

Lista di controllo per la gestione dei dati

Benvenuti nell'era dell'intelligenza artificiale (AI), in cui l'operatività dell'azienda dipende dalle tecnologie ad alta intensità di dati, quali machine learning e deep learning. Per utilizzare al meglio questi nuovi strumenti di intelligenza artificiale, è necessario assicurarsi che il luogo in cui risiedono i dati dell'organizzazione sia in ordine.

Ecco una lista di controllo per iniziare il percorso che consentirà di mettere ordine nei dati, suddiviso in due fasi fondamentali: preparazione e inferenza.

Seguire questi passaggi per diventare esperti dell'AI. Per ulteriori informazioni su come portare l'intelligenza artificiale dal "proof of concept" alla piena produzione e all'applicazione su vasta scala, dai un'occhiata a questo rapporto IDC, [*Accelerate and Operationalize AI Deployments Using AI-Optimized Infrastructure*](#).

Preparazione

Nella fase di preparazione per l'AI, si svilupperanno algoritmi atti a comprendere un set di dati. Il processo consisterà nell'acquisire i dati esistenti e utilizzare l'AI per apprendere una nuova funzionalità.

- Definire il problema aziendale specifico che si vorrebbe risolvere usando l'AI (iniziando con progetti piccoli per imparare)
- Individuare i dati in grado di risolvere il problema attingendo alle fonti pertinenti (molto probabilmente non si troveranno tutti in un unico posto)
- Preparare i dati con i tag dei metadati per ridurre significativamente il tempo necessario per trovare i dati pertinenti
- Assicurarsi che i dati siano sincronizzati correttamente e collegarli a tutti i set di dati che saranno utilizzati (e che siano sincronizzati anche cronologicamente)
- Segnalare tutti i dati sensibili dei clienti e altri dati riservati affinché sia garantita la massima sicurezza e l'osservanza di tutte le normative e le disposizioni vigenti (può essere utile al riguardo il processo di codifica dei metadati)
- Scegliere il giusto ambiente di sviluppo per il tipo di dati utilizzati e il modo in cui saranno formattati (ad es. immagini, video, testo in formato libero e audio in genere hanno ciascuno un tipo di ambiente)
- Estrarre i set di dati dall'archivio e portarli nell'ambiente di sviluppo
- Dividere i dati in due gruppi per migliorare il processo di sviluppo del modello (mantenere un set in una cartella denominata "train" e un altro in una cartella chiamata "test")
- Mantenere la tracciabilità dei dati tenendo traccia della provenienza/ dell'origine dei dati (considerare l'utilizzo di strumenti che possono aiutare ad automatizzare il processo)
- Eseguire attività elementari di pulizia dei dati per prepararli per la creazione di un modello (ad esempio, inserendo elementi di dati mancanti e rimuovendo elementi non validi)
- Utilizzare un subset di dati per i quali si conosce già la risposta all'attività di previsione (questo si chiama "set di preparazione") e identificare tutte le fasi di pre-elaborazione necessarie per preparare i dati affinché siano in grado di formulare una previsione
- Usare le conoscenze acquisite con questo set di preparazione per calcolare punteggi di precisione in modo da essere sicuri di poter applicare lo stesso modello a nuovi dati per i quali il modello non è mai stato preparato espressamente

Inferenza

Dopo aver sviluppato un modello che funziona per risolvere il problema aziendale, si passerà dalla preparazione all'inferenza. In questa fase, si prenderà il modello funzionante preparato in precedenza e lo si applicherà a nuovi dati; anche questa fase richiederà un'operazione di pulizia dei dati.

- Individuare il modello di AI adatto ai dati da gestire per ridurre la latenza, ridurre i requisiti di larghezza di banda e migliorare le prestazioni complessive del modello
- Sviluppare un processo efficiente per la pipeline dei dati e applicare le etichette ai dati dei metadati man mano che arrivano, in modo che i nuovi dati possano essere acquisiti e utilizzati per migliorare il modello in futuro
- Contrassegnare i dati in modo che siano collegati e sincronizzati (ad esempio, se i dati sono in sequenza temporale, è possibile sincronizzare set di dati o creare dei link selezionando un campo, ad esempio il nome del cliente, in tutti i dati che arrivano)
- Sviluppare un piano di storage del ciclo di vita dei dati a lungo termine per come si gestiranno il volume e la velocità dei dati man mano che arrivano e man mano che vengono archiviati
- Considerare la possibilità di assumere un Chief Data Officer che si occupi della gestione dei dati dell'organizzazione per futuri progetti di AI, deep learning e altri progetti relativi ai dati