

Data Deep

IL VALORE PROFONDO DEI DATI

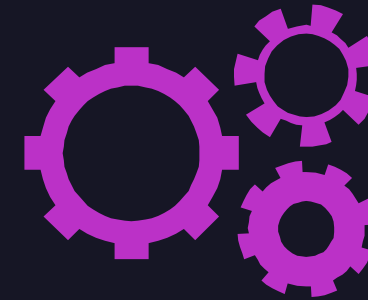
Lezione 10

**Modellizzazione, allenamento e
manutenzione dei modelli**

- Parte 2



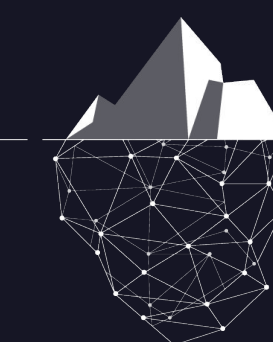
Modellizzazione



Allenamento



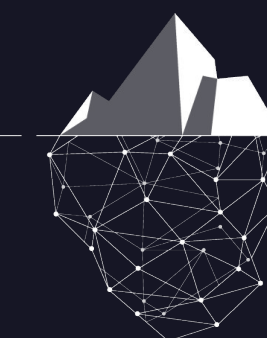
Manutenzione



DataDeep

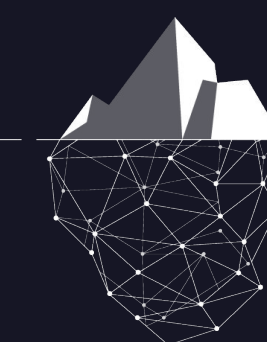
IL VALORE PROFONDO DEI DATI

Allenamento



DataDeep

IL VALORE PROFONDO DEI DATI

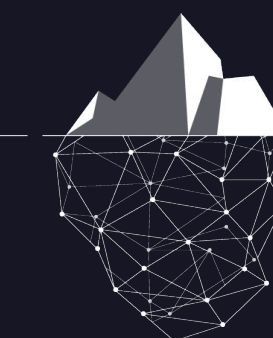


DataDeep

IL VALORE PROFONDO DEI DATI

Codifica delle variabili

ID	Marca	Modello	Anno di produzione	Produzione annua prodotti	Ore di funzionamento	Condizione
1	ABC Manufacturing	M123	2018	10.000	5.000	Buono
2	XYZ Industries	X500	2019	8.000	6.000	Buono
3	LMN Corp	L789	2017	12.000	8.000	Usurato
4	PQR Machines	P200	2020	5.000	2.000	Ottimo
5	ABC Manufacturing	A100	2015	15.000	12.000	Usurato
6	XYZ Industries	B300	2016	20.000	15.000	Mediocre
7	LMN Corp	L789	2018	9.000	5.000	Buono
8	PQR Machines	P200	2019	6.000	3.000	Ottimo
...

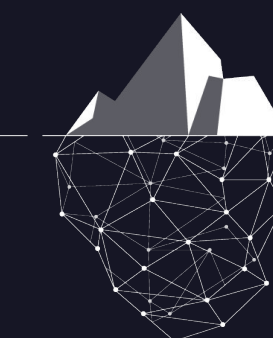


DataDeep

IL VALORE PROFONDO DEI DATI

Condizione
Buono
Buono
Usurato
Ottimo
Usurato
Mediocre
Buono
Ottimo
...

Usurato	Mediocre	Buono	Ottimo
0	0	1	0
0	0	1	0
1	0	0	0
0	0	0	1
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1
...

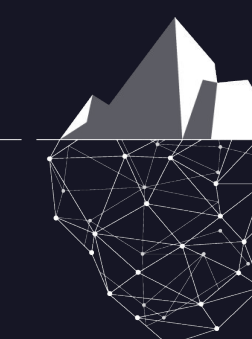


DataDeep

IL VALORE PROFONDO DEI DATI

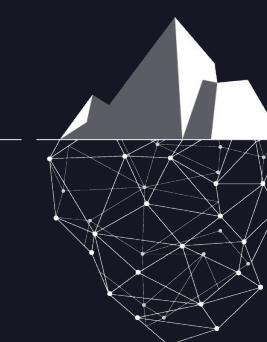
Condizione
Usurato
Mediocre
Buono
Ottimo

Condizione numerica
0
1
2
3



DataDeep

IL VALORE PROFONDO DEI DATI



DataDeep

IL VALORE PROFONDO DEI DATI

Normalizzazione

Standardizzazione

I dati vengono trasformati in modo che abbiano una media di zero e una deviazione standard di uno.

Questo viene fatto sottraendo la media da ogni valore, e dividendo per la deviazione standard dei dati.

La standardizzazione è comunemente utilizzata quando si desidera mantenere la forma e la distribuzione originale dei dati.

Normalizzazione min-max

I dati vengono ridimensionati in un intervallo specifico, solitamente compreso tra 0 e 1.

Ogni valore viene trasformato sottraendo il valore minimo, e dividendo per la differenza tra il valore massimo e il valore minimo.

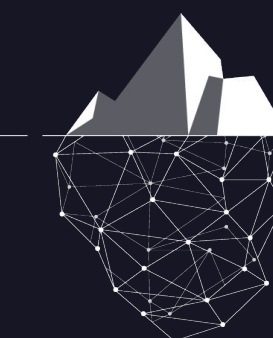
La normalizzazione min-max è utile quando si desidera che i dati siano vincolati a un intervallo specifico.

Normalizzazione L1 o L2

La normalizzazione L1 consiste nel dividere ogni valore per la somma dei valori assoluti di tutti gli elementi del vettore.

La normalizzazione L2 consiste nel dividere ogni valore per la radice quadrata della somma dei quadrati di tutti gli elementi del vettore.

Queste tecniche sono utili quando si desidera dare importanza uguale a tutte le componenti del vettore.



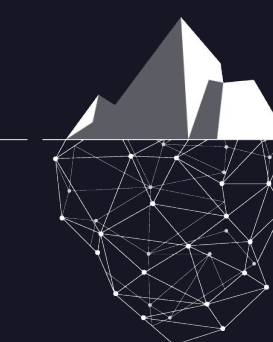
DataDeep

IL VALORE PROFONDO DEI DATI

Standardizzazione

Componente	Diametro (mm)	Peso (g)	Resistenza (MPa)	Temperatura (C°)
Comp_001	35	120	180	160
Comp_002	25	90	210	140
Comp_003	30	110	190	150
Comp_004	28	100	220	155
Comp_005	32	130	200	165

Diametro (mm)	Peso (g)	Resistenza (MPa)	Temperatura (C°)
1.0	1.0	-0.66	1.0
-1.0	-1.0	0.66	-1.0
0.0	0.0	0.33	0.0
-0.4	-0.5	1.0	0.5
0.6	1.5	0.0	1.5



DataDeep

IL VALORE PROFONDO DEI DATI

1

Scalabilità dei dati

Alcuni algoritmi di machine learning, come le reti neurali o gli algoritmi basati su calcoli di distanze, sono sensibili alla scala dei dati. Se le diverse caratteristiche hanno scale molto diverse (ad esempio, il diametro delle parti varia da 20 a 50 mm, mentre il peso varia da 80 a 150 g), gli algoritmi possono dare maggiore importanza a una caratteristica rispetto a un'altra solo a causa delle unità di misura. La standardizzazione aiuta a ridurre questa differenza di scala e a bilanciare l'influenza di ciascuna caratteristica nell'addestramento del modello.

2

Accelerazione della convergenza

Gli algoritmi di ottimizzazione utilizzati per addestrare i modelli spesso convergono più rapidamente quando i dati sono standardizzati. Questo accade perché la standardizzazione mette i dati in una scala comune, riducendo il problema della "disparità di scala" che può rallentare il processo di addestramento. Quando i dati sono standardizzati, il modello può trovare più facilmente il minimo globale o una soluzione ottimale.

3

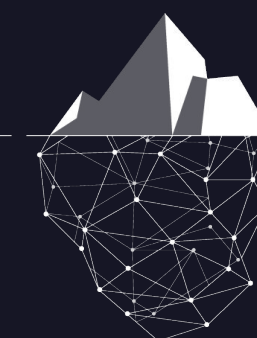
Gestione degli outlier

La standardizzazione può aiutare a gestire la presenza di outlier (valori anomali) nel dataset. Poiché la standardizzazione si basa sulla media e sulla deviazione standard, i valori outlier possono essere "neutralizzati" rispetto alla scala generale dei dati, riducendo così il loro impatto negativo sull'addestramento del modello.

4

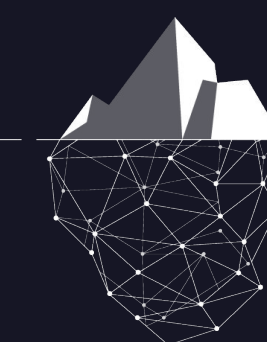
Interpretazione del modello

Quando i dati sono standardizzati, i coefficienti o i pesi del modello possono fornire un'indicazione più diretta dell'importanza relativa delle caratteristiche. Poiché tutte le caratteristiche hanno una scala comune, è più facile valutare il contributo di ciascuna caratteristica al risultato finale del modello.



DataDeep

IL VALORE PROFONDO DEI DATI



DataDeep

IL VALORE PROFONDO DEI DATI

Set di addestramento

Set di dati principale utilizzato per addestrare il modello.

Contiene gli esempi di input, insieme ai corrispondenti output desiderati o etichette.

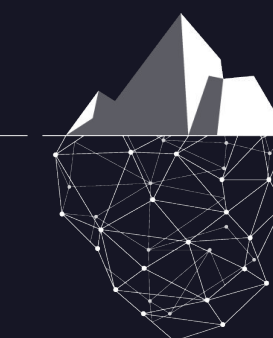
Il modello impara a riconoscere i pattern e le relazioni nei dati di addestramento, cercando di minimizzare l'errore tra le sue predizioni e gli output desiderati.

Set di convalida

Set di dati separato, che serve ad analizzare le prestazioni del modello durante il processo di allenamento.

Il set di convalida viene impiegato per valutare il modello su dati "nuovi", che non sono stati utilizzati per l'addestramento.

Ciò aiuta a stimare l'efficacia del modello nella generalizzazione, e nel fare predizioni accurate su dati sconosciuti.

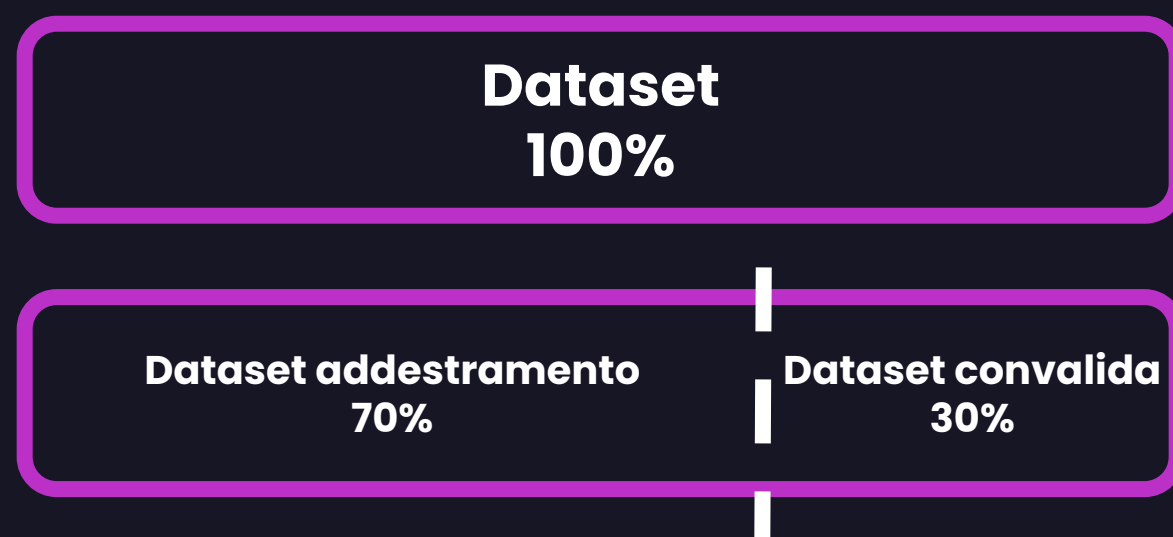


DataDeep

IL VALORE PROFONDO DEI DATI

Suddivisione 70/30 o 80/20

Pratica comune che suddivide i dati in un "set di addestramento" che rappresenta il 70% o l'80% del totale dei dati, e in un "set di convalida" che rappresenta il 30% o il 20% rimanente.



K-Fold cross validation

Tecnica utilizzata con dati limitati. I dati vengono suddivisi in K parti uguali, chiamate "fold". Il modello viene allenato K volte, utilizzando ogni volta un "fold" diverso come "set di convalida", e gli altri "fold" come "set di addestramento".

Set di dati				
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5

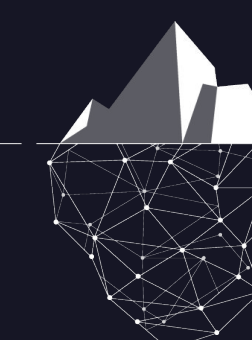
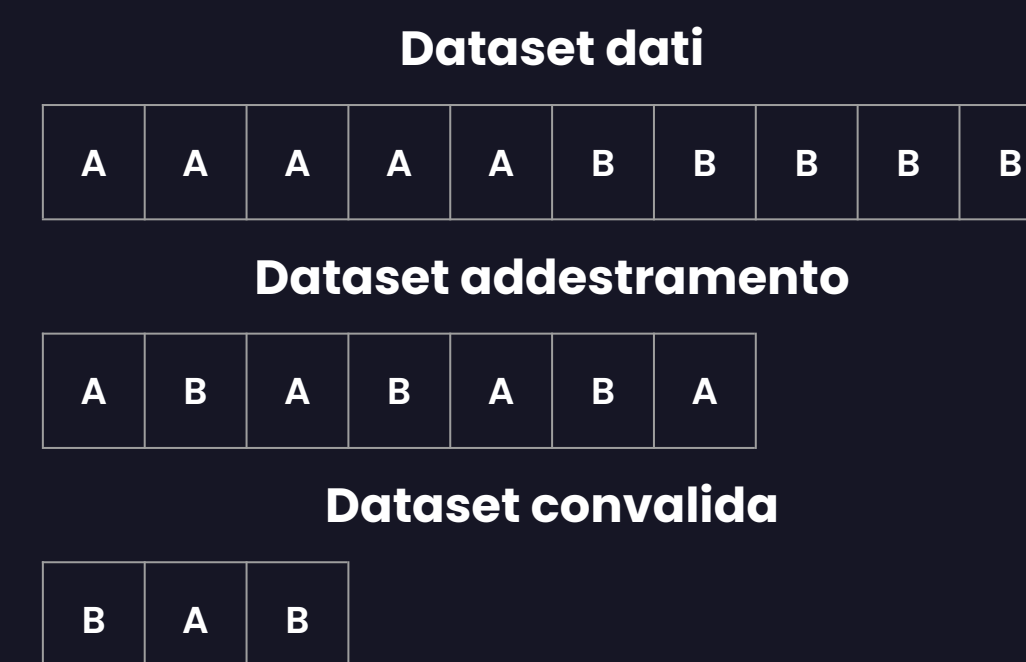
Stratificazione

Se il dataset ha una distribuzione sbilanciata di classi o etichette di output, è importante garantire che la suddivisione tra addestramento e convalida preservi questa distribuzione.

Dataset iniziale 15 esempi	Classe A 9	Classe B 6
Dataset addestramento 10 esempi	Classe A 7	Classe B 3
Dataset convalida 5 esempi	Classe A 2	Classe B 3

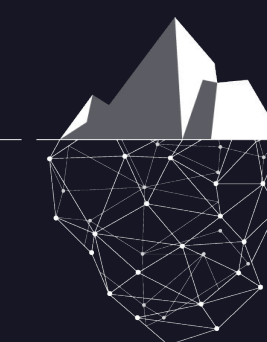
Randomizzazione

La randomizzazione aiuta a evitare qualsiasi effetto di ordine nei dati, e assicura una rappresentazione casuale delle osservazioni nel "set di addestramento" e nel "set di convalida".



DataDeep

IL VALORE PROFONDO DEI DATI



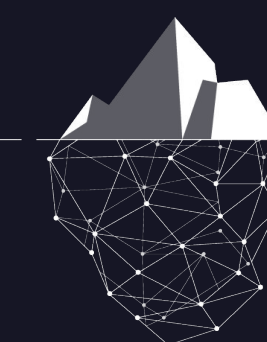
DataDeep

IL VALORE PROFONDO DEI DATI



DataDeep

IL VALORE PROFONDO DEI DATI



DataDeep

IL VALORE PROFONDO DEI DATI

Mantenimento dei modelli

1

Monitoraggio delle prestazioni

È fondamentale monitorare costantemente le prestazioni dei modelli nel tempo per assicurarsi che continuino a funzionare correttamente. Ciò implica l'analisi delle metriche di valutazione del modello. Se si osservano scostamenti significativi o deterioramenti delle prestazioni, possono essere necessarie azioni correttive.

2

Valutazione nuove situazioni

Nel corso del tempo possono verificarsi nuove situazioni o condizioni che i modelli non hanno affrontato durante l'allenamento iniziale. È importante valutare come i modelli si comportano in queste nuove circostanze e prendere provvedimenti per adattarli, se necessario. Ciò può richiedere l'acquisizione di nuovi dati per allargare il set di addestramento, o l'aggiornamento dell'architettura del modello.

3

Aggiornamento dei modelli

I modelli possono beneficiare degli aggiornamenti periodici per rimanere allineati alle nuove informazioni e alle esigenze aziendali. Ciò può comportare la raccolta di dati aggiuntivi, l'espansione del set di addestramento, l'ottimizzazione dei parametri, o l'introduzione di nuove tecniche di modellazione. L'aggiornamento dei modelli dovrebbe essere fatto con attenzione per evitare l'introduzione di nuovi problemi o bias indesiderati.

4

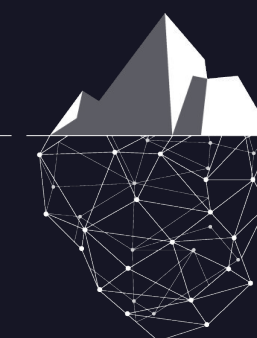
Gestione dei dati

La qualità e la rilevanza dei dati utilizzati per addestrare e mantenere i modelli sono fondamentali. È necessario assicurarsi che le informazioni siano accurate, complete e rappresentative della situazione. Inoltre, può essere necessario gestire eventuali cambiamenti nei sistemi di acquisizione dati o nelle fonti di dati per mantenere la coerenza e l'integrità del processo di apprendimento automatico.

5

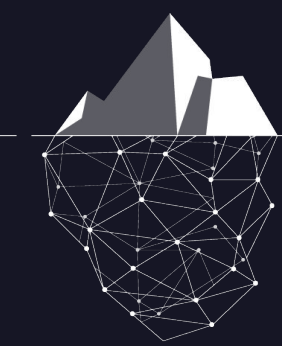
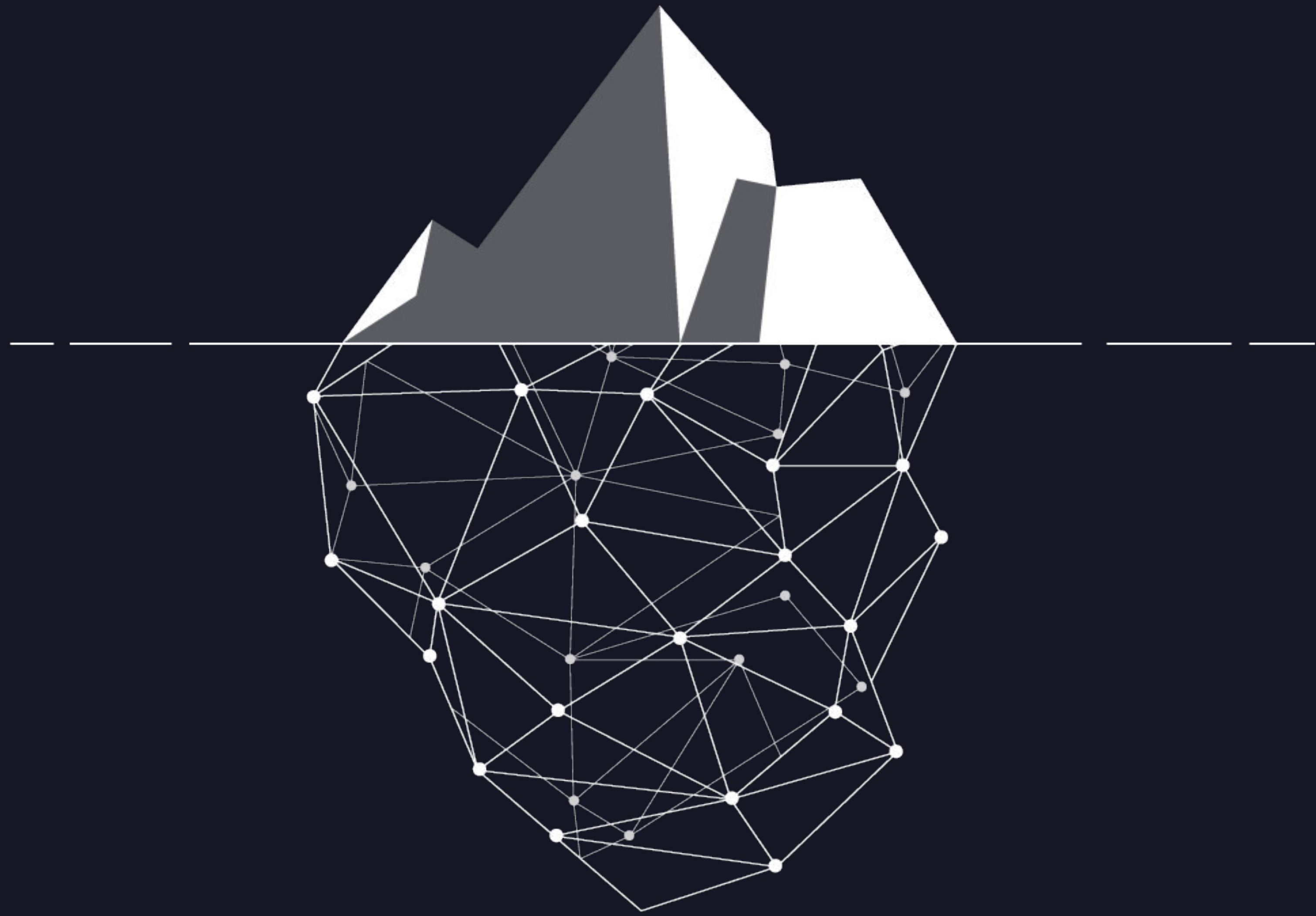
Collaborazione tra team

Il mantenimento dei modelli di apprendimento automatico richiede una collaborazione continua tra team tecnici, Data Scientist, esperti di dominio e stakeholder aziendali. La condivisione delle conoscenze, il feedback e la collaborazione continua sono fondamentali per identificare e affrontare le sfide emergenti, migliorare i modelli, e ottenere risultati ottimali nel tempo.



DataDeep

IL VALORE PROFONDO DEI DATI



DataDeep

IL VALORE PROFONDO DEI DATI

Vuoi seguire tutte le lezioni in un'unica sessione?

Possiamo organizzare la formazione dedicata a te e ai tuoi collaboratori presso la tua azienda.

Durante una sessione di mezza giornata, forniremo un quadro completo su come sviluppare un percorso di analisi dati metodico e strutturato, permettendoti di ottenere risultati concreti senza dover padroneggiare lo sviluppo di algoritmi complessi.

Inoltre, avrai la possibilità di partecipare ad un confronto diretto con noi, nel quale analizzeremo la tua realtà, i tuoi obiettivi e i risultati di business che desideri raggiungere.

Contattaci per informazioni



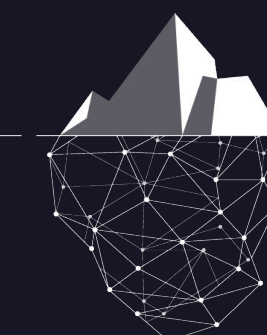
ai@datadeep.it



0163 03 50 14



datadeep.it



DataDeep

IL VALORE PROFONDO DEI DATI