

AWS Documento técnico

Principios de la previsión de serie temporal con Amazon Forecast



Principios de la previsión de serie temporal con Amazon Forecast: AWS Documento técnico

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

Las marcas comerciales y la imagen comercial de Amazon no se pueden utilizar en relación con ningún producto o servicio que no sea de Amazon, de ninguna manera que pueda causar confusión entre los clientes y de ninguna manera que menosprecie o desacredite a Amazon. Todas las demás marcas comerciales que no son propiedad de Amazon son propiedad de sus respectivos propietarios, que pueden o no estar afiliados, conectados o patrocinados por Amazon.

Table of Contents

Resumen y descripción general	i
Información general	1
¿Posee una buena arquitectura?	2
Acerca de las previsiones	3
Sistema de previsión	3
¿Dónde se producen los problemas de previsión?	3
Consideraciones antes de intentar resolver un problema de previsión	4
Caso práctico: problema de previsión de la demanda minorista para una empresa de comercio electrónico	6
Paso 1: Recopilar y agregar datos	9
Ejemplo	11
Paso 2: Preparar los datos	12
Cómo gestionar la ausencia de datos	12
Ejemplo 1	12
Ejemplo 2	15
Conceptos de caracterización y series temporales relacionadas	15
Ejemplo 3	16
Paso 3: Crear un predictor	19
Paso 4: Evaluar los predictores	21
Backtesting	21
Cuantiles de predicción y métricas de precisión	23
Pérdida cuantil ponderada (wQL)	23
Error porcentual absoluto ponderado (WAPE)	24
Raíz del error cuadrático medio (RMSE)	25
Problemas con WAPE y RMSE	25
Paso 5: Generar y utilizar previsiones para la toma de decisiones	27
Previsiones probabilísticas	27
Visualización	28
Resumen del flujo de trabajo de previsión y las API	30
Uso de Amazon Forecast para escenarios comunes	31
Implementación de previsiones en producción	32
Conclusión	34
Colaboradores	35
Documentación adicional	36

Apéndice A: Preguntas frecuentes	37
Apéndice B: Referencias	41
Historial de revisión	42
Avisos	43
Glosario de AWS	44

Principios de la previsión de serie temporal con Amazon Forecast

Fecha de publicación: 1 de septiembre de 2021 ([Historial de revisión](#))

Actualmente, las compañías usan desde simples hojas de cálculo a software de planificación financiera complejo para intentar estimar con precisión resultados empresariales futuros, como la demanda de productos, las necesidades vinculadas con recursos y el rendimiento financiero. Este documento presenta la previsión, su terminología, los desafíos y los casos prácticos. Este documento utiliza un caso práctico para reforzar los conceptos y los pasos de previsión y hace referencia a cómo [Amazon Forecast](#) puede ayudar a resolver los numerosos desafíos prácticos que plantean los problemas de previsión en el mundo real.

Información general

La predicción es la ciencia de predecir el futuro. Al utilizar datos históricos, las empresas pueden comprender las tendencias, determinar qué podría suceder y cuándo y, a su vez, incorporar esa información a sus planes futuros para todo tipo de cuestiones, desde la demanda de productos hasta la planificación del inventario y la dotación de personal.

Dadas las consecuencias de la previsión, es importante que sea precisa. Si la previsión resulta excesiva, los clientes pueden invertir demasiado en mercancía y personal, lo que puede acabar generando desperdicios en la inversión. Si la previsión es demasiado baja, es posible que los clientes no inviertan lo suficiente, lo que se podría traducir en una carencia de material y en un inventario deficiente y generaría una mala experiencia para el cliente.

Hoy en día, las empresas intentan utilizar desde simples hojas de cálculo hasta complejos programas de planificación financiera y de la demanda para generar previsiones, pero sigue siendo difícil alcanzar un alto nivel de precisión por dos motivos:

- En primer lugar, las previsiones tradicionales tienen dificultades para incorporar grandes volúmenes de datos históricos, ya que pasan por alto señales importantes del pasado que se pierden en el ruido.
- En segundo lugar, las previsiones tradicionales no suelen incorporar datos relacionados pero independientes que pueden ofrecer un contexto importante (como precios, periodos festivos o eventos, desabastecimiento, promociones de marketing, etc.). Si no cuentan con el historial

completo y un contexto más amplio, la mayoría de las previsiones no pueden predecir el futuro con precisión.

[Amazon Forecast](#) es un servicio totalmente administrado que permite superar estos problemas. Amazon Forecast proporciona los mejores algoritmos para el escenario de previsión en cuestión. Se basa en el machine learning (ML) moderno y en el aprendizaje profundo, cuando es apropiado, para ofrecer previsiones muy precisas. Amazon Forecast es fácil de usar y no requiere experiencia en machine learning. El servicio proporciona automáticamente la infraestructura necesaria, procesa los datos y crea modelos de ML personalizados/privados que se alojan en AWS y están listos para realizar predicciones. Además, a medida que los avances en las técnicas de machine learning siguen evolucionando a un ritmo acelerado, Amazon Forecast los incorpora para que los clientes sigan percibiendo mejoras en la precisión sin esfuerzos adicionales (o con un esfuerzo mínimo) por su parte.

¿Posee una buena arquitectura?

[AWS Well-Architected Framework](#) le permite comprender las ventajas y desventajas de las decisiones que tome durante la creación de sistemas en la nube. Los seis pilares de este marco permiten aprender las prácticas recomendadas arquitectónicas para diseñar y explotar sistemas confiables, seguros, eficientes, rentables y sostenibles. Con [AWS Well-Architected Tool](#), disponible gratuitamente en [AWS Management Console](#), puede revisar sus cargas de trabajo con respecto a estas prácticas recomendadas al responder a un conjunto de preguntas para cada pilar.

En [Machine Learning Lens](#) nos centramos en cómo diseñar, implementar y diseñar sus cargas de trabajo de machine learning en la Nube de AWS. Esta Lens se incorpora a las prácticas recomendadas descritas en el Well-Architected Framework.

Para obtener más orientación de expertos y prácticas recomendadas para su arquitectura de nube (despliegues de arquitectura de referencia, diagramas y documentos técnicos), consulte el [Centro de arquitectura de AWS](#).

Acerca de las previsiones

En este documento, las previsiones implican una predicción de los valores futuros de una serie temporal: los valores de entrada y salida de un problema tienen una naturaleza temporal.

Sistema de previsión

Un sistema de previsión incluye un conjunto de usuarios diverso:

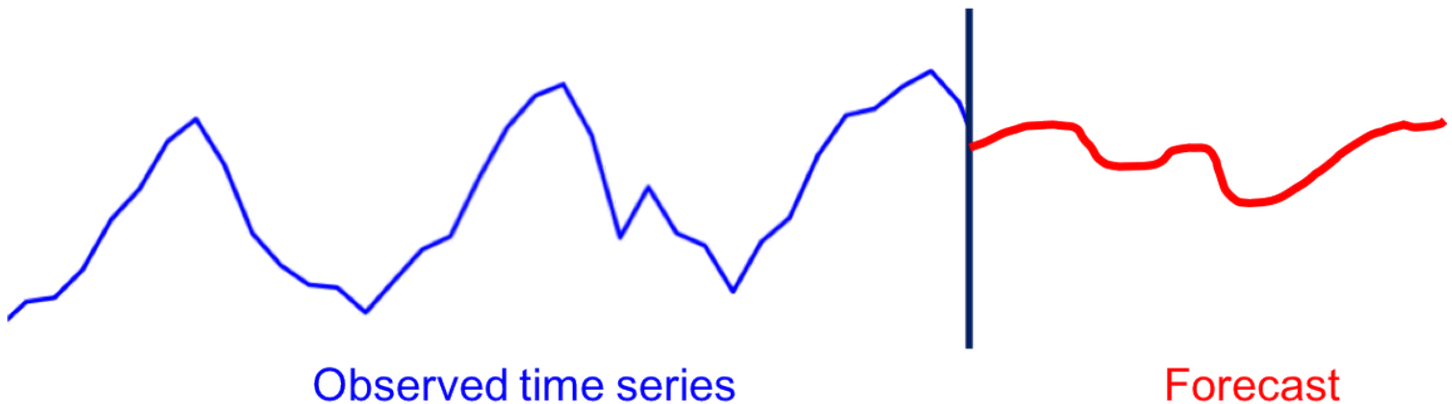
- Los usuarios finales, que consultan la previsión de un producto específico y deciden cuántas unidades comprar; pueden tratarse de una persona o un sistema automatizado.
- Los analistas de negocios/inteligencia empresarial, que ayudan a los usuarios finales, ejecutan y organizan informes agregados.
- Los científicos de datos, que analizan de forma iterativa los patrones de demanda, los efectos causales y añaden nuevas funciones para proponer mejoras graduales al modelo o mejorar el modelo de previsión.
- Los ingenieros, que configuran la infraestructura de recopilación de datos y garantizan la disponibilidad de los datos de entrada en el sistema.

Amazon Forecast facilita el trabajo de los ingenieros de software y permite a las empresas con capacidades de ciencia de datos limitadas aprovechar la tecnología de previsión más avanzada. Para las empresas con capacidades de ciencia de datos, se incluyen una serie de funciones de diagnóstico para que Amazon Forecast aborde correctamente los problemas de previsión.

¿Dónde se producen los problemas de previsión?

Los problemas de previsión se producen en muchas de las áreas que producen datos de series temporales de forma natural. Entre ellos se incluyen las ventas minoristas, el análisis médico, la planificación de la capacidad, la supervisión de redes de sensores, el análisis financiero, la minería de actividades en redes sociales y los sistemas de bases de datos. Por ejemplo, la previsión desempeña un papel clave en la automatización y optimización de los procesos operativos en la mayoría de las empresas, lo que permite la toma de decisiones basada en datos. Las previsiones de la oferta y la demanda de productos se pueden utilizar para una gestión óptima del inventario, la programación del personal y la planificación de la topología y, en términos más generales, suponen una tecnología crucial para la mayoría de los aspectos de la optimización de la cadena de suministro.

La siguiente figura contiene un resumen de los problemas que puede plantear la previsión cuando se basa en una serie temporal observada que muestra un patrón (en este ejemplo, la estacionalidad) y se crea una previsión para un período específico. El eje horizontal representa el tiempo, y avanza desde el pasado (izquierda) hacia el futuro (derecha). El eje vertical representa las unidades medidas. Teniendo en cuenta los datos del pasado (en azul) hasta la línea negra vertical, la tarea de previsión consiste en identificar los datos del futuro (en rojo).



Descripción general de la tarea de previsión

Consideraciones antes de intentar resolver un problema de previsión

Las preguntas más importantes que hay que entender antes de resolver problemas de previsión son:

- ¿Es necesario resolver un problema de previsión?
- ¿Por qué se quiere resolver el problema de previsión?

Como consecuencia de la ubicuidad de los datos de las series temporales, es fácil encontrar problemas de previsión en todas partes. Sin embargo, una pregunta clave es plantearse si realmente es necesario resolver un problema de previsión o si puede evitarse por completo sin sacrificar la eficiencia en la toma de decisiones de la empresa. Plantear esta pregunta es importante porque, desde el punto de vista científico, la predicción es uno de los problemas más difíciles para el ámbito del machine learning.

Por ejemplo, piense en las recomendaciones de productos para un vendedor minorista en línea. La recomendación de productos puede enmarcarse como un problema de previsión en el que, para cada par de unidades de mantenimiento de existencias (SKU) del cliente, se prevé la cantidad de unidades de un producto específico que podría comprar este cliente en particular. La formulación de

este problema tiene una serie de ventajas. Una de ellas es que el componente temporal se tiene en cuenta de forma explícita, por lo que es posible recomendar productos en función de los patrones de compra de los clientes.

Sin embargo, los problemas de recomendación de productos rara vez se formulan como un problema de previsión, ya que resolver este problema de previsión es mucho más difícil (debido, por ejemplo, a la escasez de información en el nivel SKU-cliente y a la magnitud del problema) que resolver directamente el problema de la recomendación. Por lo tanto, al pensar en una aplicación de previsión, es importante tener en cuenta el uso que se le va a dar posteriormente a la previsión y si es posible abordar este problema mediante un enfoque alternativo.

[Amazon Personalize](#) puede ser de ayuda en estos casos. Amazon Personalize es un servicio de machine learning que facilita a los desarrolladores la tarea de crear recomendaciones individualizadas para los clientes que utilicen sus aplicaciones.

Después de determinar que necesita resolver un problema de previsión, la siguiente pregunta que debe hacerse es: ¿por qué quiere resolver el problema de previsión? En muchos entornos empresariales, la previsión suele ser simplemente un medio para lograr un fin. Por ejemplo, para la previsión de la demanda en un contexto de venta minorista, la previsión se puede utilizar para tomar decisiones de gestión del inventario. El problema de previsión suele servir de punto de partida para un problema de decisión, que a su vez puede modelarse como un problema de optimización.

Algunos ejemplos de estos problemas de decisión pueden ser plantearse qué número de unidades se van a comprar o cuál es el mejor enfoque para gestionar el inventario existente. Otros problemas de previsión empresarial podrían ser la previsión de la capacidad del servidor o la previsión de la demanda de materias primas o piezas en un contexto de fabricación. Estas previsiones se pueden utilizar como puntos de partida para otros procesos, ya sea para problemas de decisión (como los anteriores) o para simulaciones de escenarios, que luego se utilizan para planificar sin modelos explícitos. Hay excepciones esta regla de que la previsión no es un fin en sí misma. En la previsión financiera, por ejemplo, la previsión se utiliza directamente para acumular reservas financieras o para presentársela a los inversores.

Para entender el propósito de la previsión, plantéese las siguientes preguntas:

- ¿Hasta qué punto en el futuro debería llegar la previsión?
- ¿Con qué frecuencia necesita generar previsiones?
- ¿Hay aspectos específicos de las previsiones en los que debería profundizar?

Caso práctico: problema de previsión de la demanda minorista para una empresa de comercio electrónico

Para ilustrar los conceptos de previsión con más detalle, plantéese el caso de una empresa de comercio electrónico que vende productos en línea. La optimización de las decisiones en la cadena de suministro (por ejemplo, la gestión de existencias) es fundamental para la competitividad básica de esta empresa, ya que ayuda a disponer de la cantidad exacta de productos en los centros logísticos correspondientes. Básicamente, esto significa tener una gran selección disponible con tiempos de envío más cortos y precios competitivos, lo que genera una mayor satisfacción del cliente. La entrada clave del sistema de software de la cadena de suministro es una predicción de la demanda o la previsión de las ventas potenciales de cada producto del catálogo. Esta previsión permite tomar decisiones posteriores importantes, entre las que destacan:

- Planificación a nivel macroeconómico (previsión estratégica): Para una empresa en su conjunto, ¿cuál es el crecimiento proyectado en términos de ventas e ingresos totales? ¿Dónde debería estar la empresa (más) activa geográficamente? ¿Cómo se debe dotar de personal a la empresa?
- Previsión de la demanda (o inventario): ¿Cuántas unidades de cada producto se espera que se vendan por ubicación?
- Actividad promocional (previsión táctica): ¿Cómo deben realizarse las promociones? ¿Deberían liquidarse los productos?

El resto del caso práctico se centra en el segundo problema, que forma parte de la familia de problemas de previsión operativa (Januschowski y Kolassa, 2019). Este documento aborda las principales preocupaciones: los datos, los modelos (predictores), las inferencias (previsiones) y el paso a producción.

Para este caso práctico, es importante tener en cuenta que el problema de previsión es un medio para lograr un fin. Si bien las previsiones son de vital importancia para la empresa, lo que es aún más importante son las decisiones posteriores en la cadena de suministro. En nuestro caso práctico, estas decisiones las toman sistemas de compra automatizados que se basan en modelos de optimización matemática de la investigación de operaciones. Estos sistemas intentan minimizar el coste esperado para la empresa.

La palabra clave es *esperado*, pues significa que las previsiones deben abarcar no solo un futuro posible sino todos los futuros posibles, con la ponderación adecuada en función de la probabilidad

de un resultado concreto. Con este fin, el factor clave para la toma de decisiones posteriores es una distribución completa de los valores de la previsión, en lugar de limitarse a tener una previsión puntual. La siguiente figura muestra una previsión probabilística (también denominada previsión de densidad). Tenga en cuenta que puede derivar fácilmente una previsión puntual (el futuro más probable) a partir de esta previsión probabilística, pero pasar de una previsión puntual a una previsión probabilística es más difícil.

Con una previsión probabilística dada, puede obtener diferentes estadísticas a partir de ella y adaptar los resultados para que le ayuden a tomar la decisión deseada. Es posible que el negocio del comercio electrónico tenga una serie de productos clave de los que siempre quiera tener existencias. En tal caso, utilice un cuantil alto (por ejemplo, el percentil 90), lo que se traduciría en que habría existencias del producto el 90 % del tiempo. Para otros productos, como los productos para los que es más fácil encontrar sustitutos (por ejemplo, lápices), puede ser más adecuado utilizar un percentil más bajo.

En Amazon Forecast puede obtener fácilmente diferentes cuantiles de la previsión probabilística.

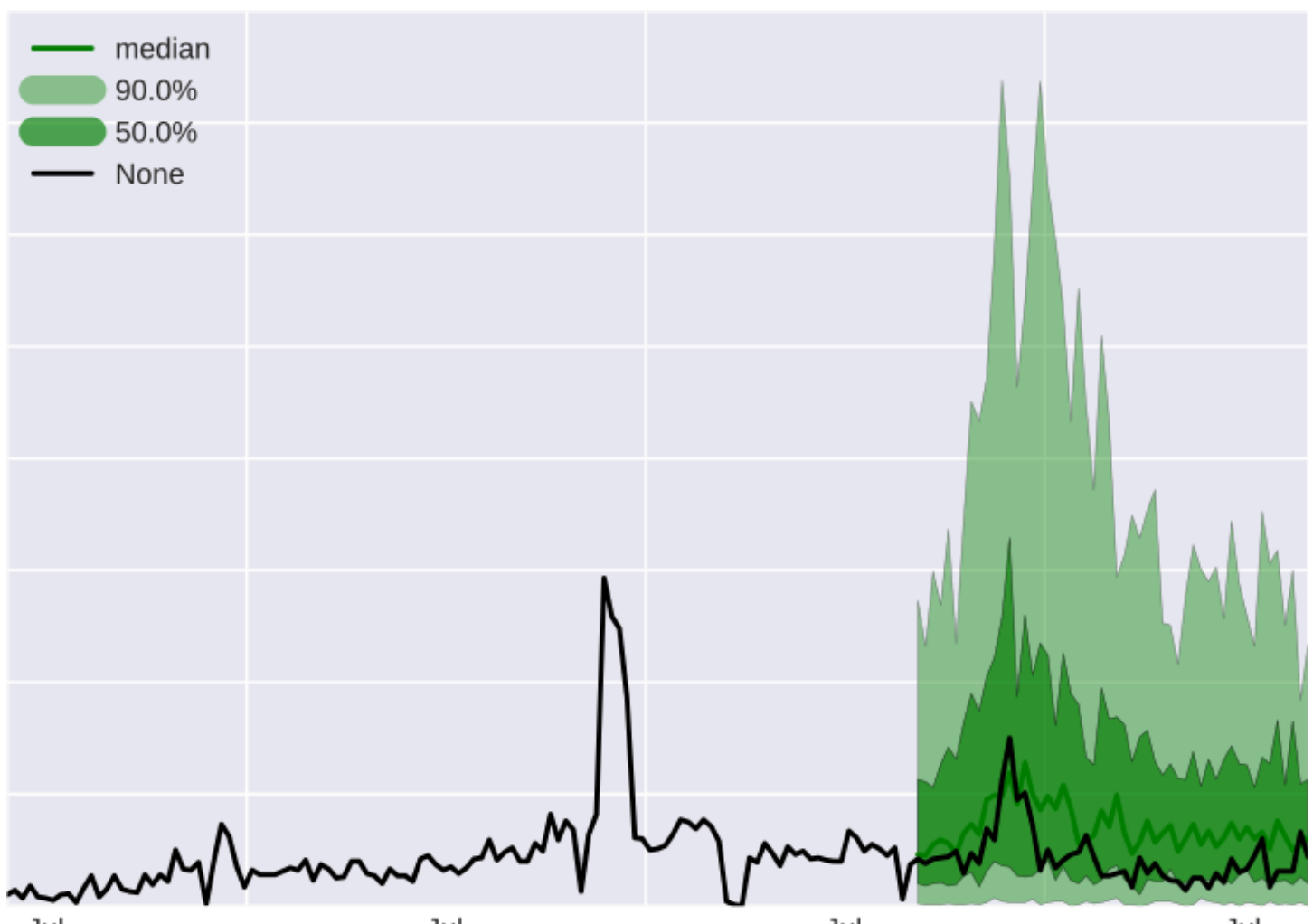


Ilustración de la previsión probabilística

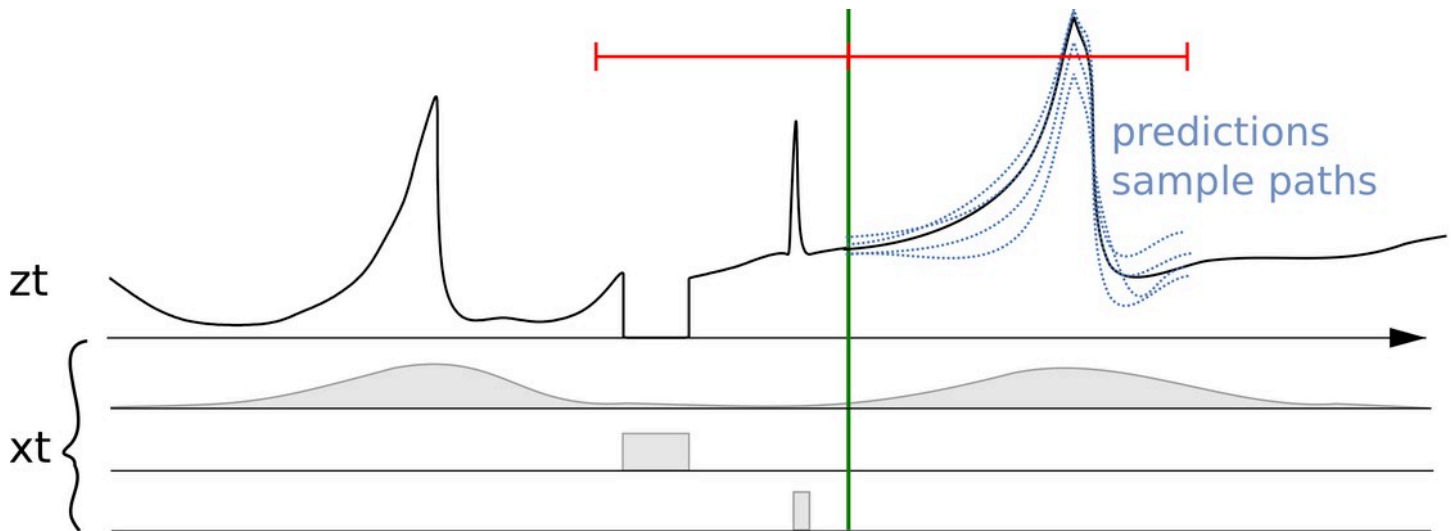
En la figura anterior, la línea negra son los valores reales; la línea verde oscura es la mediana de la distribución de previsión; el área sombreada en verde oscuro es el intervalo de predicción que se espera que comprenda el 50 % de los valores; y el área verde clara es el intervalo de predicción que se espera que comprenda el 90 % de los valores reales.

En las siguientes secciones se describen los pasos necesarios para resolver el problema de previsión de esta empresa, entre los que se incluyen:

- [Recopilación y agregación de datos \(paso 1\)](#)
- [Preparación de datos \(paso 2\)](#)
- [Creación de un predictor \(paso 3\)](#)
- [Evaluación de predictores \(paso 4\)](#)
- [Automatización de la generación de previsiones \(paso 5\)](#)

Paso 1: Recopilar y agregar datos

En la siguiente figura se muestra un modelo mental para el problema de previsión. El objetivo es prever la serie temporal z_t en el futuro, utilizando la mayor cantidad de información relevante posible para que la previsión sea lo más precisa posible. Por lo tanto, el primer paso, y el más importante, es recopilar la mayor cantidad posible de datos correctos.



Una serie temporal z_t junto con características o covariables asociadas (x_t) y múltiples previsiones

En la figura anterior, se muestran varias previsiones a la derecha de la línea vertical. Estas previsiones son muestras de la distribución de la previsión probabilística (o, a la inversa, se pueden utilizar para representar la previsión probabilística).

La información clave que debe registrar una empresa de venta minorista es:

- Los datos de ventas de las transacciones: por ejemplo, las unidades de mantenimiento de existencias (SKU), la ubicación, la marca temporal y las unidades vendidas.
- Datos de detalles del artículo de la SKU: los metadatos de un producto. Por ejemplo, el color, el departamento, el tamaño, etc.
- Datos de precios: la serie temporal de precios de cada producto con marcas temporales.
- Datos informativos sobre promociones: diferentes tipos de promociones, ya sea sobre un conjunto de productos (categoría) o sobre productos individuales con marcas temporales.
- Datos informativos sobre existencias: para cada unidad de tiempo, la información sobre si una SKU está en existencias o se puede adquirir o bien si esa SKU se ha agotado.

- **Datos de ubicación:** la ubicación de un producto o una venta en un momento determinado se puede representar como una cadena `location_id` o `store_id` , o bien como una geolocalización real. Las geolocalizaciones pueden ser un código de país más un código postal de cinco dígitos, o bien coordenadas `latitude_longitude`. La ubicación se considera una "dimensión" de las ventas transaccionales.

En [Amazon Forecast](#), los datos históricos de la cantidad que se va a prever se denominan series temporales objetivo (TTS, por sus siglas en inglés). Para las empresas minoristas, las TTS son los datos de ventas transaccionales. El resto de datos históricos, que se conocen exactamente al mismo tiempo que cada transacción de venta, se denominan series temporales relacionadas (RTS, por sus siglas en inglés). Para las empresas minoristas, las RTS incluirían variables de precio, promociones y existencias.

Tenga en cuenta que la información sobre las existencias es importante, ya que este problema se centra en la previsión de la demanda y no en las ventas, pero la empresa solo registra las ventas. Cuando una SKU se agota, el número de ventas es inferior a la demanda potencial, por lo que es importante saber y registrar cuándo se producen dichos eventos de agotamiento de existencias.

Otros conjuntos de datos que se deben tener en cuenta incluyen el número de visitas a la página web, los detalles sobre los términos de búsqueda, las redes sociales y la información meteorológica. A menudo es importante disponer de datos del pasado y del futuro para poder utilizarlos en los modelos. Este es un requisito de muchos modelos de previsión y de pruebas retrospectivas (esto se describe en la sección [Paso 4: Evaluar las previsiones](#)).

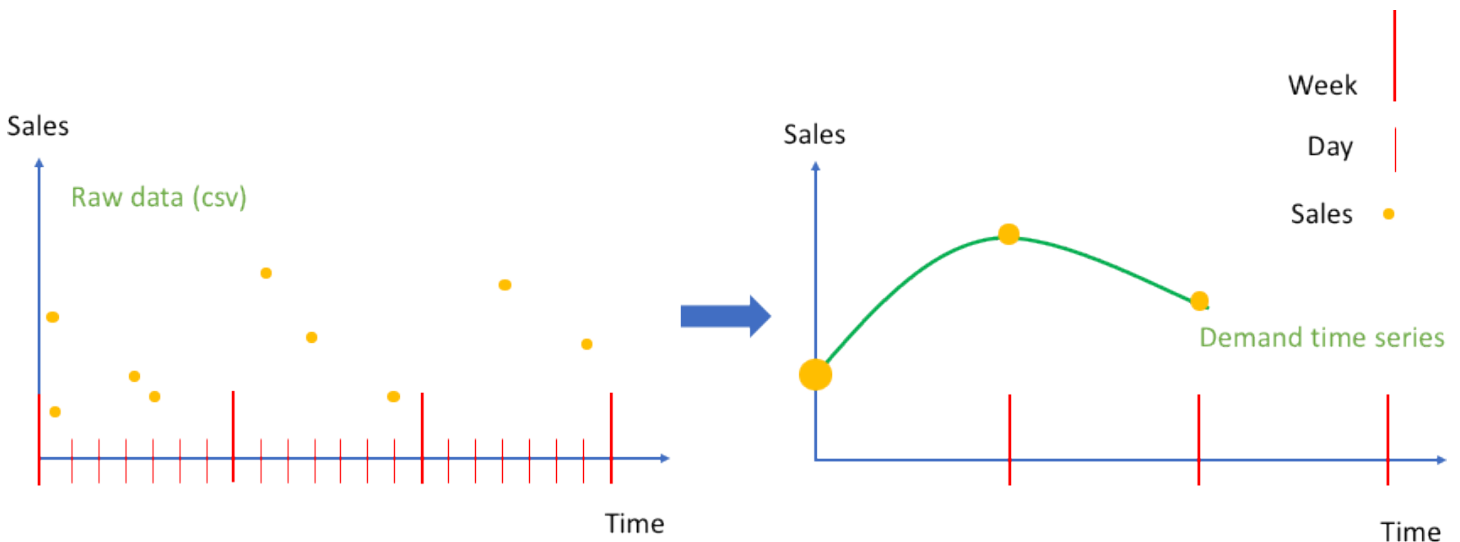
Para algunos problemas de previsión, la frecuencia de los datos sin procesar coincide naturalmente con la del problema de previsión. Entre algunos ejemplos se incluyen la solicitud del volumen del servidor, que se muestrea por minuto, cuando se desea prever con una frecuencia de minutos.

A menudo, los datos se registran con una frecuencia más precisa o simplemente en marcas temporales arbitrarias dentro de un rango de tiempo, pero el problema de la previsión pertenece a un nivel de granularidad menos específico. Esto es algo habitual en los casos prácticos de venta minorista, en los que los datos de ventas se registran normalmente como datos transaccionales; por ejemplo, el formato consiste en una marca de tiempo con una alta granularidad del momento en que se produjeron las ventas. En el caso práctico de la previsión, puede que este nivel de granularidad no sea necesario y podría ser más apropiado agregar estos datos en ventas totales por hora o por día. En este caso, el nivel de agregación debe adecuarse al problema posterior; por ejemplo, la gestión del inventario o la planificación de recursos.

Ejemplo

En la siguiente figura, el gráfico de la izquierda muestra un ejemplo de los datos de ventas de clientes sin procesar que se pueden introducir en Amazon Forecast como un archivo de valores separados por comas (CSV). En este ejemplo, los datos de ventas se definen en una cuadrícula horaria diaria más precisa y el problema consiste en prever la demanda semanal futura en una cuadrícula horaria más general. Amazon Forecast realiza la agregación de los valores diarios de una semana determinada en la llamada a la API `create_predictor`.

El resultado transforma los datos sin procesar en una colección de series temporales bien formadas con una frecuencia semanal fija. El gráfico de la derecha ilustra esta agregación en la serie temporal objetivo mediante el método de agregación por suma predeterminado. Otros métodos de agregación son por media, por máximo, por mínimo o por elección de un solo punto (por ejemplo, el primero). El nivel de granularidad y el método de agregación deben elegirse de manera que se adapten mejor al caso práctico empresarial de los datos. En este ejemplo, el valor agregado se alinea con la agregación semanal. El usuario puede configurar otros métodos de agregación mediante la clave `FeaturizationMethodParameters` del parámetro `FeaturizationConfig` de la API `create_predictor`.



Agregación de datos de ventas sin procesar como eventos (izquierda), en una serie temporal espaciada a intervalos regulares (derecha)

Paso 2: Preparar los datos

Una vez que disponga de datos sin procesar, deberá hacerse cargo de ciertas complicaciones, como la falta de datos, y asegurarse de preparar los datos para los modelos de previsión que mejor capten la interpretación prevista.

Cómo gestionar la ausencia de datos

En los problemas de previsión reales, suele ocurrir que falten ciertos valores en los datos sin procesar. Si falta un valor en una serie temporal, significa que no está disponible el valor real correspondiente en cada punto temporal con la frecuencia especificada para su posterior procesamiento. Puede haber varios motivos por los que se marquen los valores como faltantes.

Los valores faltantes pueden deberse a que no se ha realizado ninguna transacción o a posibles errores de medición (por ejemplo, porque un servicio que supervisaba ciertos datos no funcionaba correctamente o porque la medición no se pudo realizar correctamente). El principal ejemplo de esta última circunstancia en el caso práctico del comercio minorista es una situación de agotamiento de existencias en la previsión de la demanda, lo que significa que la demanda no coincide con las ventas de ese día.

Se pueden producir efectos similares en los escenarios de computación en la nube cuando un servicio ha alcanzado un límite (por ejemplo, si todas las instancias de [Amazon EC2](#) de un determinado servicio de [Región de AWS](#) están ocupadas). Otro ejemplo de valores faltantes se produce cuando un producto o servicio aún no se ha lanzado o deja de producirse.

Los componentes de procesamiento de funciones también pueden insertar valores faltantes para rellenar las series temporales y garantizar así que tengan longitudes uniformes. Si son lo suficientemente importantes, los valores faltantes pueden afectar de manera significativa a la precisión de un modelo.

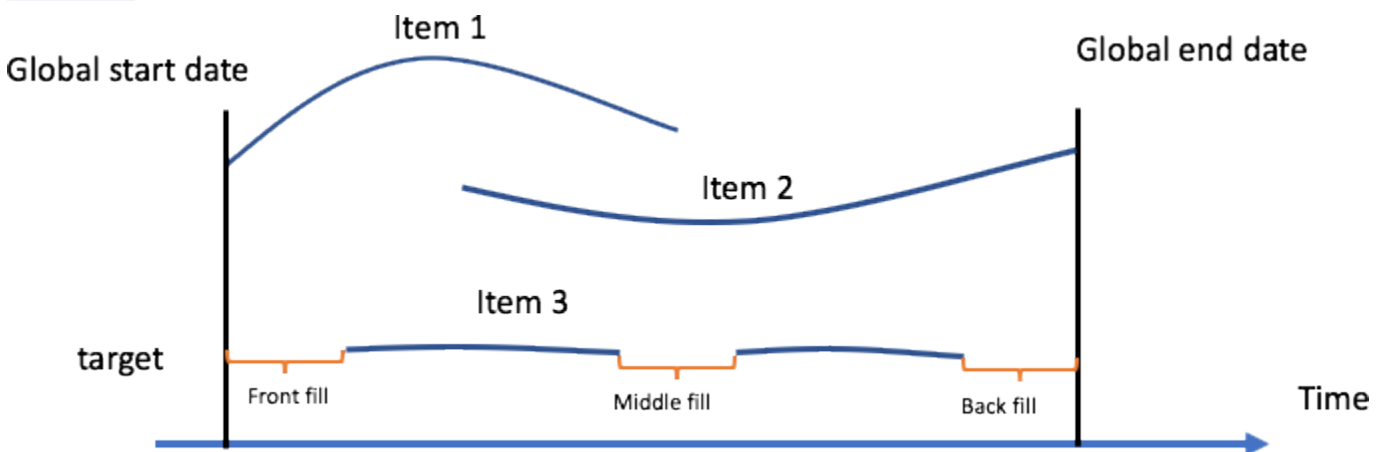
Ejemplo 1

El llenado es el proceso de añadir valores estandarizados a las entradas que faltan en su conjunto de datos. En la siguiente figura, se ilustran las diferentes estrategias para gestionar los valores faltantes en Amazon Forecast (rellenado frontal, medio, retroactivo y a futuro) para el elemento 2 de un conjunto de datos compuesto por tres elementos.

Amazon Forecast permite rellenar tanto la serie temporal objetivo (TTS) como la relacionada (RTS). La fecha de inicio global se define como la fecha de inicio más temprana de entre las fechas de inicio de todos los elementos del conjunto de datos. En el ejemplo siguiente, la fecha de inicio global corresponde a la del elemento 1. Del mismo modo, la fecha de finalización global se define como la última fecha de finalización de todos los elementos de la serie temporal, que en este caso se corresponde con la del elemento 2.

El rellenado frontal rellena todos los valores desde el inicio de esa serie temporal en particular hasta la fecha de inicio global. En el momento de publicar este documento, Amazon Forecast no realiza ningún rellenado frontal y permite que todas las series temporales comiencen en diferentes puntos temporales. El rellenado medio se refiere a la inserción de valores en la mitad de la serie temporal (por ejemplo, entre las fechas de inicio y finalización de los elementos), y el rellenado retroactivo rellena desde la última fecha de esa serie temporal hasta la fecha de finalización global.

Para la serie temporal objetivo, los métodos de rellenado medio y retroactivo tienen una lógica de rellenado predeterminada de cero. El rellenado a futuro (que solo se aplica a las series temporales relacionadas) rellena cualquier valor que falte entre la fecha de finalización global de los elementos y el horizonte de previsión especificado por el cliente. Se requieren valores futuros para usar el conjunto de datos de series temporales relacionadas con [Prophet](#) y [DeepAR+](#), y son opcionales para [CNN-QR](#).



Estrategias de gestión de valores faltantes en Amazon Forecast

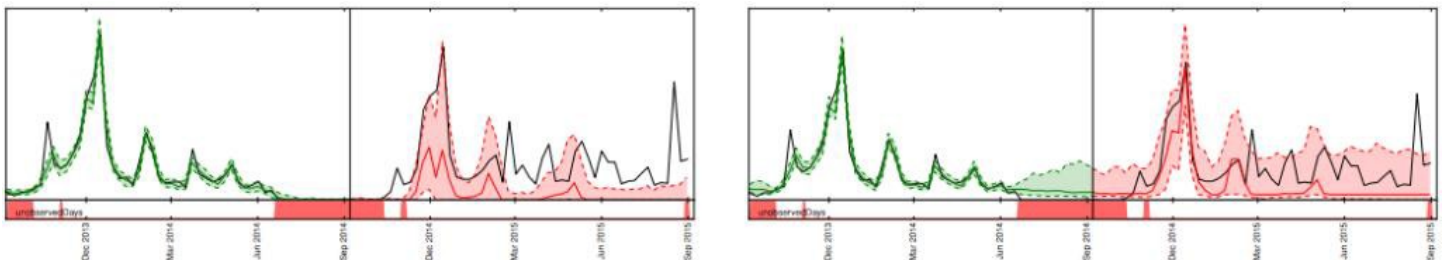
En la figura anterior, la fecha de inicio global indica la fecha de inicio más temprana de entre las fechas de inicio de todos los elementos, y la fecha de finalización global indica la fecha de finalización más tardía de entre las fechas de finalización de todos los elementos. El horizonte de previsión es el período durante el cual Forecast proporciona predicciones para el valor objetivo.

Este es un escenario común en el estudio de casos prácticos de venta minorista, y representa unas ventas de cero para los datos transaccionales de los productos disponibles. Estos valores se tratan como ceros verdaderos y se utilizan en el componente de evaluación de métricas. Amazon Forecast permite al usuario identificar los valores que faltan realmente y codificarlos como valores no numéricos (NaN) para que los procesen los algoritmos. En este siguiente artículo se examina en qué difieren estos dos casos y cuándo resulta útil cada uno de ellos.

En el caso práctico de venta minorista, la información según la cual un minorista vendió cero unidades de un artículo disponible difiere de la información de que se han vendido cero unidades de un artículo no disponible, ya sea en períodos ajenos a su existencia (por ejemplo, antes de su lanzamiento o después de su retirada de la venta) o durante períodos correspondientes a su existencia (por ejemplo, si se ha agotado parcialmente o si no se han registrado datos de ventas para este intervalo de tiempo). En el primer caso se aplica el rellenado predeterminado de cero. En el segundo caso, aunque el valor objetivo correspondiente suele ser cero, se transmite información adicional en el valor que se marca como faltante. Lo recomendable es conservar la información de que faltaban datos y no descartarla. Consulte el siguiente ejemplo para ver por qué es importante conservar la información.

Amazon Forecast es compatible con una lógica de rellenado adicional de valor, media, mediana, mínimo y máximo. Para las series temporales relacionadas (por ejemplo, precios o promociones), no se especifican valores predeterminados para los métodos de rellenado medio, retroactivo o a futuro, ya que la lógica correcta para los valores faltantes varía según el tipo de atributo y el caso práctico. La lógica de rellenado admitida para las series temporales relacionadas incluye cero, valor, media, mediana, mínimo y máximo.

Para realizar el rellenado de los valores faltantes, especifique los tipos de rellenado que quiera implementar al llamar a la operación [CreatePredictor](#). La lógica de rellenado se especifica en los objetos [FeaturizationMethod](#). Por ejemplo, para codificar un valor que no represente ventas cero de un producto no disponible en la serie temporal objetivo, marque un valor como verdaderamente faltante; para ello, establezca el tipo de rellenado como NaN. A diferencia del rellenado cero, los valores codificados como NaN se consideran verdaderamente faltantes y no se utilizan en el componente de evaluación de métricas



El efecto del relleno cero en comparación con el relleno con NaN en las previsiones para el mismo producto

En la figura anterior, en el gráfico de la izquierda, los valores a la izquierda de la línea negra vertical se rellenan con ceros, lo que genera una previsión sesgada a la baja (a la derecha de la línea negra vertical). En el gráfico de la derecha, estos valores están marcados como NaN, lo que permite realizar previsiones adecuadas.

Ejemplo 2

La figura anterior ilustra la importancia de gestionar correctamente los valores faltantes para un modelo de espacios de estado lineal, como [ARIMA o ETS](#). Representa la previsión de la demanda de un producto parcialmente sin existencias. La región de entrenamiento se muestra en verde en el gráfico de la izquierda, el rango de predicción en el panel derecho en rojo, y el objetivo real en negro. Las previsiones de mediana, p10 y p90 se muestran en la línea roja y en la región sombreada, respectivamente. En la parte inferior se muestran los productos sin existencias (80 % de los datos) marcados en rojo. En el gráfico de la izquierda, las áreas sin existencias se ignoran y se rellenan con 0.

Esto hace que los modelos de previsión supongan que hay muchos ceros que predecir y, por lo tanto, las previsiones sean demasiado bajas. En la gráfica de la derecha, las áreas sin existencias se tratan como observaciones de valores verdaderamente faltantes y la demanda se vuelve incierta en la región sin existencias. Si los valores faltantes para los artículos sin existencias se identifican debidamente como NaN, no se observa ningún sesgo a la baja en el rango de predicción de este gráfico. Amazon Forecast rellena estas lagunas en los datos, lo que le facilita la gestión adecuada de los datos faltantes sin tener que modificar explícitamente todos los datos de entrada.

Conceptos de caracterización y series temporales relacionadas

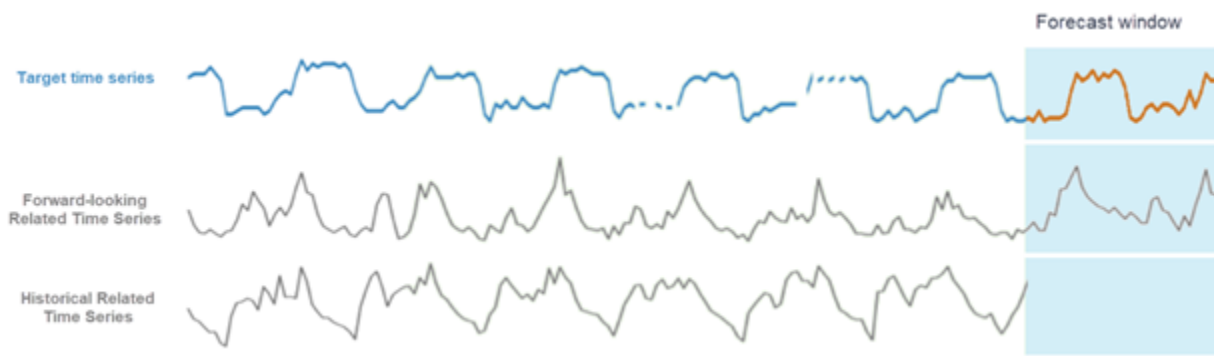
Amazon Forecast permite a los usuarios introducir datos relacionados para ayudar a mejorar la precisión de algunos modelos de previsión compatibles. Estos datos pueden ser de dos tipos: series temporales relacionadas o metadatos de elementos estáticos.

Note

Los metadatos y los datos relacionados se denominan características en machine learning y se utilizan como covariables en las estadísticas.

Las series temporales relacionadas son series temporales que tienen cierta correlación con el valor objetivo y deberían aportar cierta solidez estadística a la previsión sobre el valor objetivo, ya que ofrecen una explicación en términos intuitivos (consulte [Amazon Forecast: predicción de series temporales a escala](#) para ver un ejemplo). A diferencia de las series temporales objetivo, las series temporales relacionadas son valores conocidos del pasado que pueden afectar a la serie temporal objetivo y pueden tener valores conocidos en el futuro.

En Amazon Forecast, puede añadir dos tipos de series temporales relacionadas: series temporales históricas y series temporales prospectivas. Las series temporales relacionadas históricas contienen puntos de datos hasta el horizonte de previsión y no contienen puntos de datos dentro de ese horizonte ni en el futuro. Las series temporales relacionadas prospectivas contienen puntos de datos hasta y dentro del horizonte de previsión.



Diferentes enfoques sobre el uso de series temporales relacionadas con Amazon Forecast

Ejemplo 3

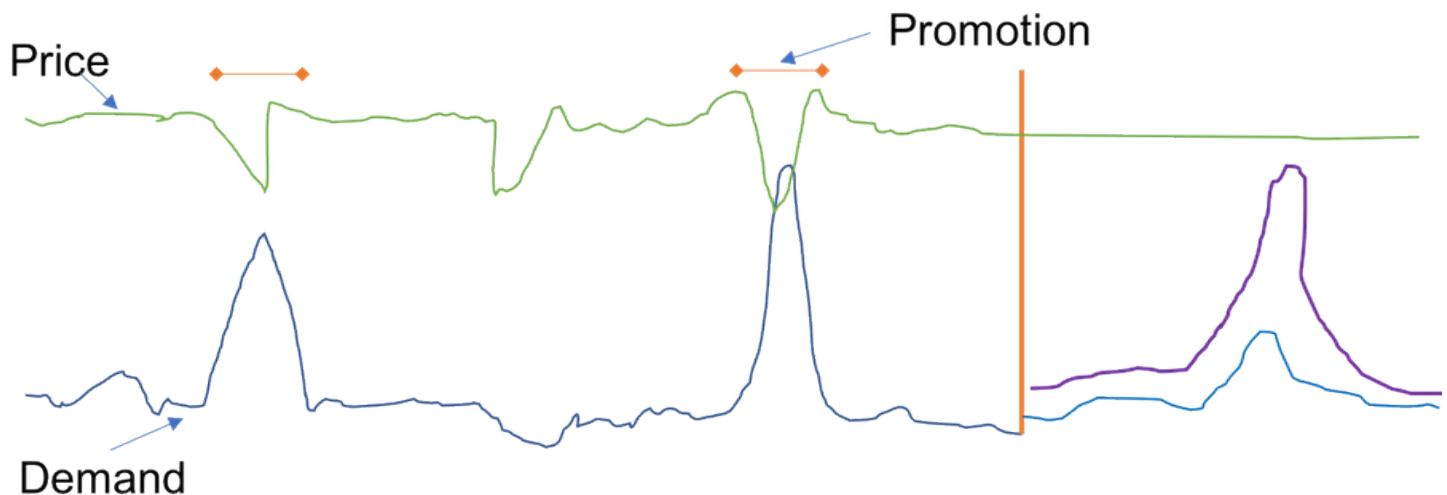
En la siguiente figura se muestra un ejemplo de cómo se pueden utilizar series temporales relacionadas para predecir la demanda futura de un libro popular. La línea azul representa la demanda en la serie temporal objetivo. El precio se muestra como la línea verde. La línea vertical representa la fecha de inicio de la previsión y las previsiones de los dos cuantiles se muestran a la derecha de la línea vertical.

En este ejemplo se utiliza una serie temporal relacionada prospectiva que se alinea con la serie temporal objetivo en cuanto a la granularidad de la previsión y que se conoce en todos los momentos futuros (o la mayoría de ellos) en el intervalo comprendido entre la fecha de inicio de la previsión y la suma de la fecha de inicio de la previsión y el horizonte de previsión (es decir, la fecha de finalización de la previsión).

En la siguiente figura también se muestra que el precio es una función adecuada para utilizar, ya que se pueden observar correlaciones entre una disminución del precio y un aumento de las ventas del producto. Pueden proporcionarse series temporales relacionadas a Amazon Forecast a través de un archivo CSV separado, que contiene la SKU del artículo, la marca temporal y los valores de las series temporales relacionadas (en este caso, el precio).

Amazon Forecast admite métodos de agregación, por ejemplo, por promedio o por suma, para las series temporales objetivo, pero no para las series temporales relacionadas. Por ejemplo, no tiene mucho sentido sumar un precio diario a un precio semanal, y lo mismo ocurre con las promociones diarias.

Amazon Forecast puede incorporar automáticamente información [meteorológica](#) y de [periodos festivos](#) en un modelo mediante la inclusión de conjuntos de datos con funciones integrados (consulte [SupplementaryFeature](#)). La información meteorológica y sobre periodos festivos puede afectar significativamente a la demanda en las ventas minoristas.



Las ventas de un producto en particular (en azul, a la izquierda de la línea roja vertical)

Los metadatos de los productos, también conocidos como variables categóricas, son otras funciones útiles que se pueden introducir en Amazon Forecast (consulte [Amazon Forecast: predicción de series temporales a escala](#) para ver un ejemplo). La principal diferencia entre las variables categóricas y las series temporales relacionadas es que las variables categóricas son estáticas, es decir, no cambian con el tiempo. Entre los ejemplos más comunes para la venta minorista se incluyen los colores de los productos, las categorías de libros y los indicadores binarios de si un televisor es un "Smart TV" o no. Los algoritmos de aprendizaje profundo pueden recopilar esta información al identificar las similitudes entre las unidades de mantenimiento de existencias (SKU), suponiendo que las SKU similares tengan ventas similares. Como estos metadatos no dependen del tiempo, cada fila del

archivo CSV de metadatos de productos solo contiene la SKU del artículo y la etiqueta o descripción de la categoría correspondiente.

Paso 3: Crear un predictor

Se puede crear un predictor de dos maneras: ejecutando [AutoML](#) o seleccionando manualmente uno de los seis algoritmos integrados de Amazon Forecast. Al ejecutar AutoML, en el momento de redactar este documento, Amazon Forecast prueba automáticamente los seis algoritmos integrados y elige el que tenga las pérdidas cuantiles medias más bajas en los cuantiles 10, 50 (mediana) y 90.

Amazon Forecast ofrece cuatro modelos locales:

- Autoregressive Integrated Moving Average (media móvil integrada autorregresiva, o [ARIMA](#))
- Exponential Smoothing (suavizado exponencial, o [ETS](#))
- Non-Parametric Time Series (serie temporal no paramétrica, o [NPTS](#))
- [Prophet](#)

Los modelos locales son métodos de previsión que ajustan un solo modelo a cada serie temporal individual (o combinación específica de elementos y dimensiones) y, a continuación, utilizan ese modelo para extrapolar las series temporales en el futuro.

ARIMA y ETS son versiones escalables de modelos locales populares del paquete de previsión R. El método NPTS, un método local desarrollado por Amazon, presenta una diferencia clave en comparación con los otros modelos locales. A diferencia de los simples previsores estacionales, que proporcionan previsiones puntuales repitiendo el último valor o el valor con una estacionalidad adecuada, el NPTS produce previsiones probabilísticas. El NPTS utiliza un índice de tiempo fijo, en el que el índice anterior ($T - 1$) o la temporada anterior ($T - \tau$) es la predicción del paso temporal T . El algoritmo muestrea aleatoriamente un índice de tiempo (t) en el conjunto $\{0, \dots, T - 1\}$ para generar una muestra para el paso temporal actual. El método NPTS es particularmente eficaz para series temporales intermitentes (a veces también denominadas dispersas) con muchos ceros. Forecast también incluye la implementación en Python de Prophet, un modelo de series temporales estructurales bayesianas.

Amazon Forecast ofrece dos algoritmos globales de aprendizaje profundo:

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

Los modelos globales entrenan un solo modelo a lo largo de toda la colección de series temporales de un conjunto de datos. Esto es especialmente útil cuando hay series temporales similares en un conjunto de unidades transversales. Por ejemplo, agrupaciones en series temporales de la demanda de distintos productos, cargas de servidor y solicitudes de páginas web.

En general, a medida que se incrementa el número de series temporales, aumenta la eficacia de CNN-QR y DeepAR+. Este no es siempre el caso de los modelos locales. Los modelos de aprendizaje profundo también se pueden utilizar para generar previsiones para nuevos SKU con pocos datos históricos de ventas (o ninguno). Esto se conoce como [previsión de arranque en frío](#).

	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)	✔	✔	✘	✘	✘	✘

Compare los algoritmos disponibles en Amazon Forecast

Para obtener más información sobre series temporales relacionadas, consulte [Series temporales relacionadas](#).

Paso 4: Evaluar los predictores

Un flujo de trabajo típico en el machine learning consiste en entrenar un conjunto de modelos o combinaciones de modelos en un conjunto de entrenamiento y evaluar su precisión en un conjunto de datos de retención. En esta sección se explica cómo dividir los datos históricos y qué métricas se deben utilizar para evaluar los modelos en la previsión de series temporales. En el ámbito de las previsiones, la técnica de backtesting es la principal herramienta para evaluar su precisión.

Backtesting

Un marco adecuado de evaluación y backtesting es uno de los factores más importantes para convertir una aplicación de machine learning en un éxito. Si consigue realizar un backtesting satisfactorio con sus modelos, podrá aumentar la confianza en su poder predictivo futuro. Además, puede ajustar los modelos mediante la optimización de hiperparámetros (HPO), aprender combinaciones de modelos y habilitar el metaaprendizaje y el AutoML.

El tiempo característico de la predicción de series temporales hace que sea diferente, en términos de metodología de evaluación y backtesting, de otros campos del machine learning aplicado. Por lo general, en las tareas de machine learning, para evaluar el error predictivo en una prueba de backtest, se divide un conjunto de datos por elementos. Por ejemplo, para la validación cruzada en tareas relacionadas con imágenes, se entrena el modelo con cierto porcentaje de las imágenes y, a continuación, se utilizan otras partes para realizar pruebas y validar. En la previsión, hay que dividir principalmente por tiempo (y, en menor medida, por elementos) para garantizar que no se filtre información del conjunto de entrenamiento al conjunto de prueba o validación y que se simule el caso de producción lo más fielmente posible.

La división por tiempo debe hacerse con cuidado, ya que no conviene elegir un solo punto en el tiempo, sino varios. De lo contrario, la precisión dependería demasiado de la fecha de inicio de la previsión, tal como se define en el punto de división. Una evaluación continua de las previsiones, en la que se realizan una serie de divisiones en varios puntos temporales y se obtiene el resultado promedio, genera resultados más sólidos y fiables en el backtest. En la siguiente figura se ilustran cuatro divisiones diferentes del backtest.

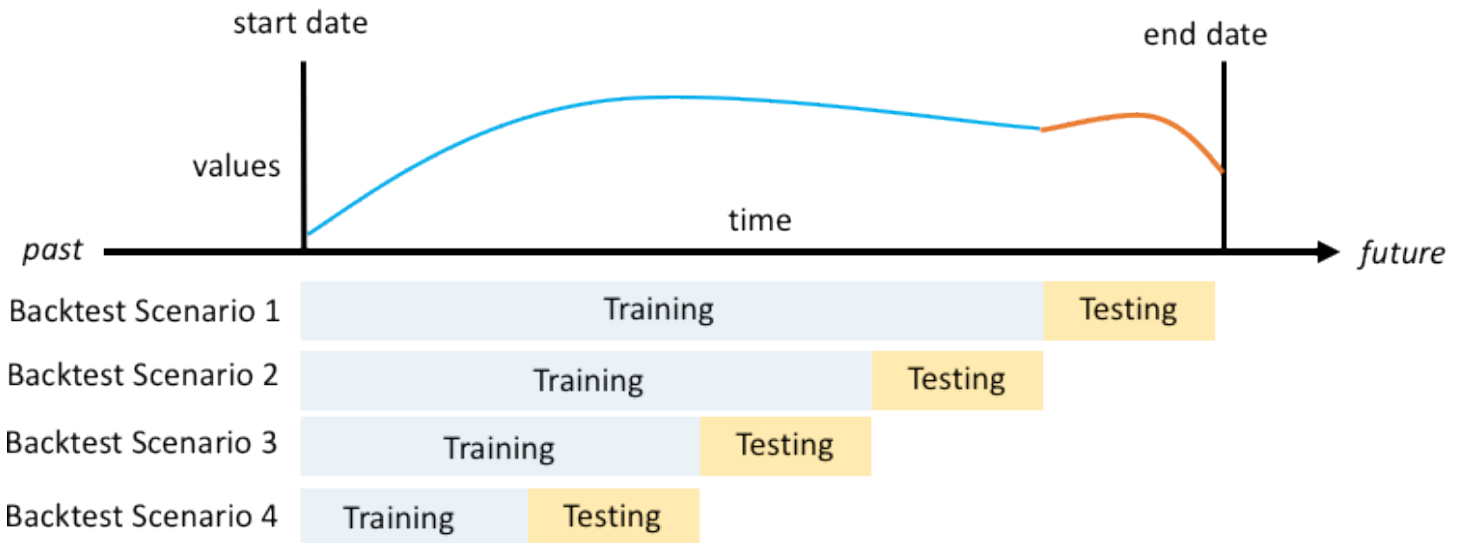


Ilustración de cuatro escenarios diferentes de backtesting con un tamaño creciente del conjunto de entrenamiento, pero un tamaño constante de las pruebas

En la figura anterior, todos los escenarios de backtesting tienen datos disponibles en su totalidad para poder evaluar los valores previstos con respecto a los reales.

La razón por la que se necesitan varias ventanas de backtesting es que la mayoría de las series temporales del mundo real normalmente no son estacionarias. La empresa de comercio electrónico del caso práctico tiene su sede en Norteamérica y gran parte de su demanda de productos se acentúa en la temporada alta del cuarto trimestre, con picos de demanda específicos en torno al Día de Acción de Gracias y antes de Navidad. En la temporada de compras del cuarto trimestre, la variabilidad de las series temporales es mayor que durante el resto del año. Al disponer de varias ventanas de backtesting, es posible evaluar los modelos de previsión en un entorno más equilibrado.

Para cada escenario de backtesting, la siguiente figura muestra los elementos básicos de la terminología de Amazon Forecast. Amazon Forecast divide automáticamente los datos en los conjuntos de datos de entrenamiento y de prueba. Amazon Forecast decide cómo dividir los datos de entrada mediante el parámetro `BackTestWindowOffset` especificado como parámetro en la API `create_predictor` o con su valor predeterminado de `ForecastHorizon`.

En la siguiente figura puede ver el primer caso, más habitual, en el que los parámetros `ForecastHorizon` y `BackTestWindowOffset` no son iguales. El parámetro `BackTestWindowOffset` define una fecha de inicio de previsión virtual, que aparece como línea vertical discontinua en la siguiente figura. Se puede utilizar para responder a la siguiente pregunta hipotética: si el modelo se implementa ese día, ¿cuál sería la previsión? El parámetro

`ForecastHorizon` define el número de pasos de tiempo que se deben predecir desde la fecha de inicio de la predicción virtual.

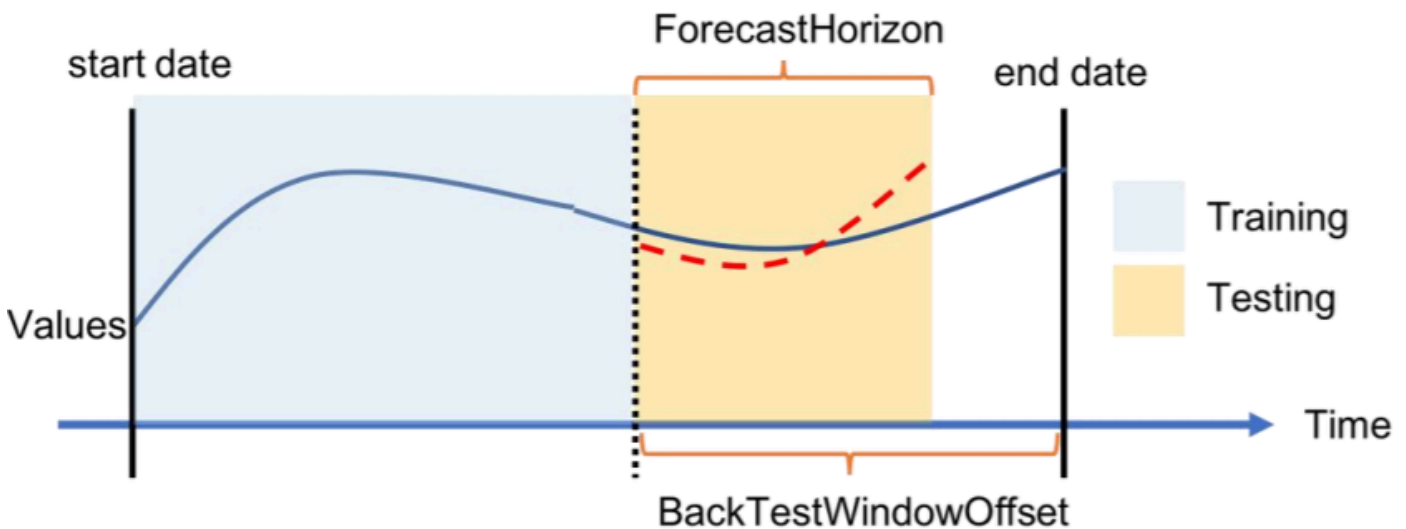


Ilustración de un único escenario de backtesting y su configuración en Amazon Forecast

Amazon Forecast puede exportar los valores previstos y las métricas de precisión generadas durante el backtesting. Los datos exportados se pueden utilizar para evaluar determinados elementos en puntos de tiempo y cuantiles específicos.

Cuantiles de predicción y métricas de precisión

Los cuantiles de predicción pueden proporcionar un límite superior e inferior para las predicciones. Por ejemplo, el uso de los tipos de predicción 0,1 (P10), 0,5 (P50) y 0,9 (P90) proporciona un rango de valores que se determina intervalo de confianza del 80 % en torno a la predicción del P50. Al generar predicciones en los cuantiles P10, P50 y P90, cabe esperar que el valor real se sitúe entre esos límites el 80 % de las veces.

En este documento se analizan con más detalle los cuantiles en el [paso 5](#).

Amazon Forecast utiliza las métricas de precisión de pérdida cuantil ponderada (wQL), raíz del error cuadrático medio (RMSE) y error porcentual absoluto ponderado (WAPE) para evaluar los predictores durante las pruebas de backtesting.

Pérdida cuantil ponderada (wQL)

La métrica de error de pérdida cuantil ponderada (wQL) mide la precisión de la predicción de un modelo en un cuantil específico. Resulta particularmente útil cuando la predicción excesiva y la

predicción insuficiente suponen costes diferentes. Al establecer la ponderación (τ) de la función wQL se incorporan automáticamente diferentes penalizaciones por quedarse corto o pasarse en la predicción.

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

Función wQL

Donde:

- τ : un cuantil del conjunto $\{0,01, 0,02, \dots, 0,99\}$
- $q_{i,t}(\tau)$: el cuantil τ que predice el modelo.
- $y_{i,t}$: el valor observado en el punto (i,t)

Error porcentual absoluto ponderado (WAPE)

El error porcentual absoluto ponderado (WAPE) es una métrica que se usa habitualmente para medir la precisión del modelo. Mide la desviación general de los valores previstos con respecto a los valores observados.

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

WAPE

Donde:

- $y_{i,t}$: el valor observado en el punto (i,t)
- $\hat{y}_{i,t}$: el valor previsto en el punto (i,t)

La previsión utiliza la previsión media como valor predicho, $\hat{y}_{i,t}$.

Raíz del error cuadrático medio (RMSE)

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

La raíz del error cuadrático medio (RMSE) es una métrica que se usa habitualmente para medir la precisión del modelo. Al igual que el WAPE, mide la desviación general de las estimaciones con respecto a los valores observados.

Donde:

- $y_{i,t}$: el valor observado en el punto (i,t)
- $\hat{y}_{i,t}$: el valor previsto en el punto (i,t)
- nT : el número de puntos de datos en un conjunto de pruebas

La previsión utiliza la previsión media como valor predicho, $\hat{y}_{i,t}$. Al calcular las métricas predictoras, nT es el número de puntos de datos en una ventana de backtesting.

Problemas con WAPE y RMSE

En la mayoría de los casos, las previsiones puntuales que se pueden generar internamente o a partir de otras herramientas de previsión deben coincidir con las previsiones medias o del cuantil p50. Tanto para WAPE como para RMSE, Amazon Forecast utiliza la previsión media para representar el valor previsto (\hat{y}).

Para una $\tau = 0,5$ en la ecuación de $w\text{QL}[\tau]$, ambas ponderaciones son iguales y el $w\text{QL}[0,5]$ se reduce al error porcentual absoluto ponderado (WAPE) que se utiliza comúnmente para las previsiones puntuales:

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

donde $\hat{y} = q(0,5)$ es la previsión informática. Se utiliza un factor de escala de 2 en la fórmula de $w\text{QL}$ para cancelar el factor de 0,5 y obtener la expresión del WAPE[mediana] exacta.

Tenga en cuenta que la definición anterior del WAPE difiere de una interpretación común del error porcentual absoluto medio ([MAPE](#)). La diferencia está en el denominador. La forma en que se define el WAPE arriba evita el problema de la división por 0, un problema que ocurre con frecuencia en escenarios del mundo real, como la empresa de comercio electrónico del caso práctico, que suele vender 0 unidades de una SKU determinada en un día determinado.

A diferencia de la métrica de pérdida cuantil ponderada para tau no igual a 0,5, el sesgo inherente a cada cuantil no se puede captar mediante un cálculo como el WAPE, en el que las ponderaciones son iguales. Otras desventajas del WAPE incluyen que no es simétrico, tiene una inflación excesiva de los errores porcentuales para números pequeños y es solo una métrica puntual.

La RMSE es el cuadrado del término de error en el WAPE y una métrica de error común en otras aplicaciones de aprendizaje automático. La métrica RMSE funciona mejor en modelos en los que los errores individuales tienen una magnitud uniforme, ya que una variación grande en los errores aumentará la RMSE de forma desproporcionada. Debido al error cuadrático, unos cuantos valores mal previstos en una buena previsión pueden aumentar el RMSE. Además, como consecuencia de la elevación al cuadrado de los términos, los términos de error más pequeños tienen menos peso en la RMSE que en el WAPE.

Las métricas de precisión permiten una evaluación cuantitativa de las previsiones. En particular, para las comparaciones a gran escala (por ejemplo, si el método A es mejor que el método B en general), resultan cruciales. Sin embargo, a menudo es importante complementar esto con elementos visuales para cada SKU.

Paso 5: Generar y utilizar previsiones para la toma de decisiones

Una vez que disponga de un modelo que se ajuste al umbral de precisión requerido para su caso de uso específico (según lo determinado mediante pruebas de backtesting), el último paso consiste en implementar el modelo y generar previsiones. Para implementar un modelo en Amazon Forecast, debe ejecutar la API `Create_Forecast`. Esta acción aloja un modelo creado mediante el entrenamiento en todo el conjunto de datos histórico (a diferencia de `Create_Predictor`, que divide los datos en un conjunto de entrenamiento y otro de prueba). Las predicciones del modelo generadas durante el horizonte de previsión se pueden utilizar, por tanto, de dos maneras:

- Puede consultar las previsiones de un elemento en particular (especificando el elemento o la combinación de elemento/dimensión) mediante la API `Query_Forecast` desde [AWS CLI](#) o directamente a través de [AWS Management Console](#).
- Puede generar las previsiones para todas las combinaciones de elementos y dimensiones en todos los cuantiles mediante la API `Create_Forecast_Export_Job`. Esta API genera un archivo CSV que se almacena de forma segura en la ubicación de [Amazon Simple Storage Service](#) (Amazon S3) de su elección. A continuación, puede utilizar los datos del archivo CSV y conectarlos a sus sistemas descendentes utilizados para la toma de decisiones. Por ejemplo, sus sistemas de cadena de suministro existentes pueden incorporar la producción de Amazon Forecast directamente para determinar la toma de decisiones sobre la fabricación de SKU específicos.

Previsiones probabilísticas

Amazon Forecast puede generar previsiones en diferentes cuantiles, lo que resulta especialmente útil cuando los costes de pasarse y de quedarse corto en una previsión son diferentes. Al igual que en la etapa de entrenamiento de predictores, se pueden generar previsiones probabilísticas para cuantiles entre p1 y p99.

De forma predeterminada, Amazon Forecast genera previsiones en los mismos cuantiles que se utilizan durante el entrenamiento de los predictores. Si no se especifican los cuantiles durante el entrenamiento del predictor, se generarán en los cuantiles p10, p50 y p90 de forma predeterminada.

Para la previsión p10, se espera que el valor real sea inferior al valor previsto el 10 % de las veces, y se puede utilizar la métrica $wQL[0,1]$ para evaluar su precisión. Esto significa que la previsión del

P10 está por debajo de lo previsto el 90 % de las veces y, si se utiliza para almacenar inventario, el 90 % de las veces el artículo estaría agotado. La previsión del P10 podría resultar útil cuando no haya mucho espacio de almacenamiento o el coste del capital invertido sea elevado.

Note

La definición formal de una previsión cuantílica es $Pr(\text{valor real} \leq \text{previsión en el cuantil } q) = q$. Técnicamente, un cuantil es un percentil/100. Los expertos en estadística suelen decir "nivel del cuantil P90", ya que es más fácil decir eso que "cuantil 0,9". Por ejemplo, una previsión a un nivel del cuantil P90 significa que cabe esperar que el valor real sea inferior a la previsión el 90 % de las veces. Concretamente, si en un tiempo = t_1 y a un nivel de cuantil = 0,9 el valor previsto = 30, significa que se esperaría que el valor real en el tiempo = t_1 , si se hubieran realizado 1000 simulaciones, fuera inferior a 30 para 900 simulaciones y superior a 30 en las 100 simulaciones restantes.

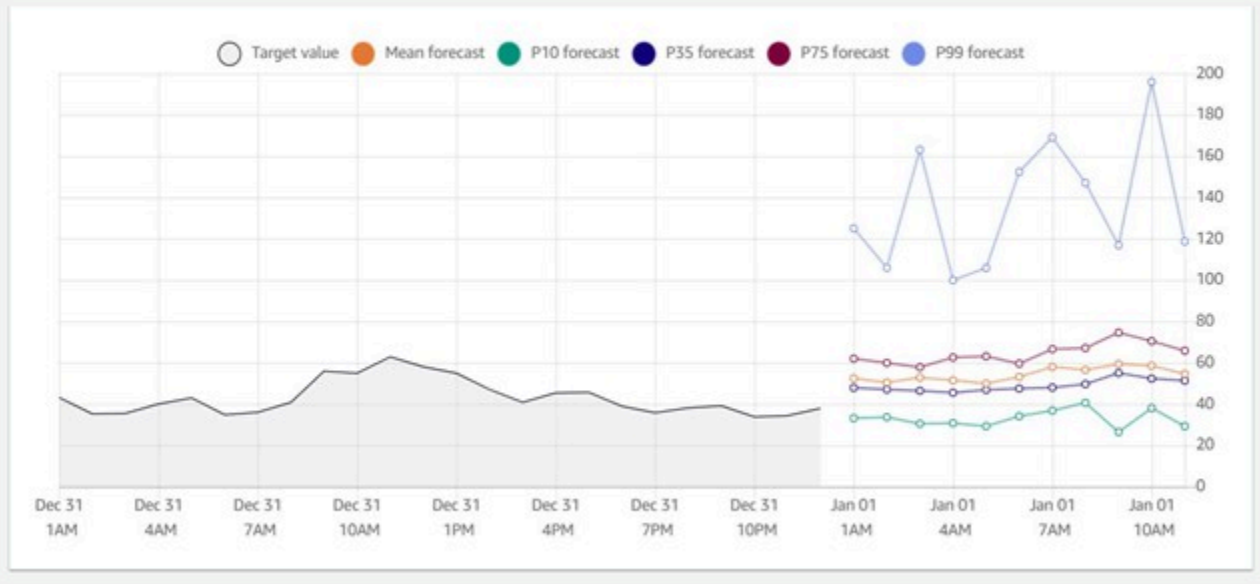
Por otro lado, la previsión del P90 es excesiva el 90 % de las veces, y resulta útil cuando el coste de oportunidad de no vender un producto es extremadamente alto o si el coste del capital invertido es bajo. Para un supermercado, por ejemplo, la previsión del P90 podría usarse para productos como leche o papel higiénico, ya que nunca es conveniente que la tienda se quede sin este tipo de productos y no le importa que siempre queden existencias en las estanterías.

Para la previsión p50 (denominada también a menudo previsión mediana), se espera que el valor real sea inferior al valor previsto el 50 % de las veces, y se puede utilizar la métrica $wQL[0,5]$ para evaluar su precisión. Cuando el exceso de existencias no resulta demasiado preocupante y hay una demanda moderada de un producto determinado, la previsión del cuantil p50 puede resultar útil.

Visualización

Amazon Forecast permite trazar previsiones de forma nativa en la AWS Management Console. Además, puede aprovechar el paquete completo de ciencia de datos de Python (consulte los [ejemplos de Amazon Forecast](#)). Amazon Forecast permite exportar las previsiones en un archivo CSV a través de la API `ExportForecastJob`, lo que permite a los usuarios visualizar la previsión en la herramienta analítica que elijan.

Item_id: client_12



Visualización proporcionada en la consola de Amazon Forecast en diferentes cuantiles

Resumen del flujo de trabajo de previsión y las API

En la siguiente tabla asociamos cada paso del flujo de trabajo de previsión con la API de Amazon Forecast correspondiente.

Tabla 1: Pasos de previsión y API de Amazon Forecast

Paso	API	Funciones de la API
Paso 1: Recopilar y agregar datos Paso 2: Preparar los datos	Create_Dataset_Group , Create_Dataset , Create_Dataset_Import_Job	<ol style="list-style-type: none"> 1. Establezca el dominio de alto nivel (venta minorista , métricas, etc.) del problema. 2. Defina el esquema de los diferentes conjuntos de datos (objetivo, relacionados, metadatos de elementos). 3. Importe datos de Amazon S3 a Amazon Forecast.
Paso 3: Crear un predictor Paso 4: Evaluar los predictores	Create_Predictor	<ol style="list-style-type: none"> 1. Realiza la ETL. 2. Divide los datos en conjuntos de entrenamiento/prueba y entrena el modelo. 3. Opcionalmente puede usar Create_predictor_backtest_Export_job para exportar los resultados del backtesting en formato CSV para calcular las métricas al nivel del elemento.

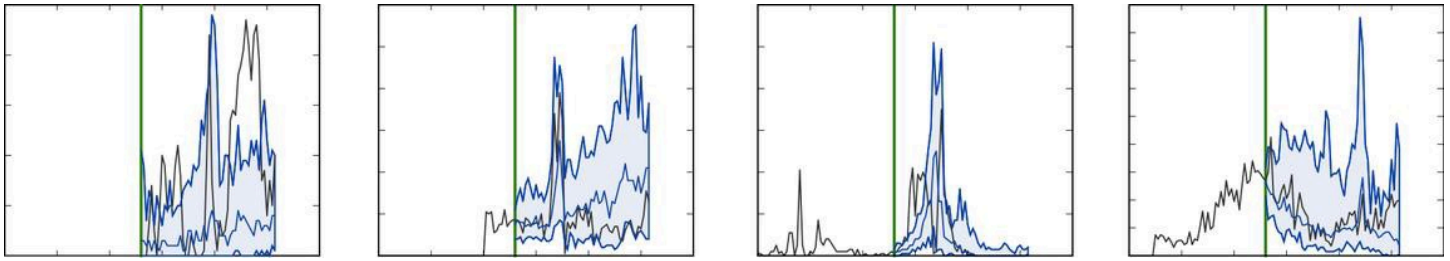
Paso	API	Funciones de la API
Paso 5: Generar y utilizar previsiones para la toma de decisiones	Create_Forecast	<ol style="list-style-type: none"> Entrena y aloja el modelo. Genera predicciones sobre el horizonte de previsión para un cuantil de interés específico (por ejemplo, cualquier número entero comprendido entre 1 y 99, incluida la media).
	Query_Forecast . Create_Forecast_Export_Job	Le permite consumir las previsiones creadas por Create_Forecast

Uso de Amazon Forecast para escenarios comunes

También puede realizar un análisis hipotético mediante la generación de diferentes previsiones en función de los cambios en las variables externas (por ejemplo, precios o promociones). Por ejemplo, en el ejemplo del caso práctico de comercio electrónico, puede crear diferentes previsiones en función de las promociones que esté planificando. Puede prever la demanda de un producto con un 10 % de descuento y luego con un 20 % de descuento para comprender la cantidad de producto que debería tener en stock para satisfacer esa demanda. Esto se puede lograr configurando grupos de conjuntos de datos únicos y actualizando las series temporales relacionadas en cada uno de ellos, según el escenario de interés.

Además, también puede generar previsiones para productos sin historial previo (lo que a veces se denomina problema de arranque en frío). Este enfoque requiere crear un predictor mediante DeepAR + o CNN-QR junto con metadatos (como un conjunto de datos de metadatos de elementos) para generar previsiones para el nuevo elemento.

La siguiente figura muestra ejemplos de cuatro SKU diferentes tal y como aparecen en problemas de previsión operativa del mundo real.



Ejemplos de cuatro SKU diferentes tal y como aparecen en problemas de previsión operativa del mundo real.

En la imagen anterior, los datos reales observados están a la izquierda de la línea vertical y las previsiones en azul están a la derecha de la línea vertical, comparadas con los datos reales en negro. Tenga en cuenta que el historial de cada SKU individual, a la izquierda de la línea vertical, no es indicativo de su evolución a la derecha de la línea verde.

Implementación de previsiones en producción

Tras completar el flujo de trabajo integral de Amazon Forecast, es fundamental identificar las diferencias clave entre las API `Create_Predictor` y `Create_Forecast` y cuándo se debe utilizar cada una de ellas.

La primera se usa principalmente durante la prueba de concepto para evaluar la precisión o las métricas del modelo, mientras que la segunda se usa para generar previsiones en un entorno de producción.

Una vez en producción, no es necesario ejecutar `Create_Predictor` cada vez que se deba generar una previsión, sino solamente cuando se deba volver a entrenar el modelo debido a cambios en los datos o como parte de una cadencia preestablecida (por ejemplo, quincenal o mensual). A medida que los conjuntos de datos se actualicen con nuevos datos, solo será necesario ejecutar `Create_Forecast` para generar previsiones para el nuevo horizonte de previsión.

En producción, también debe automatizar las operaciones de importación y previsión de sus conjuntos de datos para generar nuevas predicciones de forma continua. Hoy en día, esto se puede lograr mediante la configuración de trabajos cron mediante una combinación de registros de [Eventos de Amazon CloudWatch](#), [AWS Step Functions](#) y funciones de [AWS Lambda](#). La configuración del trabajo [cron](#), a su vez, automatiza las llamadas a la API de Amazon Forecast para la importación, el reentrenamiento o la generación de previsiones. Por último, es clave administrar los recursos y eliminarlos a intervalos regulares para no superar los [límites del sistema](#) prescritos por el servicio.

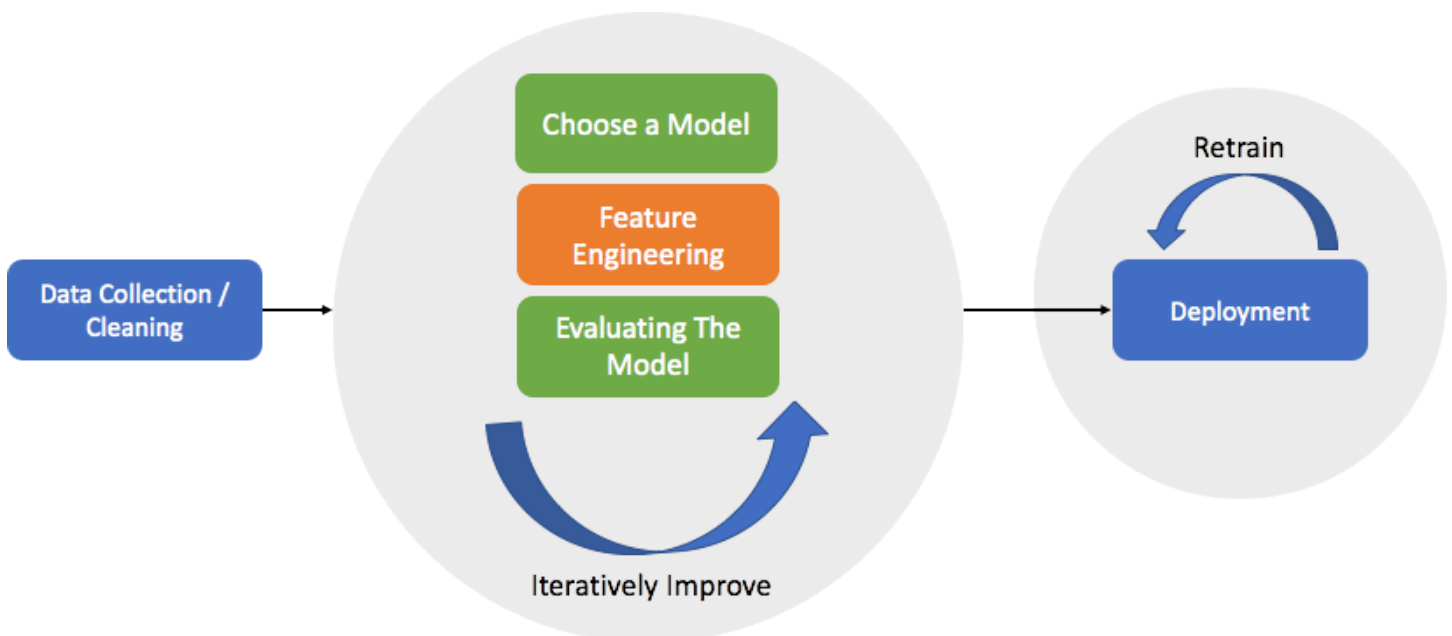
Consulte esta [entrada de blog](#), que incluye [Amazon Redshift](#), para obtener más información sobre la configuración de trabajos programados.

Conclusión

[Januschowski y Kolassa \(2019\)](#) proporcionan una clasificación de los problemas de previsión alineada con las decisiones que deben tomar las empresas, incluidas las decisiones estratégicas, tácticas y operativas. Cada nivel de decisión tiene sus tareas de previsión correspondientes.

Los problemas de previsión operativa y táctica se caracterizan por contener grandes cantidades de datos y, por lo general, por requerir un alto grado de automatización. Hay diferentes métodos de previsión que solucionan estos problemas. Los métodos de previsión locales suelen funcionar bien para los problemas de previsión estratégica, los métodos basados en el aprendizaje profundo son idóneos para los problemas de previsión operativa y, para los casos intermedios, es posible que sea necesario experimentar un poco. Si bien en este documento se abordan los problemas de previsión operativa, Amazon Forecast no es inflexible a la hora de ofrecer modelos e incluye modelos que resuelven problemas de previsión tanto estratégicos como operativos y tácticos.

El proceso de resolución de problemas de previsión operativa se puede dividir en pasos básicos, desde la recopilación y preparación de datos hasta la creación y despliegue de modelos. En general, lo más útil es considerar esto como un proceso iterativo en lugar de lineal. Por ejemplo, a medida que se comprendan mejor los modelos y los casos prácticos, puede tener sentido volver a la fase de recopilación de datos. El desarrollo de modelos también es muy iterativo en sí mismo.



Proceso de desarrollo simplificado para llevar un modelo de previsión a la producción.

Colaboradores

Entre los colaboradores de este documento, están las siguientes personas:

- Yuyang Wang, científico sénior de Machine Learning, servicios verticales de IA
- Danielle Robinson, científica aplicada, servicios verticales de IA
- Tim Januschowski, gerente de ciencia aplicada de ML
- Namita Das, directora sénior de productos, servicios verticales de IA
- Christy Bergman, arquitecta sénior de soluciones especializada en IA/ML
- Kris Tonthat, redactor técnico de documentación de IA/ML

Documentación adicional

Para obtener información adicional sobre la previsión de series temporales y los métodos de aprendizaje profundo, consulte:

- [Documentación de Amazon Forecast](#)
- [Blog de disponibilidad general de Amazon Forecast](#)
- [Ahora disponible en Amazon SageMaker: algoritmo de DeepAR para una mayor precisión](#)
- [Amazon SageMaker DeepAR ahora admite valores faltantes, funciones categóricas y de series temporales y frecuencias generalizadas](#)
- [Amazon Forecast ahora puede usar redes neuronales convolucionales \(CNN\) para entrenar modelos de previsión hasta 2 veces más rápido con una precisión hasta un 30 % mayor](#)
- [Amazon Forecast ahora admite medidas de precisión para elementos individuales](#)
- [Medición fácil de la precisión del modelo de previsión para optimizar sus objetivos comerciales con Amazon Forecast](#)
- [Amazon Forecast Weather Index: incluya automáticamente el tiempo local para aumentar la precisión de su modelo de previsión](#)
- [Artículos científicos sobre modelos de predicción de series temporales](#)
- [Página de GitHub de ejemplos de Amazon Forecast](#)
- [Centro de arquitectura de AWS](#)

Apéndice A: Preguntas frecuentes

P: ¿Cómo puedo comenzar a utilizar Amazon Forecast?

1. Primero, necesitará una Cuenta de AWS.
2. A continuación, abra el servicio de previsión en la [AWS Management Console](#), cree un grupo de conjuntos de datos e importe un archivo .csv en el conjunto de datos de la serie temporal objetivo (obligatorio). Los datos mínimos necesarios para empezar son los datos históricos de la cantidad que quiera predecir, como la electricidad por marca temporal y hogar.
3. Por último, cree un modelo ejecutando [CreatePredictor](#) y genere resultados ejecutando [CreateForecast](#). Consulte la página de documentación de [Introducción](#) para obtener más detalles.

Además, consulte la [Introducción a GitHub y la Guía de prácticas recomendadas](#).

P: ¿Amazon Forecast es una buena opción para mí?

No todos los problemas del machine learning son problemas de previsión. La primera pregunta que hay que plantearse es: "¿Mi problema empresarial incluye series temporales en su planteamiento?" Por ejemplo, ¿necesita un valor concreto solo en una fecha y hora determinadas en el futuro? La previsión no es idónea para problemas generales y estáticos (en los que la fecha y la hora en particular no importan), como la detección de fraudes o la recomendación de títulos de películas a los usuarios. Hay soluciones mucho más rápidas para esos problemas estáticos.

Además de tener datos de series temporales, los datos en sí deben ser "densos" y tener historiales nutridos. Esto se resume en la siguiente tabla:

Tabla 2: Criterios y clases de algoritmos de Amazon Forecast

Criterios	Clase de algoritmo de Amazon Forecast
Amplio conjunto de datos con hasta cinco millones de series temporales con patrones subyacentes similares, efectos estacionales y datos relacionados. Cada serie temporal debe tener un historial nutrido, idealmente de más de dos años si se intenta capturar eventos	Productos de aprendizaje profundo propiedad de Amazon Forecast: DeepAR+ y CNN-QR

Criterios	Clase de algoritmo de Amazon Forecast
anuales, y cada serie temporal debe tener más de 300 puntos de datos, 1000 en casos ideales.	
Conjunto de datos pequeño con entre 1 y varios cientos de series temporales, donde la mayoría de las series temporales tienen más de 300 puntos de datos, efectos estacionales y datos relacionados.	Prophet
Conjunto de datos pequeño con entre 1 y varias decenas de series temporales, donde la mayoría de las series temporales tienen más de 300 puntos de datos y efectos estacionales.	ETS, ARIMA
Intermitente (dispersa con muchos ceros) con entre 1 y varias decenas de series temporales, donde la mayoría de las series temporales tienen más de 300 puntos de datos.	NPTS propiedad de Amazon Forecast
Conjunto de datos pequeño (regular o disperso) con entre 1 y varias decenas de series temporales, donde la mayoría de las series temporales tienen menos de 300 puntos de datos.	Los datos son demasiado pequeños para Amazon Forecast. Pruebe ETS en Excel o los modelos estadísticos tradicionales ARIMA y Prophet en su lugar.

Lo recomendable es entrenar con el modo AutoML en su Predictor la primera vez que utilice sus datos. AutoML ejecutará automáticamente todos los algoritmos (los algoritmos de DL se ejecutan con HPO activado) para averiguar qué algoritmo funciona mejor con sus datos.

P: ¿Cómo me planteo la carencia de datos? ¿Cuándo es demasiado importante como para generar previsiones razonables?

Puede darse el caso de que haya problemas en el registro de los datos o que el nivel de agregación de los datos sea demasiado bajo o demasiado alto. La regla general es que la duración de la previsión no puede superar un tercio de los datos de entrenamiento.

Además de la cantidad de datos que falten, también hay que plantearse cómo imputar los datos que no tengamos. Puede convertir todos los 0 en valores nulos y dejar que Amazon Forecast se encargue de imputar automáticamente los valores faltantes. Amazon Forecast detectará automáticamente si los valores faltantes se deben a la introducción de nuevos productos (arranques en frío) o a productos que han llegado al final de su vida útil. Puede utilizar varias lógicas para los valores faltantes, como el valor, la mediana, el mínimo, el máximo, el cero, la media y los valores NaN (solo series temporales objetivo). Consulte [la documentación para conocer la sintaxis de relleno de valores nulos](#).

- "frontfill" (relleno frontal): (solo TTS) hace referencia a los elementos nuevos o de arranque en frío y a la forma en que se quieren tratar los valores nulos antes de que el elemento comience a tener un historial
- "middlefill" (relleno medio): se refiere a valores nulos en la mitad de los valores de las series temporales
- "backfill" (relleno retroactivo): hace referencia a los productos que han llegado al final de su vida útil y a la forma en que se quieren tratar los valores nulos cuando un producto ha dejado de venderse
- "futurefill" (relleno a futuro): (solo RTS) se refiere a los valores nulos que se producen una vez finalizados los datos de entrenamiento

P: Mis datos históricos de entrada no tienen valores negativos, pero veo valores negativos en las previsiones de demanda. ¿Por qué ocurre esto? ¿Qué puedo hacer para evitar esto?

No hay ninguna garantía de generar números positivos en ninguno de los modelos, excepto NPTS (que está entrenado con datos no negativos) y DeepAR (que tiene una función de verosimilitud binomial negativa). La solución consiste en cambiar a uno de los modelos mencionados o en limitar los valores de previsión a valores no negativos.

P: ¿Por qué las métricas de precisión difieren en los cuantiles? ¿El error no debería ser el mismo, ya que el modelo es el mismo?

Consulte la [Pérdida cuantil ponderada \(wQL\)](#) para obtener más información sobre cómo la ponderación depende del cuantil.

Imagine que tienes todas las previsiones en tres cuantiles diferentes: p10, p50, p90. Las tres predicciones en sí mismas son variables aleatorias. Los niveles de precisión se calculan por separado entre los datos reales y las previsiones en el nivel de cada cuantil. Es posible que vea una tabla de "wQL", pérdidas cuantiles ponderadas, como la que se muestra a continuación. Los valores

de las wQL no tienen una relación determinista entre sí. (Las pérdidas por retirada implican un error, por lo que no están ordenadas; sin embargo, las previsiones de cuantiles están ordenadas). Por lo tanto, no hay ninguna razón por la que la wQL del p90 deba ser mayor que la wQL del p50, por ejemplo.

Tabla 3 — Ejemplo de cuantiles de previsión

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

P: ¿Cómo puedo mejorar la precisión de las previsiones?

La precisión de la previsión depende de si los datos correctos están disponibles en la cantidad y calidad adecuadas. Si la precisión no es satisfactoria, puede ser recomendable intentar comprender cómo de predecible es el problema (o cómo de aleatorios / imprecisos / estacionarios son los datos). Otros factores que se deben tener en cuenta incluyen la evaluación de diferentes modelos, la configuración de los hiperparámetros y la incorporación de funciones adicionales mediante el uso de los conjuntos de datos de metadatos de elementos y las series temporales relacionados. Para obtener sugerencias específicas, [consulte este documento de prácticas recomendadas en GitHub](#).

P: Tengo un algoritmo que funciona muy bien en mi caso práctico pero no está en Amazon Forecast. ¿Qué tengo que hacer?

El equipo de Amazon Forecast estará encantado de ayudarle con este caso práctico. Póngase en contacto con el equipo de servicio de Amazon Forecast por correo electrónico: <amazonforecast-poc@amazon.com>.

Apéndice B: Referencias

[Januschowski, Tim y Kolassa, Stephan. Una clasificación de los problemas de previsión empresarial. Foresight: Revista internacional de previsión aplicada. 2019](#)

[Salinas, David, Flunkert, Valentin, Gasthaus, Jan y Januschowski, Tim. DeepAR: previsión probabilística con redes recurrentes autorregresivas. Revista internacional de previsiones. 2019](#)

[Gasthaus, Jan, Benidis, Konstantinos, Wang, Yuyang, Rangapuram, Syama Sundar, Salinas, David, Flunkert, Valentin y Januschowski, Tim. {Previsión probabilística con RNN de funciones spline cuantílicas. 22ª Conferencia Internacional sobre Inteligencia Artificial y Estadística. 2019](#)

[Januschowski, Tim, Gasthaus, Jan, Wang, Yuyang, Salinas, David, Flunkert, Valentin, Bohlke-Schneider, Michael y Callot, Laurent. Criterios para clasificar los métodos de previsión. Revista internacional de previsiones. 2019 \(es necesario iniciar sesión\)](#)

[Januschowski, Tim, Gasthaus, Jan, Wang, Yuyang, Rangapuram, Syama y Callot, Laurent. Aprendizaje profundo para la previsión. Foresight: Revista internacional de previsión aplicada. 2018](#)

[Januschowski, Tim, Gasthaus, Jan, Wang, Yuyang, Rangapuram, Syama Sundar y Callot, Laurent. Aprendizaje profundo para la previsión: tendencias y desafíos actuales. Foresight: Revista internacional de previsión aplicada. 2018](#)

[Bose, Joos-Hendrik, Flunkert, Valentin, Gasthaus, Jan, Januschowski, Tim, Lange, Dustin, Salinas, David, Schelter, Sebastian, Seeger, Matthias y Wang, Yuyang. Previsión probabilística de la demanda a escala. Actas de VLDB Endowment. 2017](#)

Historial de revisión

Para recibir notificaciones sobre las actualizaciones de este documento técnico, suscríbase a la fuente RSS.

Cambio	Descripción	Fecha
Actualización del documento técnico	Actualizaciones.	1 de septiembre de 2021
Publicación inicial	Primera edición del documento técnico.	4 de febrero de 2020

Note

Para suscribirse a las actualizaciones de RSS, debe disponer de un complemento de RSS habilitado para el navegador que utilice.

Avisos

Los clientes son responsables de realizar sus propias evaluaciones de la información contenida en este documento. Este documento: (a) solo tiene fines informativos, (b) representa las prácticas y las ofertas de productos vigentes de AWS, que están sujetas a cambios sin previo aviso, y (c) no crea ningún compromiso ni garantía de AWS y sus empresas afiliadas, proveedores o concesionarios de licencias. Los productos o servicios de AWS se proporcionan "tal cual", sin garantías, representaciones ni condiciones de ningún tipo, ya sean explícitas o implícitas. Las responsabilidades y obligaciones de AWS en relación con sus clientes se rigen por los acuerdos de AWS, y este documento no modifica ni forma parte de ningún acuerdo entre AWS y sus clientes.

© 2021 Amazon Web Services, Inc. o sus empresas afiliadas. Todos los derechos reservados.

Glosario de AWS

Para ver la terminología más reciente de AWS, consulte el [glosario de AWS](#) en la referencia general de AWS.