163 03 50 14

Email: ai@datadeep.it

Prenota una videocall:

<https://calendly.com/karonsrl/datadeep-raccontaci-le-tue-esigenze>

www.datadeep.it