

Whitepaper AWS

Principi della previsione di serie temporali con Amazon Forecast



Principi della previsione di serie temporali con Amazon Forecast:

Whitepaper AWS

Copyright © 2023 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

I marchi e il trade dress di Amazon non possono essere utilizzati in relazione a prodotti o servizi che non siano di Amazon, in qualsiasi modo che possa causare confusione tra i clienti o in qualsiasi modo che denigri o discrediti Amazon. Tutti gli altri marchi non di proprietà di Amazon sono di proprietà dei rispettivi proprietari, che possono o meno essere affiliati, collegati o sponsorizzati da Amazon.

Table of Contents

Riassunto e panoramica	i
Panoramica	1
Hai già provato Well-Architected?	2
Informazioni sulla previsione	3
Sistema di previsione	3
Quali settori riscontrano problemi di previsione?	3
Aspetti da considerare prima di provare a risolvere un problema di previsione	4
Caso di studio: problema di previsione della domanda al dettaglio per un'azienda di e-commerce	6
Fase 1: raccolta e aggregazione dei dati	9
Esempio	11
Fase 2: preparazione dei dati	12
Come gestire i dati mancanti	12
Esempio 1	12
Esempio 2	15
Concetti relativi alla caratterizzazione e alle serie temporali correlate	15
Esempio 3	16
Fase 3: creazione di un predittore	18
Fase 4: valutazione dei predittori	20
Backtesting	20
Quantili di previsione e metriche di precisione	22
Perdita quantilica ponderata (wQI)	22
Errore percentuale assoluto ponderato (WAPE)	23
Errore quadratico medio (RMSE)	23
Problemi con WAPE e RMSE	24
Fase 5: generazione e uso di previsioni per il processo decisionale	26
Previsioni probabilistiche	26
Visualizzazione	27
Riepilogo del flusso di lavoro di previsione e delle API	29
Uso di Amazon Forecast per scenari comuni	30
Implementazione di Forecast nell'ambiente di produzione	31
Conclusione	32
Collaboratori	33
Approfondimenti	34

Appendice A: Domande frequenti 35

Appendice B: Riferimenti 39

Cronologia documenti 40

Avvisi 41

Glossario di AWS 42

Principi della previsione di serie temporali con Amazon Forecast

Data di pubblicazione: 1 settembre 2021 ([Cronologia documenti](#))

Oggi le aziende usano ogni strumento possibile, da semplici fogli di calcolo a complessi software di pianificazione finanziaria, per tentare di prevedere con precisione i risultati aziendali futuri, tra cui la domanda di prodotti, le esigenze in termini di risorse o le prestazioni finanziarie. Questo documento presenta le previsioni, insieme alla terminologia, alle sfide e ai casi d'uso correlati. Questo documento usa un caso di studio per avvalorare i concetti correlati alle previsioni e presentare le fasi del processo e informazioni su come [Amazon Forecast](#) permette di affrontare molte sfide pratiche in problemi di previsione reali.

Panoramica

La previsione è la scienza di prevedere il futuro. Usando dati cronologici, le aziende possono identificare le tendenze, stabilire che cosa può succedere e quando e quindi integrare queste informazioni nei piani futuri per qualsiasi scenario, dalla domanda di prodotti alla pianificazione dell'inventario fino alla gestione del personale.

A causa delle conseguenze delle previsioni, la precisione è un aspetto importante. Se una previsione è sovrastimata, i clienti possono investire eccessivamente in prodotti e personale, con il conseguente spreco di investimenti. Se la previsione è sottostimata, i clienti possono investire in modo insufficiente, producendo una carenza di materie prime e scorte e creando un'esperienza insoddisfacente per il cliente.

Oggi le aziende provano a usare qualsiasi strumento, da semplici fogli di calcolo a complessi software di pianificazione finanziaria o della domanda, per generare previsioni, ma la precisione elevata resta un aspetto problematico per due motivi:

- Innanzitutto, le previsioni tradizionali hanno difficoltà a includere grandi volumi di dati cronologici, perdendo importanti segnali dal passato.
- In secondo luogo, le previsioni tradizionali raramente includono dati correlati ma indipendenti, che possono offrire un contesto importante (come prezzi, festività ed eventi, esaurimento delle scorte, promozioni di marketing e così via). Senza la cronologia completa e un contesto più ampio, la maggior parte delle previsioni non riesce a prevedere il futuro con precisione.

[Amazon Forecast](#) è un servizio completamente gestito che permette di risolvere questi problemi. Amazon Forecast fornisce i migliori algoritmi per ogni scenario di previsione da gestire. Si affida a moderne tecniche di machine learning e deep learning in base alle esigenze per fornire previsioni ad alta precisione. Amazon Forecast è facile da usare e non richiede alcuna competenza di machine learning. Il servizio fornisce automaticamente l'infrastruttura necessaria, elabora i dati e crea modelli di machine learning personalizzati/privati ospitati in AWS e pronti per generare previsioni. Inoltre, man mano che le tecniche di machine learning continuano a evolversi rapidamente, vengono integrate in Amazon Forecast, in modo che i clienti possano continuare a osservare miglioramenti in termini di precisione, con attività minime o nulle da svolgere.

Hai già provato Well-Architected?

Il [Framework AWS Well-Architected](#) ti aiuta a identificare i pro e i contro delle tue decisioni durante la creazione di sistemi nel cloud. I sei principi del framework permettono di apprendere best practice architetturali per la progettazione e l'esecuzione di sistemi affidabili, sicuri, efficienti, convenienti e sostenibili. Usando lo [AWS Well-Architected Tool](#), disponibile gratuitamente nella [AWS Management Console](#), puoi esaminare i carichi di lavoro rispetto alla best practice rispondendo a una serie di domande per ogni principio.

Nella pagina di [approfondimento sul machine learning](#) ci concentriamo su come progettare, implementare e costruire i carichi di lavoro di machine learning nel Cloud AWS. Questo approfondimento si aggiunge alle best practice incluse nel Framework Well-Architected.

Per ulteriori indicazioni specializzate e le best practice per la tua architettura cloud, inclusi diagrammi, implementazioni e whitepaper, consulta il [Centro di architettura AWS](#).

Informazioni sulla previsione

In questo documento, previsione significa prevedere i valori futuri di una serie temporale: l'input o l'output di un problema è per natura costituito da serie temporali.

Sistema di previsione

Un sistema di previsione include una serie diversificata di utenti:

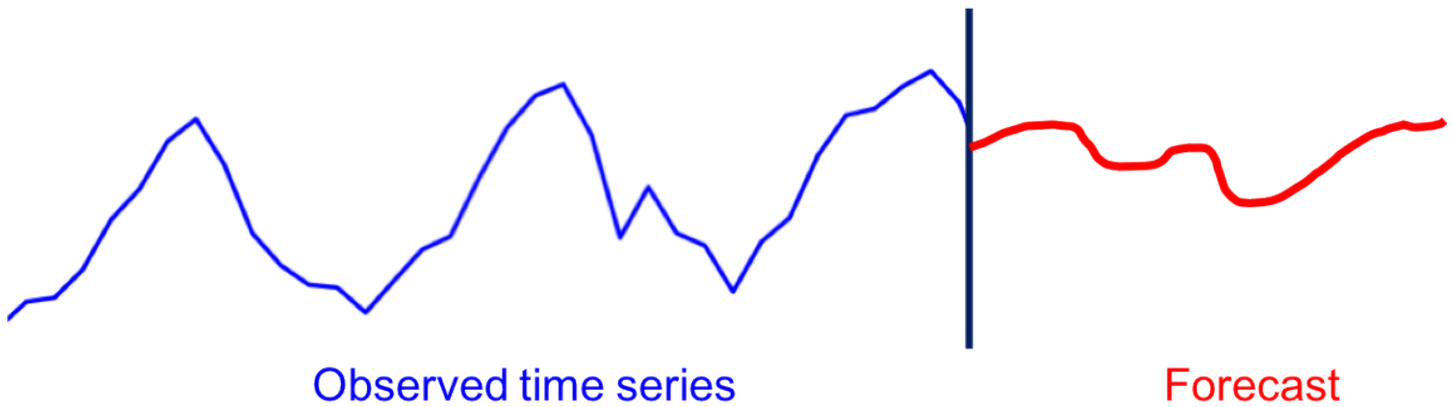
- Utenti finali, che eseguono query sul sistema di previsione relativamente a un prodotto specifico e decidono quante unità acquistare. Può trattarsi di una persona o di un sistema automatizzato.
- Analisti aziendali/business intelligence, che supportano gli utenti finali ed eseguono e organizzano report aggregati.
- Data scientist, che analizzano in modo iterativo i modelli della domanda e gli effetti causali e aggiungono nuove caratteristiche per fornire miglioramenti incrementali al modello o ottimizzare il modello di previsione.
- Data engineer, che configurano l'infrastruttura di raccolta dei dati e garantiscono la disponibilità dei dati di input per il sistema.

Amazon Forecast alleggerisce il lavoro dei software engineer e permette ad aziende con competenze limitate in data science di trarre il massimo vantaggio da una tecnologia di previsione all'avanguardia. Per le aziende con competenze in data science, è inclusa una serie di funzionalità di diagnostica per affrontare con efficienza i problemi di previsione con Amazon Forecast.

Quali settori riscontrano problemi di previsione?

I problemi di previsione toccano molte aree che per definizione producono dati di serie temporali. Questi includono vendita al dettaglio, analisi medica, pianificazione della capacità, monitoraggio di reti di sensori, analisi finanziaria, raccolta di dati su attività di social media e sistemi di database. Ad esempio, le previsioni svolgono un ruolo chiave nell'automazione e nell'ottimizzazione dei processi operativi nella maggior parte delle aziende che implementano processi decisionali basati sui dati. Le previsioni per la domanda e l'approvvigionamento di prodotti possono essere usate per attività ottimali di gestione dell'inventario, programmazione del personale e pianificazione della topologia e sono più in generale una tecnologia essenziale per la maggior parte degli aspetti legati all'ottimizzazione della catena di approvvigionamento.

La figura seguente contiene un riepilogo di un problema di previsione osservato e basato su una serie temporale che presenta un modello (in questo caso, la stagionalità) e la creazione di una previsione per un periodo specificato. L'asse orizzontale rappresenta il tempo che va dal passato (a sinistra) al futuro (a destra). L'asse verticale rappresenta le unità misurate. Considerando il passato (in blu) fino alla linea verticale nera, l'identificazione del futuro (in rosso) corrisponde all'attività di previsione.



Panoramica dell'attività di previsione

Aspetti da considerare prima di provare a risolvere un problema di previsione

Le domande più importanti a cui rispondere prima di risolvere problemi di previsione sono le seguenti:

- Devi risolvere un problema di previsione?
- Perché vuoi risolvere il problema di previsione?

A causa dell'ubiquità dei dati delle serie temporali, è facile trovare problemi di previsione ovunque. Tuttavia, una domanda importante è se vi sia davvero l'esigenza di risolvere un problema di previsione o se sia possibile eluderlo completamente senza sacrificare l'efficienza del processo decisionale dell'azienda. La risposta a questa domanda è importante perché, in termini scientifici, la previsione pone alcuni dei problemi più complessi nel machine learning.

Ad esempio, pensiamo ai suggerimenti sui prodotti per un rivenditore online. Questo problema può essere inquadrato come problema di previsione in cui, per ogni coppia cliente-SKU (codice di riferimento del prodotto), è necessario prevedere il numero di unità di un articolo specifico che un determinato cliente potrebbe acquistare. Questa formulazione del problema presenta diversi

vantaggi. Un vantaggio è che tiene esplicitamente conto della componente temporale e permette quindi di suggerire prodotti in base ai modelli di acquisto dei clienti.

Tuttavia, i problemi correlati ai suggerimenti sui prodotti raramente vengono formulati come problemi di previsione, in quanto la risoluzione di un problema di questo tipo è molto più difficile, ad esempio per l'esiguità di informazioni a livello di cliente-SKU e l'entità del problema, di quanto non sia affrontare direttamente il problema originale. Di conseguenza, nel riflettere sull'applicazione di una previsione, è importante tenere conto dell'utilizzo a valle della previsione e chiedersi se non sia invece possibile risolvere il problema usando un approccio alternativo.

[Amazon Personalize](#) può essere utile in questi casi. Amazon Personalize è un servizio di machine learning che permette agli sviluppatori di creare facilmente suggerimenti individuali per gli utenti delle loro applicazioni.

Dopo aver stabilito che è necessario risolvere un problema di previsione, la domanda successiva a cui rispondere è perché vogliamo risolverlo. In molti contesti aziendali la previsione è in genere solo un mezzo per raggiungere un fine. Ad esempio, per la previsione della domanda in un contesto di vendita al dettaglio, la previsione può essere usata per prendere decisioni sulla gestione dell'inventario. Il problema di previsione è in genere l'input per un problema decisionale, che a sua volta può essere ricondotto a un problema di ottimizzazione.

Alcuni esempi di tali problemi decisionali includono il numero di unità da acquistare o il miglior approccio per gestire l'inventario esistente. Altri problemi di previsione aziendale includono la previsione della capacità dei server o della domanda di materie prime/pezzi in un contesto di produzione. Queste previsioni possono essere usate come input per altri processi, sia per i problemi decisionali presentati sopra, sia per simulazioni di scenari, da usare quindi per la pianificazione senza modelli espliciti. Esistono alcune eccezioni alla regola secondo cui una previsione non è in se stessa un fine. Nelle previsioni finanziarie, ad esempio, la previsione viene usata direttamente per consolidare riserve finanziarie o viene presentata agli investitori.

Per determinare lo scopo della previsione, considera le domande seguenti:

- Quanto in là nel futuro deve arrivare la previsione?
- Con che frequenza devono essere generate le previsioni?
- Vi sono aspetti specifici da approfondire riguardo alle previsioni?

Caso di studio: problema di previsione della domanda al dettaglio per un'azienda di e-commerce

Per illustrare più dettagliatamente i concetti correlati alle previsioni, consideriamo il caso di un'azienda di e-commerce che vende prodotti online. L'ottimizzazione delle decisioni nella catena di approvvigionamento (ad esempio, la gestione delle scorte) è fondamentale per la competitività dell'azienda, in quanto permette di ottenere con precisione il numero di prodotti necessari in ogni centro di distribuzione. Di conseguenza, l'azienda può avere a disposizione un'ampia selezione di prodotti con tempi di consegna ridotti e prezzi competitivi, per una maggiore soddisfazione dei clienti. L'input principale nel sistema software della catena di approvvigionamento è una stima della domanda o la previsione delle potenziali vendite di ogni prodotto nel catalogo. Questa previsione determina importanti decisioni a valle, tra cui:

- Pianificazione a livello macro (previsione strategica): per un'azienda nel suo complesso, qual è la crescita prevista in termini di vendite/ricavi totali? Dove deve essere (più) geograficamente attiva l'azienda? Come deve essere impiegata la manodopera?
- Previsione della domanda (o dell'inventario): quante unità di ogni prodotto verranno probabilmente vendute in ogni sede?
- Attività promozionale (previsione tattica): come devono essere gestite le promozioni? I prodotti devono essere liquidati?

Il resto del caso di studio è incentrato sul secondo problema, che fa parte della famiglia di problemi di previsione di natura operativa (Januschowski e Kolassa, 2019). Questo documento è incentrato sulle questioni principali: dati, modelli (predittori), inferenze (previsioni) e produttivizzazione.

Ai fini di questo caso di studio, è importante ricordare che il problema di previsione è un mezzo per raggiungere un fine. Benché le previsioni siano di fondamentale importanza per l'azienda, le decisioni a valle per la catena di approvvigionamento lo sono ancora di più. Nel nostro caso di studio queste decisioni vengono prese da sistemi di acquisto automatizzati che si basano su modelli di ottimizzazione matematica derivanti da ricerche operative. Questi sistemi cercano di ridurre al minimo i costi previsti per l'azienda.

La parola chiave è previsti, ovvero le previsioni devono riguardare non solo un possibile futuro, ma tutti i futuri possibili, con la ponderazione appropriata in base alla probabilità di un determinato risultato. A questo scopo, il fattore chiave per il processo decisionale a valle è una distribuzione completa dei valori di previsione anziché una semplice previsione puntuale. La figura seguente

mostra una previsione probabilistica, nota anche come previsione di densità. Tieni presente che è possibile ricavare una singola previsione puntuale (il futuro più probabile) da questa previsione probabilistica, mentre sarà più difficile passare da una previsione puntuale a una probabilistica.

Data una previsione probabilistica, è possibile ricavarne statistiche diverse e personalizzare i risultati in modo da semplificare la decisione da prendere. L'azienda di e-commerce può avere una serie di prodotti principali che devono sempre essere disponibili in magazzino. In questo caso, usiamo un quantile elevato (ad esempio, il 90° percentile), che si traduce nel 90% delle volte in cui i prodotti saranno disponibili in magazzino. Per altri prodotti, ad esempio quelli per cui è più facile reperire articoli sostitutivi (come per le matite), l'uso di un percentile inferiore può essere più appropriato.

Amazon Forecast permette di ottenere facilmente diversi quantili dalla previsione probabilistica.

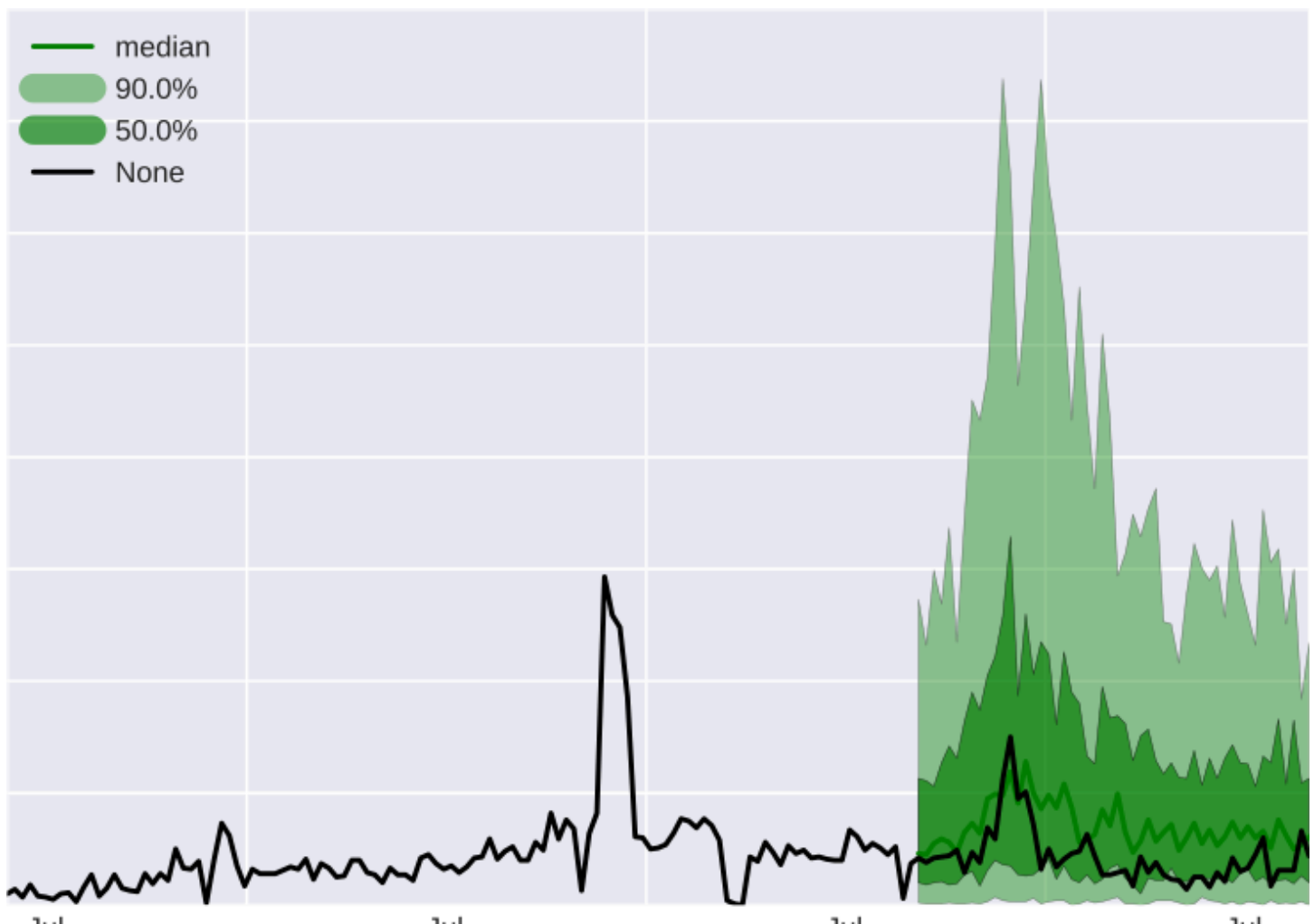


Figura della previsione probabilistica

Nella figura precedente la linea nera rappresenta i valori effettivi, la linea verde scuro è la mediana della distribuzione della previsione; l'area verde scuro ombreggiata è l'intervallo di previsione in cui

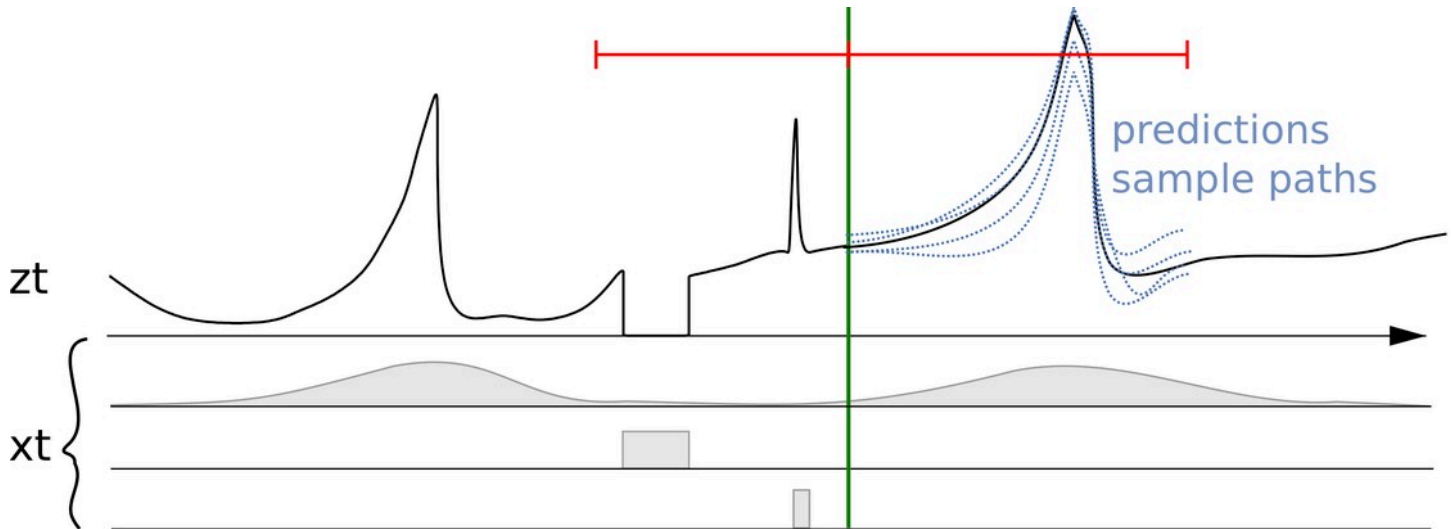
prevediamo una diminuzione del 50% dei valori e l'area verde chiaro è l'intervallo di previsione in cui prevediamo il 90% dei valori effettivi.

Le sezioni seguenti presentano le fasi per la risoluzione del problema di previsione per l'azienda, tra cui:

- [Raccolta e aggregazione dei dati \(fase 1\)](#)
- [Preparazione dei dati \(fase 2\)](#)
- [Creazione di un predittore \(fase 3\)](#)
- [Valutazione dei predittori \(fase 4\)](#)
- [Automazione della generazione di previsioni \(fase 5\)](#)

Fase 1: raccolta e aggregazione dei dati

La figura seguente mostra un modello mentale per il problema di previsione. L'obiettivo è prevedere la serie temporale z_t nel futuro, usando quante più informazioni pertinenti per rendere la previsione il più accurata possibile. Di conseguenza, la prima e più importante fase consiste nel raccogliere quanti più dati corretti possibile.



Serie temporale z_t insieme a caratteristiche o covariate (x_t) associate e previsioni multiple

La figura precedente mostra previsioni multiple a destra della linea verticale. Queste previsioni sono esempi della distribuzione della previsione probabilistica o, viceversa, possono essere usate per rappresentare la previsione probabilistica.

Le informazioni principali che un'azienda di vendita al dettaglio deve registrare sono le seguenti:

- Dati sulle vendite transazionali: ad esempio, SKU (codice di riferimento del prodotto), ubicazione, timestamp e unità vendute.
- Dati relativi ai dettagli degli articoli degli SKU: metadati di un articolo. Alcuni esempi includono colore, reparto, dimensioni e così via.
- Dati sui prezzi: serie temporale dei prezzi di ogni articolo con timestamp.
- Dati informativi sulle promozioni: diversi tipi di promozioni, su una raccolta di articoli (categoria) o su singoli articoli con timestamp.
- Dati informativi sulla disponibilità in magazzino: per ogni unità di tempo, informazioni relative alla disponibilità in magazzino o alla possibilità di acquisto di uno SKU rispetto alla non disponibilità.

- **Dati sull'ubicazione:** l'ubicazione di un articolo o la vendita in un momento specifico può essere rappresentata come stringa `location_idstore_id` o come effettiva geolocalizzazione. Le geolocalizzazioni possono essere costituite dal codice del paese più il codice postale o da coordinate `latitude_longitude`. L'ubicazione è considerata una "dimensione" delle vendite transazionali.

In [Amazon Forecast](#) i dati cronologici della quantità da prevedere sono noti come serie temporale di destinazione (TTS, Target Time Series). Nell'ambito della vendita al dettaglio la TTS corrisponde ai dati sulle vendite transazionali. Altri dati cronologici, identificati esattamente nello stesso momento di ogni transazione di vendita, sono noti come serie temporale correlata (RTS, Related Time Series). Nell'ambito della vendita al dettaglio la RTS includerebbe le variabili di prezzo, promozione e disponibilità in magazzino.

Le informazioni sulla disponibilità in magazzino sono importanti perché questo problema è incentrato sulla previsione della domanda e non delle vendite, ma l'azienda registra solo le vendite. Quando uno SKU non è più disponibile, il numero di vendite è inferiore alla potenziale domanda ed è quindi importante identificare e registrare quando si verificano questi eventi di esaurimento delle scorte.

Altri set di dati da considerare includono il numero di visite nelle pagine Web, i dettagli sui termini di ricerca, attività nei social media e le informazioni sul meteo. Spesso è importante avere a disposizione dati per il passato e per il futuro da usare nei modelli. Questo è un requisito di molti modelli di previsione e per il backtesting (descritto nella sezione [Fase 4: valutazione delle previsioni](#)).

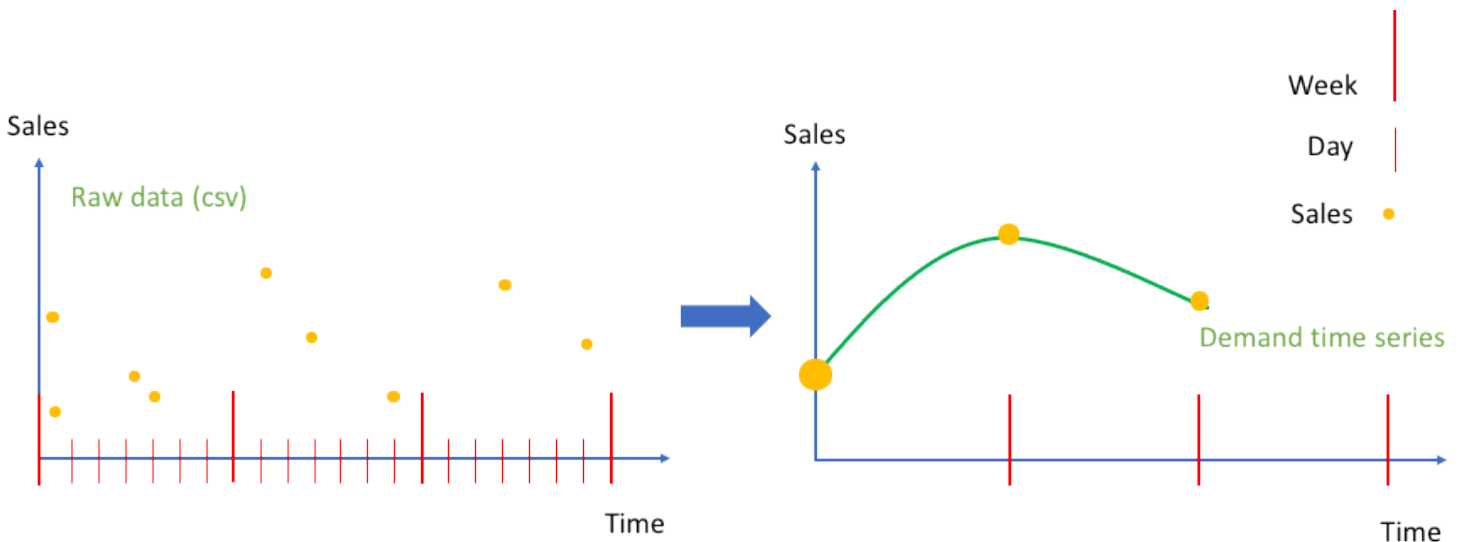
Per alcuni problemi di previsione, la frequenza dei dati non elaborati corrisponde naturalmente a quella del problema di previsione. Tra gli esempi è inclusa la richiesta del volume di server, campionata al minuto, per eseguire previsioni alla frequenza di un minuto.

I dati vengono spesso registrati a una frequenza maggiore o semplicemente con timestamp arbitrari all'interno di un intervallo di tempo, ma il problema di previsione avviene a un livello di dettaglio più grossolano. Si tratta di un evento comune nel caso di studio per la vendita al dettaglio, in cui i dati di vendita vengono registrati in genere come dati transazionali. Ad esempio, il formato è costituito da un timestamp con un livello di dettaglio elevato relativamente a quando hanno avuto luogo le vendite. Nel caso d'uso delle previsioni, questo basso livello di dettaglio può non essere necessario e può essere più appropriato aggregare i dati in vendite orarie o giornaliere. Qui il livello di aggregazione corrisponde al problema a valle, ad esempio la gestione dell'inventario o la pianificazione delle risorse.

Esempio

Il grafico a sinistra nella figura seguente mostra un esempio dei dati non elaborati sulle vendite ai clienti, che possono essere immessi in Amazon Forecast come file con valori separati da virgole (CSV). In questo esempio i dati di vendita sono definiti su una griglia temporale giornaliera più precisa e il problema consiste nel prevedere la domanda settimanale in futuro sulla griglia temporale più grossolana. Amazon Forecast esegue l'aggregazione dei valori giornalieri in una determinata settimana nella chiamata API `create_predictor`.

Il risultato trasforma i dati non elaborati in una raccolta di serie temporali dal formato corretto con una frequenza settimanale fissa. Il grafico a destra mostra questa aggregazione nella serie temporale di destinazione usando il metodo di aggregazione sommatoria predefinito. Altri metodi di aggregazione includono media, massimo, minimo o la scelta di un singolo punto, ad esempio il primo. Il livello di dettaglio e il metodo di aggregazione devono essere scelti in modo che corrispondano al meglio ai dati del caso d'uso aziendale. In questo esempio il valore aggregato è allineato all'aggregazione settimanale. Altri metodi di aggregazione possono essere impostati dall'utente usando la chiave `FeaturizationMethodParameters` del parametro `FeaturizationConfig` dell'API `create_predictor`.



Aggregazione di dati di vendita non elaborati come eventi (a sinistra), in una serie temporale equidistante (a destra)

Fase 2: preparazione dei dati

Una volta ottenuti dati non elaborati, devi gestire le complicazioni, ad esempio i dati mancanti, e preparare i dati per i modelli di previsione per acquisire al meglio l'interpretazione prefissata.

Come gestire i dati mancanti

Un elemento comune nei problemi di previsione reali consiste nell'esistenza di valori mancanti nei dati non elaborati. Un valore mancante in una serie temporale significa che l'effettivo valore corrispondente in ogni momento alla frequenza specificata non è disponibile per l'ulteriore elaborazione. I motivi per cui i valori vengono contrassegnati come mancanti possono essere diversi.

I valori mancanti possono essere dovuti all'assenza di transazioni o a possibili errori di misurazione, ad esempio perché un servizio che monitora certi dati non ha funzionato correttamente o perché non è stato possibile eseguire correttamente la misurazione. L'esempio principale di quest'ultima circostanza nel caso di studio per la vendita al dettaglio è una situazione di esaurimento delle scorte nella previsione della domanda, per cui la domanda non equivale alle vendite nel giorno specifico.

Effetti simili possono verificarsi in scenari di cloud computing quando un servizio raggiunge un limite, ad esempio se le istanze [Amazon EC2](#) in una determinata [Regione AWS](#) sono tutte occupate. Un altro esempio di valori mancanti è il caso di un prodotto o un servizio non ancora lanciato o fuori produzione.

Valori mancanti possono essere inseriti anche da componenti di elaborazione delle caratteristiche, per garantire la stessa lunghezza delle serie temporali con il padding. Se numerosi, i valori mancanti possono avere un forte impatto sulla precisione di un modello.

Esempio 1

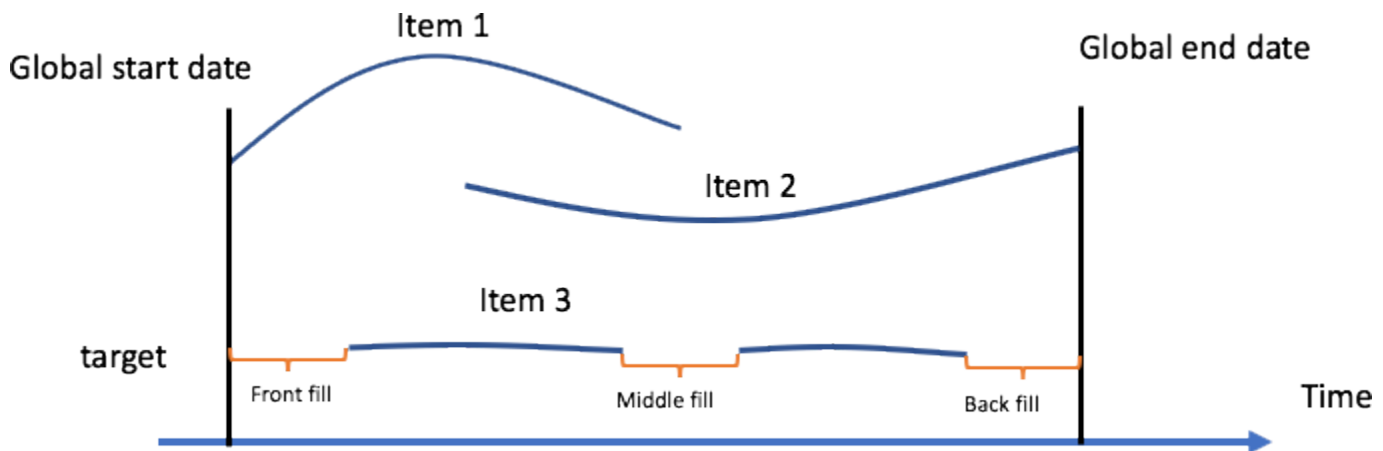
Il riempimento è il processo di aggiunta di valori standardizzati alle voci mancanti nel set di dati. La figura seguente mostra le diverse strategie per gestire i valori mancanti in Amazon Forecast (riempimento in avanti, centrale, all'indietro e futuro), per l'articolo 2 in un set di dati di tre articoli.

Amazon Forecast supporta il riempimento per serie temporali di destinazione e correlate. La data di inizio globale è definita come la prima data di inizio rispetto alle date di inizio di tutti gli articoli nel set di dati. Nell'esempio seguente la data di inizio globale è in corrispondenza dell'articolo 1.

Analogamente, la data di fine globale è definita come l'ultima data di fine della serie temporale di tutti gli articoli, in corrispondenza dell'articolo 2.

Il riempimento in avanti immette ogni valore dall'inizio della serie temporale specifica fino alla data di inizio globale. Al momento della pubblicazione di questo documento, Amazon Forecast non attiva alcun riempimento in avanti e permette l'inizio delle serie temporali in momenti diversi. Il riempimento centrale indica i valori che sono stati immessi a metà della serie temporale, ad esempio tra le date di inizio e di fine degli articoli, mentre il riempimento all'indietro immette i valori dall'ultima data della serie temporale alla data di fine globale.

Per le serie temporali di destinazione, i metodi di riempimento centrale e all'indietro hanno la logica predefinita di riempimento a zero. Il riempimento futuro, che si applica solo a serie temporali correlate, immette qualsiasi valore mancante tra la data di fine globale degli articoli e l'orizzonte di previsione specificato dal cliente. I valori futuri sono obbligatori per il set di dati delle serie temporali correlate con [Prophet](#) e [DeepAR+](#) e facoltativi per [CNN-QR](#).



Strategie di gestione dei valori mancanti in Amazon Forecast

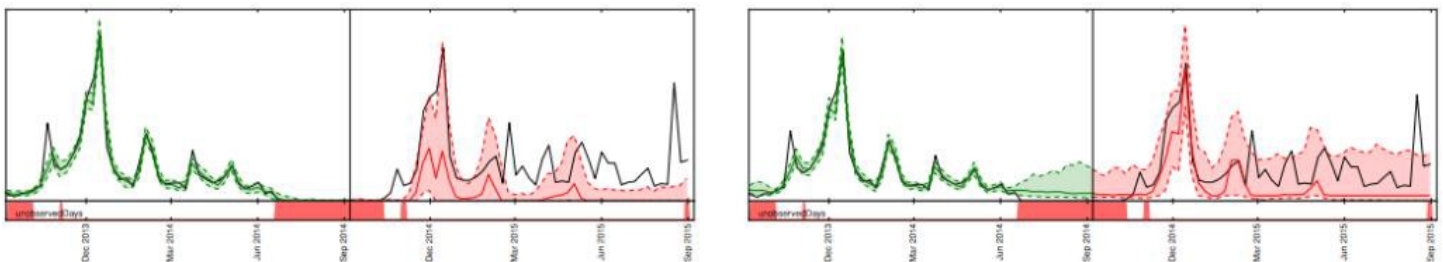
Nella figura precedente la data di inizio globale indica la prima data di inizio rispetto alle date di inizio di tutti gli articoli, mentre la data di fine globale indica l'ultima data di fine rispetto alle date di fine di tutti gli articoli. L'orizzonte di previsione è il periodo durante il quale Forecast fornisce previsioni per il valore di destinazione.

Si tratta di uno scenario comune nel caso di studio per la vendita al dettaglio, che rappresenta zero vendite per i dati transazionali per gli articoli disponibili. Questi valori vengono gestiti come veri zeri e usati nel componente di valutazione delle metriche. Amazon Forecast permette all'utente di identificare i valori effettivamente mancanti e di codificarli come diversi da numeri (NaN, Not a Number) durante l'elaborazione da parte degli algoritmi. Questo articolo esamina i motivi per cui i due casi differiscono e quando ognuno è utile.

Nel caso di studio per la vendita al dettaglio l'informazione indicante che un rivenditore ha venduto zero unità di un articolo disponibile differisce da quella che specifica che sono state vendute zero unità di un articolo non disponibile, nei periodi oltre la sua esistenza (ad esempio, prima della sua introduzione o dopo la sua obsolescenza) o durante la sua esistenza (ad esempio, quando è parzialmente non disponibile o quando non sono stati registrati dati di vendita per l'intervallo di tempo). Il riempimento a zero predefinito è applicabile in questo primo caso. Nell'ultimo caso, anche se il valore di destinazione corrispondente è in genere zero, il valore contrassegnato come mancante comunica informazioni aggiuntive. La best practice consiste nel preservare le informazioni indicanti dati mancanti anziché eliminarle. Fai riferimento all'esempio seguente per un'illustrazione del motivo per cui è importante mantenere le informazioni.

Amazon Forecast supporta logica di riempimento di altri tipi, tra cui valore, media, mediano, minimo e massimo. Per le serie temporali correlate, ad esempio prezzo o promozione, non sono specificati valori predefiniti per i metodi di riempimento centrale, all'indietro o futuro, perché la logica corretta per i valori mancanti varia in base al tipo di attributo e al caso d'uso. La logica di riempimento supportata per le serie temporali correlate include zero, valore, media, mediano, minimo e massimo.

Per eseguire il riempimento dei valori mancanti, specifica i tipi di riempimento da implementare quando chiami l'operazione [CreatePredictor](#). La logica di riempimento è specificata in oggetti [FeaturizationMethod](#). Ad esempio, per codificare un valore che non rappresenta zero vendite di un prodotto non disponibile nella serie temporale di destinazione, puoi contrassegnare un valore come effettivamente mancante impostando il tipo di riempimento su NaN. A differenza del riempimento a zero, i valori codificati con NaN vengono considerati effettivamente mancanti e non vengono usati nel componente di valutazione delle metriche



Effetto del riempimento a zero rispetto al riempimento tramite NaN sulle previsioni per lo stesso articolo

Nel grafico a sinistra nella figura precedente i valori a sinistra della linea verticale nera vengono riempiti con 0, producendo una previsione con distorsione per difetto (a destra della linea verticale nera). Nel grafico a destra questi valori sono contrassegnati come NaN, producendo previsioni appropriate.

Esempio 2

La figura precedente mostra l'importanza di una corretta gestione dei valori mancanti per un modello spaziale lineare, come [ARIMA o ETS](#). Il grafico rappresenta la previsione della domanda per un articolo parzialmente esaurito. L'area di addestramento viene mostrata nel grafico a sinistra in verde, l'intervallo di previsione nel riquadro a destra in rosso e l'effettiva destinazione in nero. Le previsioni mediana, p10 e p90 vengono mostrate rispettivamente nella linea rossa e nell'area ombreggiata. La parte inferiore mostra gli articoli esauriti (80% dei dati) in rosso. Nel grafico a sinistra le aree relative all'esaurimento degli articoli vengono ignorate e riempite con 0.

In questo modo, i modelli di previsione presuppongono la presenza di molti zeri da stimare e di conseguenza le previsioni sono troppo basse. Nel grafico a destra le aree relative all'esaurimento degli articoli vengono considerate informazioni effettivamente mancanti e la domanda diventa incerta in questa regione. Quando i valori mancanti per gli articoli non disponibili sono contrassegnati correttamente come NaN, l'intervallo di previsione in questo grafico non mostra una distorsione per difetto. Amazon Forecast colma queste lacune nei dati, semplificando la corretta gestione dei dati mancanti, senza dover modificare esplicitamente tutti i dati di input.

Concetti relativi alla caratterizzazione e alle serie temporali correlate

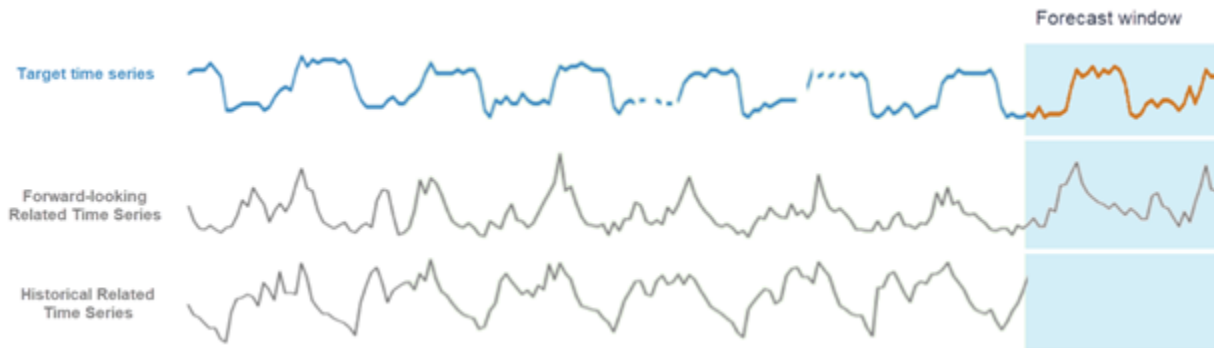
Amazon Forecast permette agli utenti di immettere dati correlati per migliorare la precisione di alcuni modelli di previsione supportati. Questi dati possono essere di due tipi: serie temporali correlate o metadati statici degli articoli.

Note

I metadati e i dati correlati sono noti come caratteristiche nell'ambito del machine learning e covariate in statistica.

Le serie temporali correlate sono serie temporali che hanno una correlazione con il valore di destinazione e dovrebbero conferire una certa efficacia alla previsione nel valore di destinazione, perché forniscono una spiegazione in termini intuitivi. Per un esempio, consulta [Amazon Forecast: predicting time-series at scale](#). A differenza delle serie temporali di destinazione, le serie temporali correlate sono valori noti nel passato che possono avere impatto sulle serie temporali di destinazione e includere valori noti nel futuro.

In Amazon Forecast puoi aggiungere due tipi di serie temporali correlate: serie temporali cronologiche e serie temporali orientate al futuro. Le serie temporali correlate cronologiche contengono punti dati fino all'orizzonte di previsione e nessuno all'interno dell'orizzonte di previsione nel futuro. Le serie temporali correlate orientate al futuro contengono punti dati fino all'orizzonte di previsione e al suo interno.



Diversi approcci all'uso di serie temporali correlate con Amazon Forecast

Esempio 3

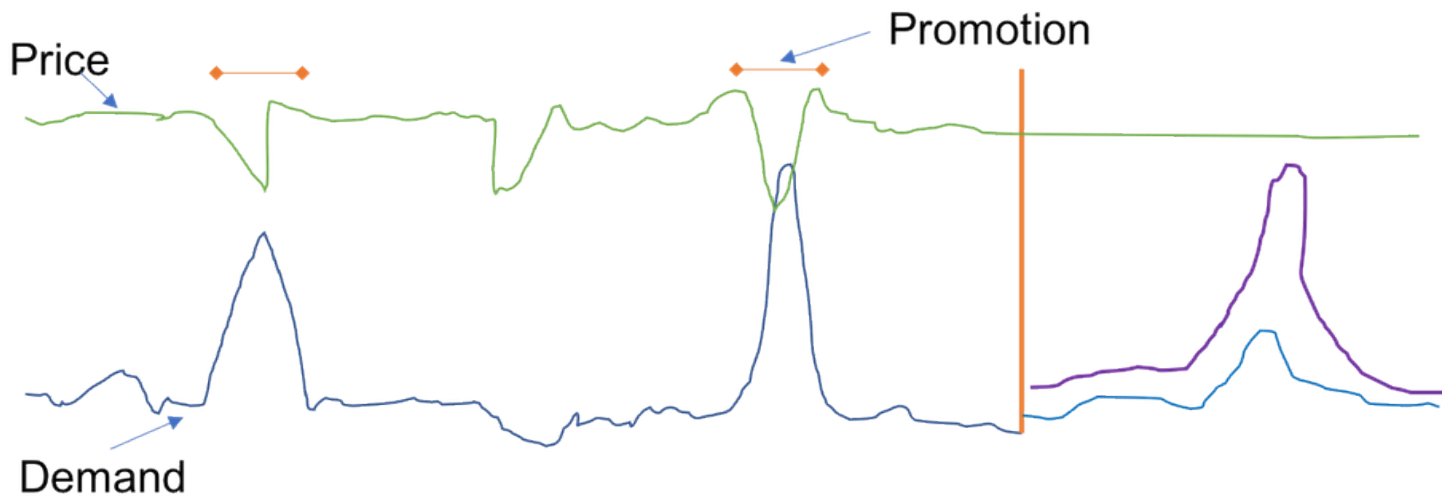
La figura seguente mostra un esempio di un possibile uso delle serie temporali correlate per prevedere la futura domanda di un libro di successo. La linea blu rappresenta la domanda nella serie temporale di destinazione. Il prezzo è indicato dalla linea verde. La linea verticale rappresenta la data di inizio della previsione e le previsioni ai due quantili vengono mostrate a destra della linea verticale.

Questo esempio usa una serie temporale correlata orientata al futuro in linea con la serie temporale di destinazione per il livello di dettaglio della previsione ed è nota in tutti i momenti (o nella maggior parte) nel futuro nell'intervallo tra la data di inizio della previsione fino alla data di inizio della previsione incrementata dall'orizzonte di previsione (data di fine della previsione).

La figura seguente mostra anche che il prezzo è una caratteristica utile da usare, in quanto possiamo osservare le correlazioni tra una riduzione del prezzo e un aumento delle vendite del prodotto. Le serie temporali correlate possono essere fornite ad Amazon Forecast tramite un file CSV separato, contenente lo SKU dell'articolo, il timestamp e i valori della serie temporale correlata (in questo caso, il prezzo).

Amazon Forecast supporta metodi di aggregazione, come media e sommatoria, per le serie temporali di destinazione, ma non per le serie temporali correlate. Ad esempio, non ha molto senso sommare un prezzo giornaliero a un prezzo settimanale e lo stesso vale per le promozioni giornaliere.

Amazon Forecast può incorporare automaticamente le informazioni sul [meteo](#) e sulle [festività](#) in un modello includendo set di dati caratterizzati predefiniti (consulta [SupplementaryFeature](#)). Le informazioni sul meteo e le festività possono influire notevolmente sulla domanda al dettaglio.



Vendite di un articolo specifico (in blu, a sinistra della linea verticale rossa)

I metadati degli articoli, noti anche come variabili categoriche, sono altre caratteristiche utili che possono essere immesse in Amazon Forecast. Per un esempio, consulta [Amazon Forecast: predicting time-series at scale](#). La differenza principale tra le variabili categoriche e le serie temporali correlate è che le prime sono statiche, ovvero non cambiano nel tempo. Gli esempi più comuni nell'ambito della vendita al dettaglio includono i colori degli articoli, le categorie di libri e indicatori binari che specificano, ad esempio, se un televisore è o meno una smart TV. Queste informazioni possono essere acquisite dagli algoritmi di deep learning per apprendere le analogie tra SKU, presupponendo che SKU simili avranno vendite analoghe. Poiché questi metadati non hanno una dipendenza temporale, ogni riga del file CSV dei metadati degli articoli è costituita solo dallo SKU dell'articolo e dall'etichetta di categoria o dalla descrizione corrispondente.

Fase 3: creazione di un predittore

Un predittore può essere creato in due modi: eseguendo [AutoML](#) o selezionando manualmente uno dei sei algoritmi Amazon Forecast predefiniti. Al momento della stesura di questo documento, durante l'esecuzione di AutoML Amazon Forecast prova automaticamente i sei algoritmi predefiniti e sceglie quello con le perdite quantiliche medie più basse rispetto al 10°, 50° (mediano) e 90° quantile.

Amazon Forecast offre quattro modelli locali:

- ARIMA ([Autoregressive Integrated Moving Average, autoregressivo integrato a media mobile](#))
- ETS ([Exponential Smoothing, livellamento esponenziale](#))
- NPTS ([Non-Parametric Time Series, serie temporali non parametriche](#))
- [Prophet](#)

I modelli locali sono metodi di previsione che adattano un singolo modello a ogni serie temporale (o a ogni specifica combinazione di articolo/dimensione) e quindi usano questi modelli per estrapolare la serie temporale nel futuro.

I modelli ARIMA ed ETS sono versioni scalabili dei modelli locali più diffusi del pacchetto di previsioni R. Il modello NPTS, un metodo locale sviluppato da Amazon, presenta una differenza fondamentale rispetto agli altri modelli locali. A differenza dei semplici modelli di previsione stagionali, che forniscono previsioni puntuali ripetendo l'ultimo valore o il valore in corrispondenza di una stagionalità appropriata, NPTS produce previsioni probabilistiche. NPTS usa un indice temporale fisso, in cui l'indice precedente ($T - 1$) o la stagione passata ($T - \tau$) è la previsione della fase temporale T . L'algoritmo campiona in modo casuale un indice temporale (t) nel set $\{0, \dots, T - 1\}$ per generare un campione per l'attuale fase temporale T . NPTS è particolarmente efficace per serie temporali intermittenti (a volte note anche come sparse) con molti zeri. Forecast include anche l'implementazione Python di Prophet, un modello di serie temporale strutturale bayesiano.

Amazon Forecast offre due algoritmi di deep learning globali:

- [DeepAR+](#)
- [CNN-QR](#)

I modelli globali addestrano un singolo modello sull'intera raccolta di serie temporali in un set di dati. Sono particolarmente utili quando sono presenti serie temporali simili in un insieme di unità

trasversali, ad esempio raggruppamenti di serie temporali della domanda per diversi prodotti, carichi di server e richieste di pagine Web.

In generale, con l'aumentare del numero di serie temporali, aumenta l'efficacia dei modelli CNN-QR e DeepAR+. Questo non sempre avviene per i modelli locali. I modelli di deep learning possono essere usati anche per generare previsioni per nuovi SKU con nessuno o pochi dati di vendita cronologici.

Queste sono note come [previsioni di chiamate a freddo](#).

	Neural Networks		Flexible Local Algorithms	Baseline Algorithms		
	CNN-QR	DeepAR+	Prophet	NPTS	ARIMA	ETS
Computationally intensive training process	High	High	Medium	Low	Low	Low
Accepts historical related time series*	✓	✗	✗	✗	✗	✗
Accepts forward-looking related time series*	✓	✓	✓	✗	✗	✗
Accepts item metadata (product color, brand, etc)	✓	✓	✗	✗	✗	✗
Suitable for sparse datasets	✓	✓	✗	✓	✗	✗
Performs Hyperparameter Optimization (HPO)	✓	✓	✗	✗	✗	✗
Allows overriding default hyperparameter values	✓	✓	✗	✓	✗	✗
Suitable for What-if analysis	✓	✓	✓	✗	✗	✗
Suitable for Cold Start scenarios (forecasting with little to no historical data)	✓	✓	✗	✗	✗	✗

Confronto degli algoritmi disponibili in Amazon Forecast

Per ulteriori informazioni sulle serie temporali correlate, consulta questa [pagina](#).

Fase 4: valutazione dei predittori

Un tipico flusso di lavoro nel machine learning consiste nell'addestrare un set di modelli o una combinazione di modelli su un set di addestramento e quindi valutarne la precisione su un set di dati di controllo. Questa sezione spiega come dividere i dati cronologici e quali metriche usare per valutare i modelli nella previsione di serie temporali. Per la previsione, la tecnica del backtesting è lo strumento principale per valutare la precisione delle previsioni.

Backtesting

Un framework di valutazione e backtesting adeguato è tra i fattori più importanti per realizzare un'applicazione di machine learning di successo. Puoi usare test retrospettivi efficaci con i tuoi modelli per acquisire familiarità con le capacità predittive future dei modelli. Inoltre, puoi ottimizzare i modelli tramite l'ottimizzazione degli iperparametri, apprendere le combinazioni dei modelli e abilitare il meta-learning e AutoML.

La caratteristica tempo della previsione di serie temporali si distingue in termini di metodologia di valutazione e backtesting dagli altri campi di applicazione del machine learning. In genere, nelle attività di machine learning per valutare l'errore predittivo in un test retrospettivo, è necessario dividere un set di dati in base agli articoli. Ad esempio, per la convalida incrociata in attività correlate alle immagini, eseguirai l'addestramento su una certa percentuale delle immagini e quindi userai altre parti per i test e la convalida. Nella previsione devi dividere i dati principalmente in base al tempo (e in misura minore in base agli articoli) per evitare la fuoriuscita di informazioni dal set di addestramento al set per i test e la convalida e per simulare il caso di produzione il più fedelmente possibile.

La divisione in base al tempo deve essere eseguita con attenzione, perché è bene non scegliere un singolo momento, ma più momenti. In caso contrario, la precisione dipenderà eccessivamente dalla data di inizio della previsione, definita dal punto di divisione. Una valutazione continuativa della previsione, in cui viene eseguita una serie di divisioni su più momenti e che restituisce il risultato medio, produce risultati più solidi e affidabili durante i test retrospettivi. La figura seguente mostra quattro diverse divisioni per i test retrospettivi.

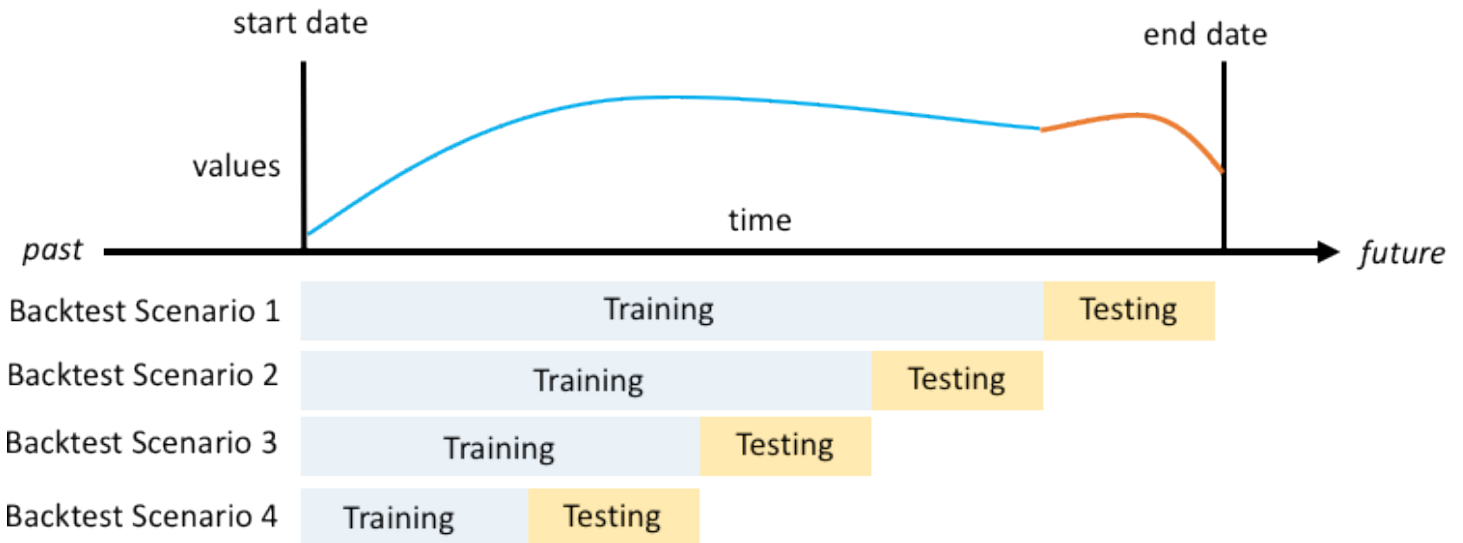


Figura di quattro scenari di backtesting diversi con set di addestramento di dimensioni crescenti, ma test con dimensioni costanti

Nella figura precedente tutti gli scenari di backtesting hanno dati interamente disponibili in modo da valutare i valori previsti rispetto a quelli effettivi.

Il motivo per cui sono necessarie più finestre di backtesting è che la maggior parte delle serie temporali reali è normalmente non stazionaria. L'azienda di e-commerce presentata nel case study ha sede in Nord America e gran parte della sua domanda di prodotti è determinata dal picco del quarto trimestre, con picchi particolari attorno al Ringraziamento e prima di Natale. Nella stagione degli acquisti del quarto trimestre la variabilità delle serie temporali è maggiore rispetto al resto dell'anno. La presenza di più finestre di backtesting permette di valutare i modelli di previsione in un contesto più equilibrato.

Per ogni scenario di backtesting, la figura seguente mostra gli elementi di base della terminologia di Amazon Forecast. Amazon Forecast divide automaticamente i dati in set di dati di addestramento e test. Amazon Forecast decide come dividere i dati di input usando il parametro `BackTestWindowOffset` specificato come parametro nell'API `create_predictor` oppure usa il valore predefinito di `ForecastHorizon`.

Nella figura seguente puoi osservare il caso precedente più generale quando i parametri `ForecastHorizon` e `BackTestWindowOffset` non sono uguali. Il parametro `BackTestWindowOffset` definisce una data di inizio della previsione virtuale, mostrata come linea verticale tratteggiata nella figura seguente. Può essere usato per rispondere a questa domanda ipotetica: se il modello viene implementato in questo giorno, quale sarà la previsione?

`ForecastHorizon` definisce il numero di fasi temporali da prevedere dalla data di inizio della previsione virtuale.

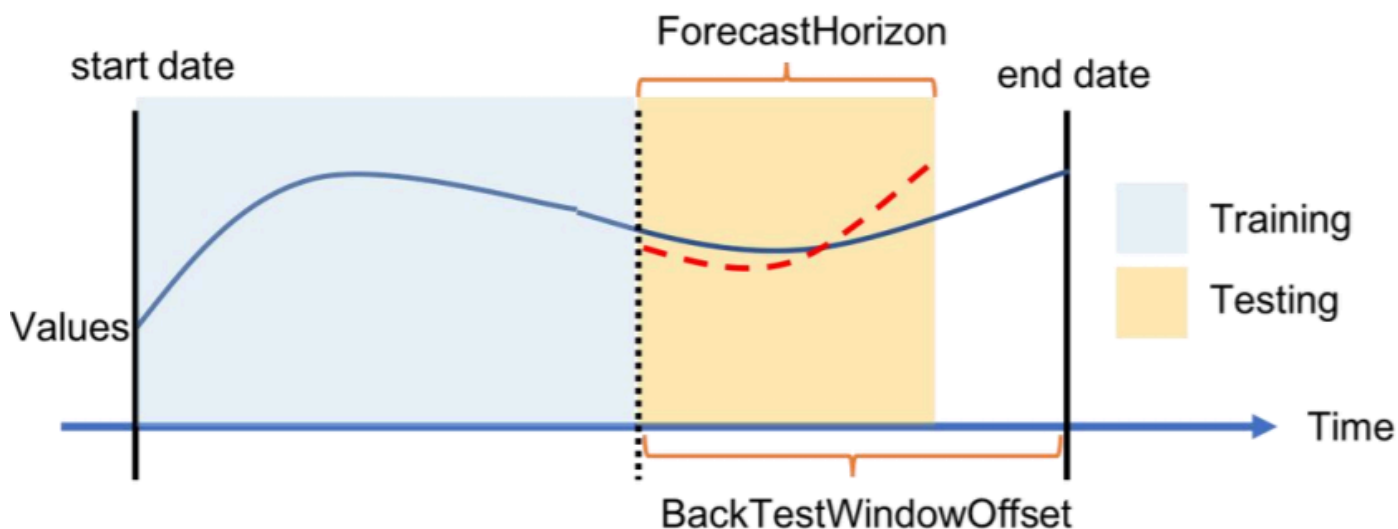


Figura di un singolo scenario di backtesting e della sua configurazione in Amazon Forecast

Amazon Forecast può esportare i valori previsti e le metriche di precisione generate durante il backtesting. I dati esportati possono essere usati per valutare determinati articoli in momenti e a quantili specifici.

Quantili di previsione e metriche di precisione

I quantili di previsione possono fornire un limite superiore e uno inferiore per le previsioni. Ad esempio, usando i tipi di previsione 0,1 (p10), 0,5 (p50) e 0,9 (p90), viene fornito un intervallo di valori noto come intervallo di confidenza dell'80% intorno alla previsione p50. Generando previsioni a p10, p50 e p90, possiamo aspettarci che il valore reale rientri tra questi limiti l'80% delle volte.

Continueremo a parlare dei quantili nella sezione [Fase 5](#).

Amazon Forecast usa le metriche di precisione di perdita quantilica ponderata (wQL, Weighted Quantile Loss), errore quadratico medio (RMSE, Root Mean Square Error) ed errore percentuale assoluto ponderato (WAPE, Weighted Absolute Percentage Error) per valutare i predittori durante il backtesting.

Perdita quantilica ponderata (wQL)

La metrica di errore della perdita quantilica ponderata (wQL) misura la precisione della previsione di un modello a un quantile specificato. È particolarmente utile in presenza di costi diversi per

le previsioni sottostimate e quelle sovrastimate. L'impostazione del peso (τ) della funzione wQL incorpora automaticamente penalità diverse per le previsioni sottostimate e quelle sovrastimate.

$$wQL[\tau] = 2 \frac{\sum_{i,t} [\tau \max(y_{i,t} - q_{i,t}^{(\tau)}, 0) + (1 - \tau) \max(q_{i,t}^{(\tau)} - y_{i,t}, 0)]}{\sum_{i,t} |y_{i,t}|}$$

Funzione wQL

Dove:

- τ : quantile nel set $\{0,01, 0,02, \dots, 0,99\}$
- $q_{i,t}(\tau)$: quantile τ previsto dal modello.
- y_i, t : valore osservato in corrispondenza del punto (i,t)

Errore percentuale assoluto ponderato (WAPE)

L'errore percentuale assoluto ponderato (WAPE) è una metrica comunemente usata per misurare la precisione del modello. La metrica misura la deviazione complessiva dei valori previsti rispetto ai valori osservati.

$$WAPE = \frac{\sum_{i,t} |y_{i,t} - \hat{y}_{i,t}|}{\sum_{i,t} |y_{i,t}|}$$

WAPE

Dove:

- y_i, t : valore osservato in corrispondenza del punto (i, t)
- $\hat{y}_{i,t}$: valore previsto in corrispondenza del punto (i,t)

Per la previsione viene usata la previsione media come valore previsto $\hat{y}_{i,t}$.

Errore quadratico medio (RMSE)

$$RMSE = \sqrt{\frac{1}{nT} \sum_{i,t} (\hat{y}_{i,t} - y_{i,t})^2}$$

L'errore quadratico medio (RMSE) è una metrica comunemente usata per misurare la precisione del modello. Come la metrica WAPE, misura la deviazione complessiva delle stime rispetto ai valori osservati.

Dove:

- $y_{i,t}$: valore osservato in corrispondenza del punto (i, t)
- $\hat{y}_{i,t}$: valore previsto in corrispondenza del punto (i,t)
- nT : numero di punti dati in un set di test

Per la previsione viene usata la previsione media come valore previsto $\hat{y}_{i,t}$. Nel calcolo delle metriche dei predittori, nT è il numero di punti dati in una finestra di backtesting.

Problemi con WAPE e RMSE

Nella maggior parte dei casi, le previsioni puntuali che possono essere generate internamente o da altri strumenti di previsione devono corrispondere al quantile p50 o alle previsioni medie. Per le metriche WAPE ed RMSE, Amazon Forecast usa la previsione media per rappresentare il valore previsto ($\hat{y}$).

Per $\tau = 0,5$ nell'equazione $wQL[\tau]$, i due pesi sono uguali e $wQL[0,5]$ si riduce all'errore percentuale assoluto ponderato (WAPE) comunemente usato per le previsioni puntuali:

$$wQL[0.5] = 2 \frac{\sum_{i,t} 0.5 [\max(y_{i,t} - q_{i,t}^{(0.5)}, 0) + \max(q_{i,t}^{(0.5)} - y_{i,t}, 0)]}{\sum_{i,t} |y_{i,t}|} = \frac{\sum_{i,t} |y_{i,t} - q_{i,t}^{(0.5)}|}{\sum_{i,t} |y_{i,t}|}$$

Dove $\hat{y} = q(0,5)$ è la previsione del calcolo. Nella formula wQL viene usato il fattore di scala 2 per annullare il fattore 0,5 e ottenere l'esatta espressione $WAPE[\text{mediano}]$.

La definizione precedente della metrica WAPE differisce dall'interpretazione comune dell'errore assoluto medio percentuale ([MAPE](#)). La differenza consiste nel denominatore. La definizione precedente della metrica WAPE evita il problema della divisione per 0, comune in scenari reali come l'azienda di e-commerce oggetto del case study, che spesso vende 0 unità di un determinato SKU in un certo giorno.

A differenza della metrica di perdita quantilica ponderata per τ diverso da 0,5, la distorsione intrinseca di ogni quantile non può essere acquisita da un calcolo come WAPE, in cui i pesi sono

uguali. Altri svantaggi della metrica WAPE: non è simmetrica, ha un'inflazione eccessiva degli errori percentuali per numeri piccoli ed è solo una metrica puntuale.

La metrica RMSE è il quadrato del termine di errore in WAPE e una metrica di errore comune in altre applicazioni di machine learning. La metrica RMSE favorisce un modello in cui i singoli errori hanno un ordine di grandezza coerente, perché grandi variazioni nell'errore aumentano la sovrapporzione di RMSE. A causa dell'errore quadratico, pochi valori previsti erroneamente in una previsione altrimenti valida possono aumentare la metrica RMSE. Inoltre, a causa dei termini al quadrato, i termini di errore più piccoli hanno un peso inferiore in RMSE che non in WAPE.

Le metriche di precisione permettono una valutazione quantitativa delle previsioni. In particolare per i confronti su larga scala (il metodo A è migliore del metodo B in generale?), sono fondamentali. Tuttavia, spesso è importante integrare questi risultati con immagini per singoli SKU.

Fase 5: generazione e uso di previsioni per il processo decisionale

Una volta ottenuto un modello che soddisfa la soglia di precisione necessaria per il caso d'uso specifico (in base al backtesting), la fase finale comporta l'implementazione del modello e la generazione di previsioni. Per implementare un modello in Amazon Forecast, devi eseguire l'API `Create_Forecast`. Questa azione ospita un modello creato tramite l'addestramento sull'intero set di dati cronologico, diversamente da `Create_Predictor`, che divide i dati in un set di addestramento e uno di test. Le previsioni del modello generate nell'orizzonte di previsione possono quindi essere utilizzate in due modi:

- Puoi eseguire query sulle previsioni per un determinato articolo (specificando l'articolo o la combinazione di articolo/dimensione) tramite l'API `Query_Forecast` dall'[AWS CLI](#) o direttamente tramite la [AWS Management Console](#).
- Puoi generare le previsioni per tutte le combinazioni di articoli e dimensioni in tutti i quantili usando l'API `Create_Forecast_Export_Job`. Questa API genera un file CSV che viene archiviato in modo sicuro in una posizione a tua scelta in [Amazon Simple Storage Service](#) (Amazon S3). Puoi quindi usare i dati del file CSV e inserirli nei sistemi a valle usati per il processo decisionale. Ad esempio, i sistemi della catena di approvvigionamento esistenti possono importare direttamente l'output di Amazon Forecast per orientare il processo decisionale riguardo alla produzione di SKU specifici.

Previsioni probabilistiche

Amazon Forecast può generare previsioni a diversi quantili, un'opzione particolarmente utile quando i costi delle previsioni sottostimate e di quelle sovrastimate sono diversi. Analogamente alla fase di addestramento dei predittori, vengono generate previsioni probabilistiche per quantili compresi tra p1 e p99.

Per impostazione predefinita, Amazon Forecast genera previsioni agli stessi quantili usati durante l'addestramento dei predittori. Se i quantili non vengono specificati durante l'addestramento dei predittori, per impostazione predefinita le previsioni vengono generate ai quantili p10, p50 e p90.

Per la previsione p10, l'effettivo valore sarà probabilmente inferiore al valore previsto il 10% delle volte ed è possibile usare la metrica $wQL[0,1]$ per valutarne la precisione. Di conseguenza, la

previsione p10 è sottostimata il 90% delle volte e, se usata per rifornire l'inventario, il 90% delle volte l'articolo risulterà esaurito. La previsione p10 può essere utile quando lo spazio in magazzino è ridotto o il costo del capitale investito è elevato.

Note

La definizione formale di una previsione quantilica è $\Pr(\text{valore effettivo} \leq \text{previsione al quantile } q) = q$. Tecnicamente un quantile è un percentile/100. Gli statistici tendono a dire "livello di quantile p90", perché è più facile rispetto a "quantile 0,9". Ad esempio, una previsione a livello di quantile p90 significa che ci possiamo aspettare che il valore effettivo sia inferiore alla previsione il 90% delle volte. In particolare se in corrispondenza di tempo = t1 e livello di quantile = 0,9 il valore previsto = 30, ci aspetteremo che il valore effettivo in corrispondenza di tempo = t1 con 1.000 simulazioni sia inferiore a 30 per 900 simulazioni e superiore a 30 per 100 simulazioni.

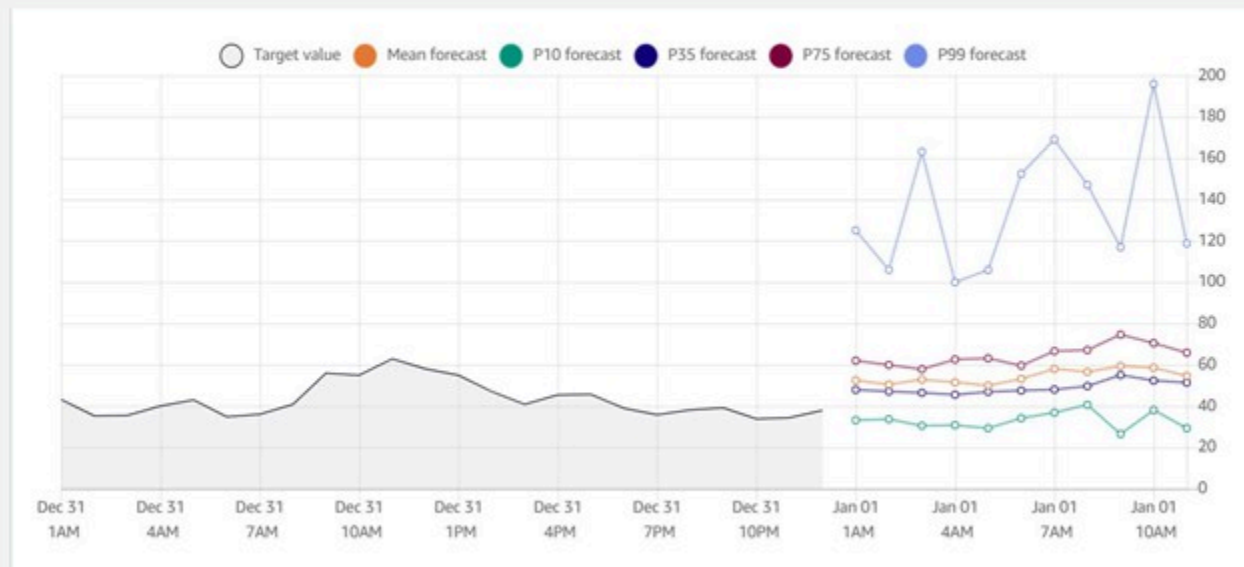
D'altro canto, la previsione p90 è una previsione sovrastimata il 90% delle volte, utile quando il costo opportunità della mancata vendita di un articolo è estremamente elevato o il costo del capitale investito è basso. Per un negozio di alimentari, la previsione p90 potrebbe essere usata per prodotti come il latte o la carta igienica, di cui il negozio non vuole mai essere sprovvisto e non si preoccupa di averne sempre una certa quantità disponibile.

Per la previsione p50 (nota spesso come previsione mediana), il valore reale è in genere inferiore al valore previsto il 50% delle volte ed è possibile usare la metrica $wQL[0,5]$ per valutarne la precisione. Quando l'eccesso di scorte non è motivo di preoccupazione e vi è una domanda moderata per un determinato articolo, la previsione quantilica p50 può essere una scelta utile.

Visualizzazione

Amazon Forecast permette di rappresentare le previsioni in modo nativo nella AWS Management Console. Inoltre, puoi trarre vantaggio dallo stack di data science Python completo (consulta [Amazon Forecast Examples](#)). Amazon Forecast esegue l'esportazione delle previsioni come file CSV tramite l'API `ExportForecastJob`, che permette agli utenti di visualizzare la previsione nel proprio strumento di analisi preferito.

Item_id: client_12



Visualizzazione fornita nella console Amazon Forecast a diversi quantili

Riepilogo del flusso di lavoro di previsione e delle API

Nella tabella seguente ogni fase del flusso di lavoro di previsione viene associata alla rispettiva API Amazon Forecast.

Tabella 1: fasi di previsione e API Amazon Forecast

Fase	API	Funzioni delle API
Fase 1: raccolta e aggregazione dei dati Fase 2: preparazione dei dati	Create_Dataset_Group , Create_Dataset , Create_Dataset_Import_Job	1. Stabiliscono l'ambito generale (vendita al dettaglio, metriche e così via) per il problema. 2. Definiscono lo schema per i diversi set di dati (destinazione, correlazione, metadati degli articoli). 3. Importa i dati da Amazon S3 ad Amazon Forecast.
Fase 3: creazione di un predittore Fase 4: valutazione dei predittori	Create_Predictor	1. Esegue le operazioni di estrazione, trasformazione e caricamento (ETL). 2. Divide i dati in set di addestramento/test e addestra il modello. 3. Puoi usare Create_predictor_backtest_Export_job per esportare i risultati dei test retrospettivi in un file CSV per il calcolo delle metriche a livello di articolo (facoltativo).

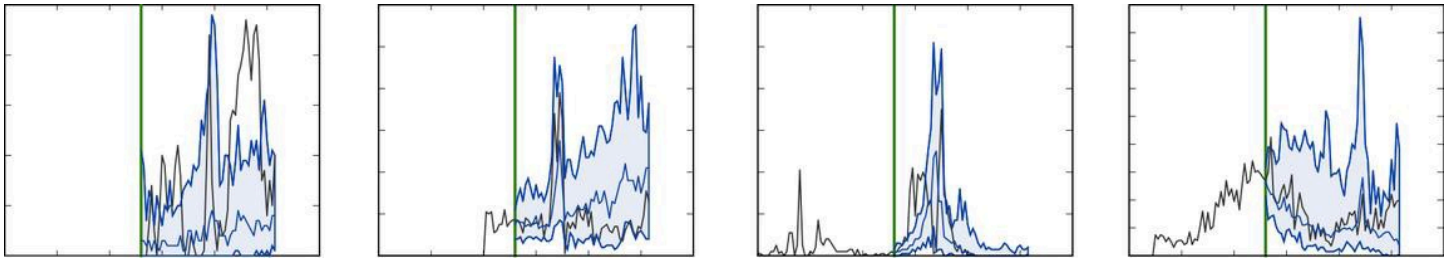
Fase	API	Funzioni delle API
<u>Fase 5: generazione e uso di previsioni per il processo decisionale</u>	Create_Forecast	1. Addestra/ospita il modello. 2. Genera previsioni nell'orizzonte di previsione per uno specifico quantile di interesse (ad esempio, qualsiasi numero intero compreso tra 1 e 99 inclusa la media).
	Query_Forecast , Create_Forecast_Export_Job	Permette di utilizzare le previsioni create da Create_Forecast

Uso di Amazon Forecast per scenari comuni

Puoi anche svolgere analisi what-if generando diverse previsioni in base alle variazioni in variabili esterne, ad esempio i prezzi o le promozioni. Nell'esempio del caso di studio per l'e-commerce, ad esempio, potresti creare previsioni diverse in base alle promozioni pianificate. Puoi prevedere la domanda di un prodotto scontato al 10% e quindi al 20% per identificare la quantità di prodotto che deve essere disponibile in magazzino per soddisfare la domanda. A questo scopo, puoi configurare gruppi di set di dati univoci e aggiornare le serie temporali correlate in ciascuno, in base allo scenario di interesse.

Puoi anche generare previsioni per articoli privi di cronologia precedente (problema noto talvolta come chiamata a freddo). Questo approccio richiede la creazione di un predittore tramite DeePar+ o CNN-QR insieme ai metadati (ad esempio, un set di metadati degli articoli) per generare previsioni per il nuovo articolo.

La figura seguente mostra esempi di quattro diversi SKU così come appaiono in reali problemi di previsione di natura operativa reali.



Esempi di quattro diversi SKU così come appaiono in reali problemi di previsione di natura operativa.

Nell'immagine precedente i valori effettivi osservati sono a sinistra della linea verticale e le previsioni in blu sono a destra della linea verticale, confrontate con gli effettivi valori in nero. La cronologia di ogni singolo SKU, a sinistra della linea verticale, non è indicativa della sua evoluzione a destra della linea verde.

Implementazione di Forecast nell'ambiente di produzione

Al termine del flusso di lavoro completo di Amazon Forecast, è fondamentale identificare le differenze principali tra le API `Create_Forecast` e `Create_Predictor` e quando usare ciascuna.

La prima viene usata principalmente in fase di proof of concept per valutare la precisione e le metriche del modello, mentre la seconda viene usata per generare previsioni in un ambiente di produzione.

Nell'ambiente di produzione l'API `Create_Predictor` non deve essere eseguita ogni volta che è necessario generare una previsione, ma solo quando il modello deve essere riaddestrato a causa di variazioni nei dati o in base a una cadenza prestabilita, ad esempio bisettimanale o mensile. Poiché i set di dati vengono aggiornati con nuovi dati, è necessario eseguire solo l'API `Create_Forecast` per generare previsioni per il nuovo orizzonte di previsione.

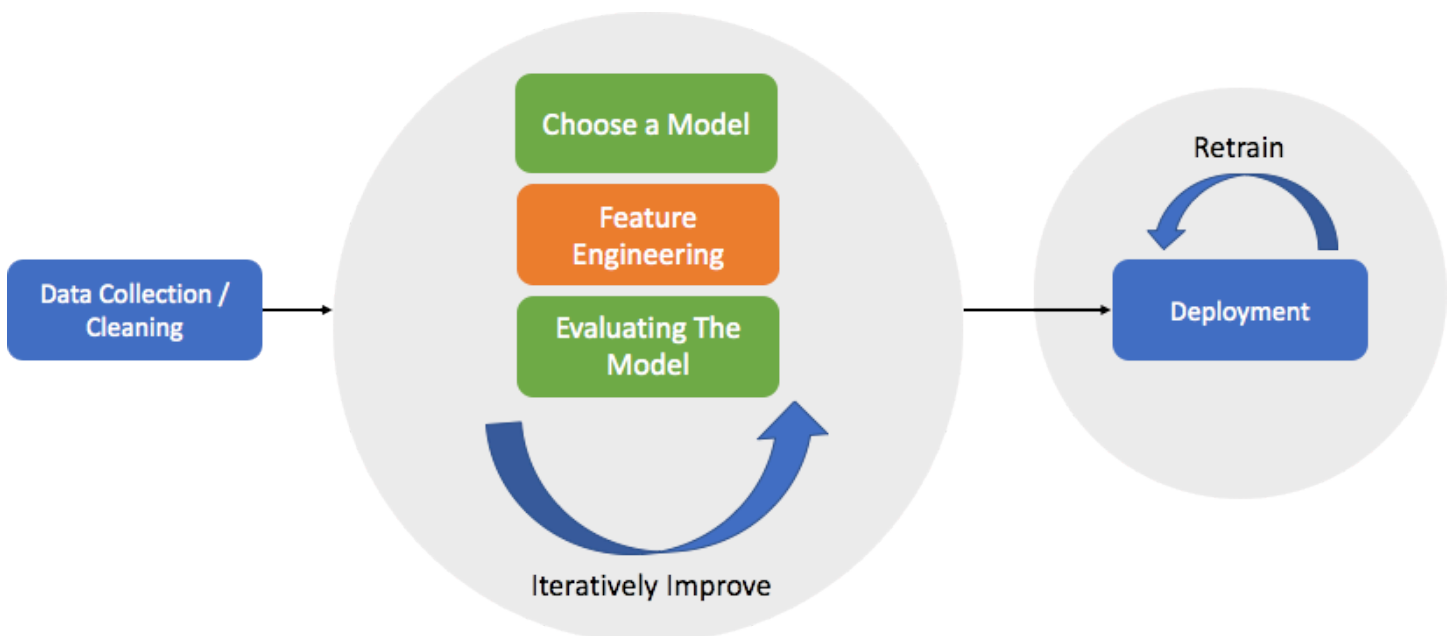
Nell'ambiente di produzione devi anche automatizzare le importazioni dei set di dati e le operazioni di previsione per poter generare nuove previsioni su base continuativa. A questo scopo, oggi è possibile configurare processi Cron usando una combinazione di log [Eventi Amazon CloudWatch](#), [AWS Step Functions](#) e funzioni [AWS Lambda](#). La configurazione di processi [Cron](#) a sua volta automatizza le chiamate API Amazon Forecast per l'importazione e/o il riaddestramento o la generazione di previsioni. Infine, è fondamentale gestire le risorse ed eliminarle a intervalli regolari in modo da non superare i [limiti di sistema](#) prescritti dal servizio. Per ulteriori informazioni sulla configurazione di processi pianificati, consulta questo [post di blog](#) e la pagina [Amazon Redshift](#).

Conclusione

[Januschowski e Kolassa \(2019\)](#) forniscono una classificazione dei problemi di previsione in linea con le decisioni che devono essere prese dalle aziende, tra cui le decisioni strategiche, tattiche e operative. Per ogni livello decisionale esistono attività di previsione corrispondenti.

I problemi di previsione di natura operativa e tattica sono caratterizzati dalla presenza di grandi quantità di dati e richiedono in genere un elevato livello di automazione. Diversi metodi di previsione permettono di risolvere questi problemi. I metodi di previsione locali sono in genere appropriati per i problemi di previsione di natura strategica, i metodi basati sul deep learning sono appropriati per problemi di natura operativa, mentre per tutte le situazioni intermedie può essere necessaria una certa sperimentazione. Benché questo articolo abbia affrontato i problemi di previsione di natura operativa, Amazon Forecast non è dogmatico riguardo ai modelli che offre e ne include di utili anche per problemi strategici e tattici.

Il processo di risoluzione dei problemi operativi può essere suddiviso in fasi di base, a partire dalla raccolta e preparazione dei dati fino alla creazione e all'implementazione del modello. In generale, è più utile considerarlo un processo iterativo anziché lineare. Ad esempio, man mano che i modelli e i casi d'uso vengono compresi meglio, può avere senso tornare indietro alla fase di raccolta. Anche lo sviluppo del modello è di per sé un processo altamente iterativo.



Processo di sviluppo semplificato per l'implementazione di un modello di previsione nell'ambiente di produzione.

Collaboratori

I collaboratori di questo documento includono:

- Yuyang Wang, Senior Machine Learning Scientist, AI Vertical Services
- Danielle Robinson, Applied Scientist, AI Vertical Services
- Tim Januschowski, responsabile, ML Applied Science
- Namita Das, Senior Product Manager, AI Vertical Services
- Christy Bergman, Senior AI/ML Specialist Solutions Architect
- Kris Tonthat, redattore tecnico, documentazione su intelligenza artificiale/machine learning

Approfondimenti

Per ulteriori informazioni sulla previsione di serie temporali e sui metodi di deep learning, consulta:

- [Documentazione di Amazon Forecast](#)
- [Blog sulla disponibilità generale di Amazon Forecast](#)
- [Ora disponibile in Amazon SageMaker: algoritmo DeepAR per una maggiore precisione](#)
- [Amazon SageMaker DeepAR supporta ora valori mancanti, caratteristiche categoriche e di serie temporali e frequenze generalizzate](#)
- [Amazon Forecast può ora usare reti neurali convoluzionali \(CNN\) per addestrare i modelli di previsione fino a 2 volte più rapidamente con precisione fino al 30% maggiore](#)
- [Amazon Forecast supporta ora misurazioni di precisione per singoli articoli](#)
- [Misurazione della precisione del modello di previsione per ottimizzare gli obiettivi aziendali con Amazon Forecast](#)
- [Amazon Forecast Weather Index: include automaticamente informazioni sul meteo locale per aumentare la precisione del modello di previsione](#)
- [Articoli scientifici sui modelli di previsione di serie temporali](#)
- [Pagina GitHub di esempi di Amazon Forecast](#)
- [Centro di architettura AWS](#)

Appendice A: Domande frequenti

D.: Come inizio a usare Amazon Forecast?

1. Per prima cosa, è necessario un Account AWS.
2. Apri quindi il servizio Forecast nella [AWS Management Console](#), crea un gruppo di set di dati e importa un file .csv nel set di dati della serie temporale di destinazione (obbligatorio). I dati minimi necessari per iniziare sono dati cronologici per la quantità che vuoi prevedere, ad esempio l'elettricità per timestamp per nucleo familiare.
3. Crea infine un modello eseguendo [CreatePredictor](#) e genera i risultati eseguendo [CreateForecast](#). Per ulteriori informazioni, consulta la pagina della [documentazione introduttiva](#).

Consulta anche la pagina [Amazon Forecast Introduction, Best Practices, and Cheat Sheet Tutorial](#).

D.: Amazon Forecast è una soluzione adatta per il mio caso d'uso?

Non tutti i problemi correlati al machine learning sono problemi di previsione. La prima domanda da porsi è se il problema aziendale includa una o più serie temporali nella sua definizione. Ad esempio, devi prevedere un determinato valore solo in una data e un'ora specifiche nel futuro? Le previsioni non sono la scelta migliore per problemi statici generali, in cui la data e l'ora specifiche non sono importanti, ad esempio per il rilevamento di frodi o il suggerimento di titoli di film agli utenti. Esistono soluzioni molto più rapide per i problemi statici.

Oltre alla presenza di dati di serie temporale, i dati stessi devono essere "densi" e con cronologie estese. La tabella seguente fornisce un riepilogo:

Tabella 2 – Criteri e classi di algoritmi Amazon Forecast

Criteri	Classe di algoritmi Amazon Forecast
Set di dati di grandi dimensioni con fino a cinque milioni di serie temporali con modelli sottostanti simili + effetti stagionali + dati correlati. Ogni serie temporale deve essere caratterizzata da una cronologia estesa, in teoria più di due anni per acquisire eventi	Algoritmi di deep learning Amazon Forecast proprietari DeepAr+ e CNN-QR

Criteri	Classe di algoritmi Amazon Forecast
annuali, e includere più di 300 punti dati, idealmente almeno 1000.	
Piccolo set di dati da 1 a più centinaia di serie temporali, di cui la maggior parte include 300 punti dati + effetti stagionali + dati correlati.	Prophet
Piccolo set di dati da 1 a più decine di serie temporali, di cui la maggior parte include più di 300 punti dati + effetti stagionali.	ETS, ARIMA
Set di dati intermittente (di tipo sparse, contenenti molti 0) da 1 a più decine di serie temporali, di cui la maggior parte include più di 300 punti dati.	Algoritmo Amazon Forecast proprietario NPTS
Piccolo set di dati (di tipo normale o sparse) da 1 a più decine di serie temporali, di cui la maggior parte include meno di 300 punti dati.	I dati sono insufficienti per Amazon Forecast. Prova invece ETS in Excel o i modelli statistici tradizionali ARIMA e Prophet.

La soluzione migliore è fare pratica con la modalità AutoML nel predittore, la prima volta sui tuoi dati. AutoML proverà automaticamente tutti gli algoritmi (gli algoritmi di deep learning vengono eseguiti con ottimizzazione degli iperparametri attivata), per identificare quello che funziona meglio sui dati.

D.: Come devo gestire i dati mancanti? Quando diventano troppi per generare previsioni sensate?

È possibile che vi siano problemi di registrazione dei dati o che il livello di aggregazione dei dati sia troppo basso o troppo alto. La regola generale è che la lunghezza della previsione non può essere maggiore di 1/3 dei dati di addestramento.

Oltre alla quantità di dati mancanti, un aspetto di cui tenere conto è l'attribuzione dei dati mancanti. Puoi convertire tutti gli 0 in valori null e lasciare che Amazon Forecast si occupi dell'attribuzione automatica dei valori mancanti. Amazon Forecast rileverà automaticamente se i valori mancanti siano dovuti all'introduzione di nuovi prodotti (chiamate a freddo) o a prodotti a fine vita. Puoi usare diversi tipi di logica per i valori mancanti: valore, mediano, minimo, massimo, zero, media e NaN (solo per serie temporali di destinazione). Consulta la [documentazione per la sintassi di riempimento di valori null](#).

- "Riempimento in avanti" (solo TTS): fa riferimento ad articoli nuovi o con chiamata a freddo e al modo in cui gestisci i valori null prima che l'articolo inizi ad avere una cronologia
- "Riempimento centrale": si riferisce a valori null a metà dei valori delle serie temporali
- "Riempimento all'indietro": si riferisce ad articoli a fine vita e al modo in cui gestisci i valori null dopo che un articolo non viene più venduto
- "riempimento futuro" (solo RTS): si riferisce ai valori null restituiti dopo la fine dei dati di addestramento

D.: I miei dati cronologici di input non includono valori negativi, ma trovo valori negativi nelle previsioni della domanda. Qual è il motivo? Come posso evitarlo?

Per tutti i modelli diversi da NPTS (addestrato su dati non negativi) e DeepAR (con funzione di probabilità binomiale negativa), la generazione di numeri positivi non è garantita. La soluzione consiste nel passare a uno dei modelli indicati sopra oppure troncare i valori di previsione in valori non negativi.

D.: Perché le metriche di precisione differiscono a quantili diversi? L'errore non dovrebbe essere identico se il modello è lo stesso?

Per ulteriori spiegazioni sulla dipendenza della ponderazione dal quantile, consulta la pagina relativa alla [metrica wQL](#).

Immagina di ottenere tutte le previsioni a tre diversi quantili: p10, p50, p90. Le tre previsioni stesse sono variabili casuali. Le precisioni vengono calcolate separatamente tra valori effettivi e previsioni a ogni livello di quantile. Potresti ottenere una tabella di "wQL" (weighted Quantile Loss, perdita quantilica ponderata), come quella mostrata di seguito. I valori wQL non hanno alcuna relazione deterministica gli uni con gli altri. Poiché la perdita di recupero significa errore, non è ordinata, mentre le previsioni quantiliche sono ordinate. Di conseguenza, la metrica wQL p90 non deve necessariamente essere maggiore, ad esempio, di wQL p50.

Tabella 3 – Esempi di quantili previsti

	A	B	C
1	wQL p10	wQL p50	wQL p90
2	0,18647	0,50879	0,30428

D.: Come posso migliorare la precisione delle previsioni?

La precisione delle previsioni dipende dalla disponibilità di dati corretti al giusto grado di quantità e qualità. Se la precisione non è soddisfacente, può avere senso identificare quanto sia prevedibile il problema (o quanto siano casuali/rumorosi/stazionari i dati). Tra gli altri fattori di cui tenere conto vi sono la valutazione di modelli diversi, le impostazioni degli iperparametri e l'integrazione di caratteristiche aggiuntive usando le serie temporali e i set di metadati degli articoli correlati. Per suggerimenti specifici, consulta [questo documento sulle best practice su GitHub](#).

D.: Il mio algoritmo funziona molto bene per il mio caso d'uso, ma non è disponibile in Amazon Forecast. Che cosa devo fare?

Il team di Amazon Forecast sarà lieto di aiutarti per questo caso d'uso. Contatta il team di assistenza di Amazon Forecast all'indirizzo e-mail <amazonforecast-poc@amazon.com>.

Appendice B: Riferimenti

[Tim Januschowski e Stephan Kolassa. A classification of business forecasting problems. Foresight: The International Journal of Applied Forecasting. 2019](#)

[David Salinas, Valentin Flunkert, Jan Gasthaus e Tim Januschowski. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting. 2019](#)

[Jan Gasthaus, Konstantinos Benidis, Yuyang Wang, Syama Sundar Rangapuram, David Salinas, Valentin Flunkert e Tim Januschowski. Probabilistic Forecasting with Spline Quantile Function RNNs. International Conference on Artificial Intelligence and Statistics, 22a edizione. 2019](#)

[Tim Januschowski, Jan Gasthaus, Yuyang Wang, David Salinas, Valentin Flunkert, Michael Bohlke-Schneider e Laurent Callot. Criteria for classifying forecasting methods. International Journal of Forecasting. 2019](#) (account di accesso necessario)

[Tim Januschowski, Jan Gasthaus, Yuyang Wang, Syama Sundar Rangapuram e Laurent Callot. Deep Learning for Forecasting. Foresight: The International Journal of Applied Forecasting. 2018](#)

[Tim Januschowski, Jan Gasthaus, Yuyang Wang, Syama Sundar Rangapuram e Laurent Callot. Deep Learning for Forecasting: Current Trends and Challenges. Foresight: The International Journal of Applied Forecasting. 2018](#)

[Joos-Hendrik Bose, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, Dustin Lange, David Salinas, Sebastian Schelter, Matthias Seeger e Yuyang Wang. Probabilistic demand forecasting at scale. Atti di VLDB Endowment. 2017](#)

Cronologia dei documenti

Per ricevere una notifica sugli aggiornamenti di questo Whitepaper, iscriviti al feed RSS.

Modifica	Descrizione	Data
Aggiornamento del whitepaper	Aggiornamenti.	1 settembre 2021
Pubblicazione iniziale	Prima pubblicazione del whitepaper	4 febbraio 2020

Note

Per iscriverti e ricevere aggiornamenti RSS, devi aver abilitato un plug-in RSS per il browser in uso.

Avvisi

I clienti sono responsabili della propria valutazione autonoma delle informazioni contenute in questo documento. Questo documento: (a) è solo a scopo informativo, (b) presenta le attuali procedure e offerte di prodotti AWS, soggette a modifiche senza preavviso, e (c) non implica alcun impegno o garanzia da parte di AWS e dei suoi affiliati, fornitori o licenziatari. I prodotti o servizi AWS vengono forniti "così come sono", senza garanzie, rappresentazioni o condizioni di alcun tipo, espresse o implicite. Le responsabilità e gli obblighi di AWS verso i propri clienti sono disciplinati dagli accordi AWS e questo documento non fa parte di né modifica alcun accordo tra AWS e i suoi clienti.

© 2023, Amazon Web Services, Inc., o sue affiliate. Tutti i diritti riservati.

Glossario di AWS

Per la terminologia di AWS più recente, consulta [Glossario di AWS](#) in Riferimenti generali di AWS.