



aws INNOVATE

AI/ML EDITION

24 February 2022

Amazon Forecast로 비즈니스 시계열 예측 모델을 손쉽게 개발해보자!

김영민

데이터 사이언티스트

AWS



Agenda

- 비즈니스 현황과 예측 작업
- Amazon Forecast 소개
- 콘솔 데모

비즈니스 현황

다양한 취향의 고객들이 빠르고 저렴한 서비스를 점점 더 원하고 있습니다.

유연하고 정확한 예측 작업이 중요해졌습니다. 그러나 기존 방법으로는 다양하고 복잡한 수요 패턴을 포착하기 어렵습니다.

시장 선도 기업들은 시장 수요를 보다 효과적으로 충족시킬 수 있게 ML 기반 예측 작업에 투자하기 시작했습니다.

정확도 20% 향상

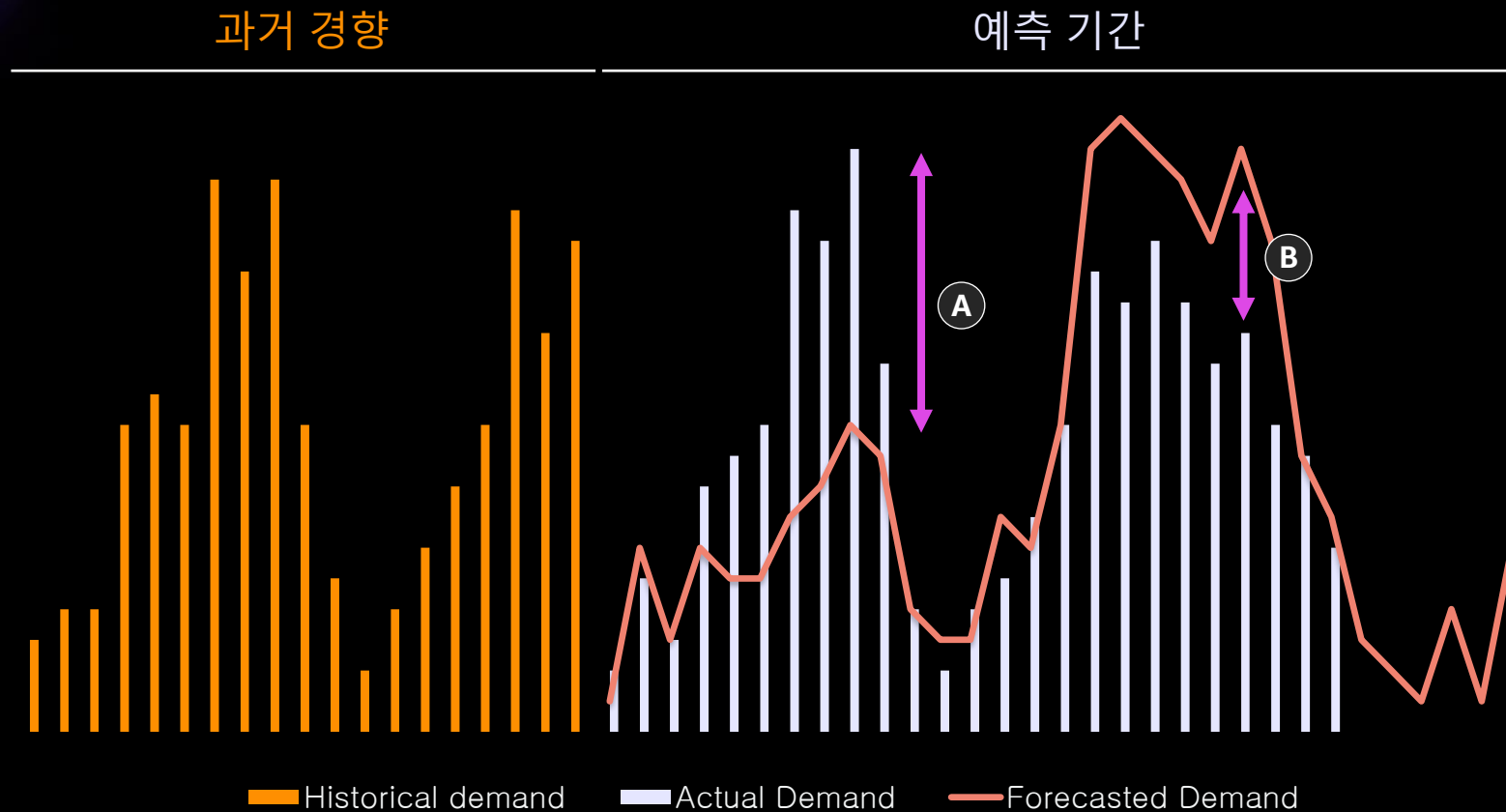
재고 5% 감소

매출 총 이익 3% 증대



예측 *Forecasting*이란?

과거 경향에 기반한 예측 과학



예측으로 다음을 최소화하려고 합니다.

- Ⓐ 비즈니스 기회를 놓치는 과소 예측 오류
- Ⓑ 리소스 낭비를 초래하는 과대 예측 오류

비즈니스 성과 향상하기

예측의 힘

재고 계획



세분화된 수준에서 수요 계획 개선
낭비 감소, 재고 회전율 증가, 재고 가용성 향상

인력 계획



다양한 수요를 충족하기 위해 보다 효과적으로
직원 양성
활용도, 서비스 제공 시간과 고객 만족도 향상

생산 능력 계획



더 높은 신뢰도로 장기적인 의사 결정 가능
자본 활용도 향상

재정 계획



매출과 매출 총 이익에 대한 더 정밀한 계획
현금 흐름을 효과적으로 관리

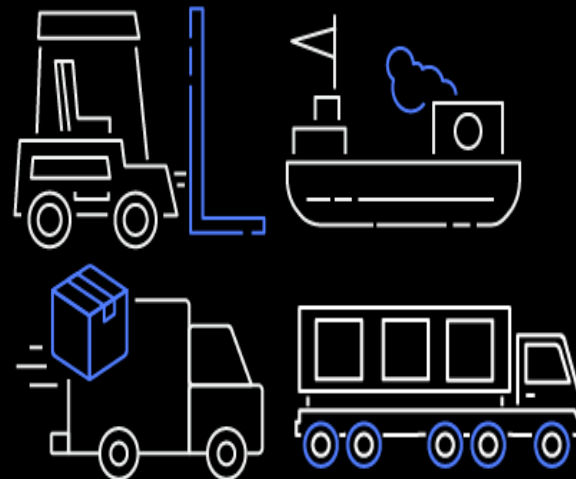
Amazon.com에서 사용하는 예측 기술 기반으로 소비자 니즈 충족을 위해 예측 활동은 필수



더 높은 제품 가용성



더 낮은 가격



더 빠른 배송

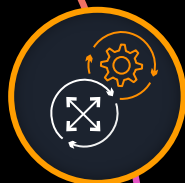
Amazon.com에서 사용하는 예측 기술 기반으로 20년 이상 동안 발전되어 온 예측 기술



ML 활용하여 비즈니스 성과 지표 개선 Amazon Forecast의 장점



높은 정확도 기존 방법보다 최대 50% 더 정확해진 ML 모델을 훈련하고 배포합니다. 딥 러닝 경험이 필요하지 않습니다.



완전 관리형 서비스 Amazon Forecast에서 클릭 몇 번으로 사용자 지정 ML 모델을 구축, 훈련 및 배포하는 모든 작업을 수행하여 대규모 ML 솔루션을 쉽게 구현할 수 있습니다.



사용과 통합 용이성 Amazon Forecast를 기존 공급망 관리 시스템에 통합하여 모든 프로세스에 ML 예측 작업을 추가할 수 있습니다.

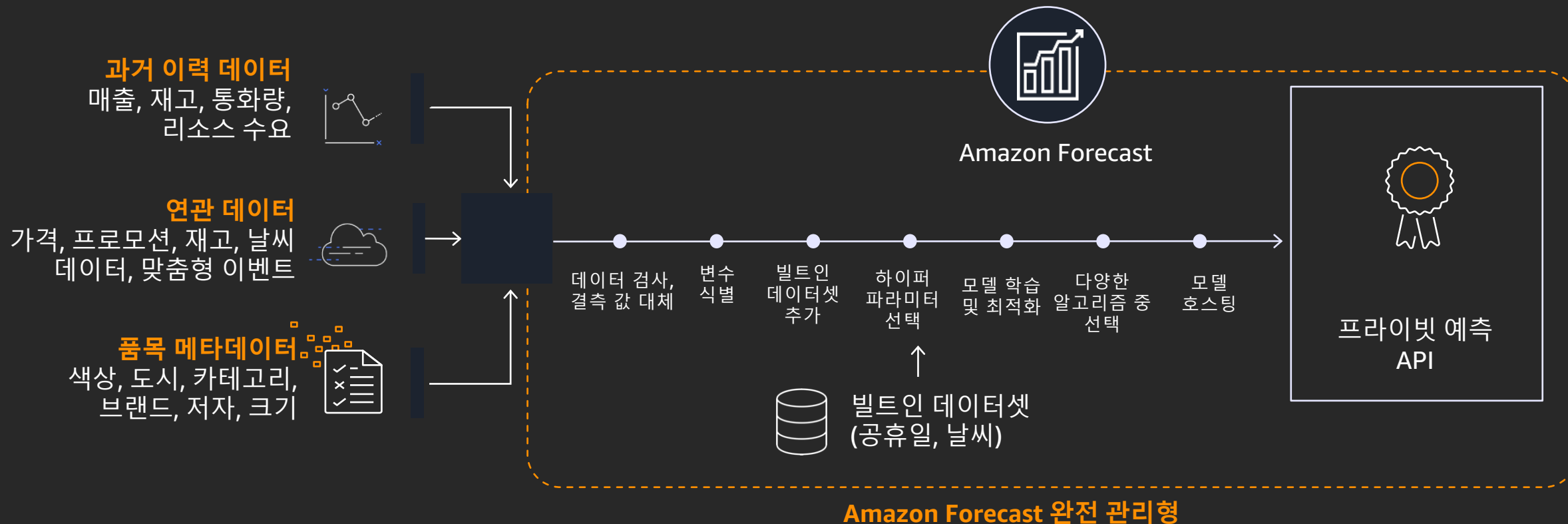


데이터와 모델에 대한 보안 모든 데이터는 비공개로 안전하게 암호화되며 고객을 위한 서비스를 만드는 데만 사용됩니다. 데이터는 다른 고객과, 또는 Amazon.com과 공유되지 않습니다.

Amazon Forecast는 Amazon.com에서 사용하는 예측 기술 기반으로 개발한 완전 관리형 ML 서비스입니다.



Amazon Forecast 어떻게 동작할까?



콘솔 데모

Amazon Forecast

Easily build accurate forecasting models

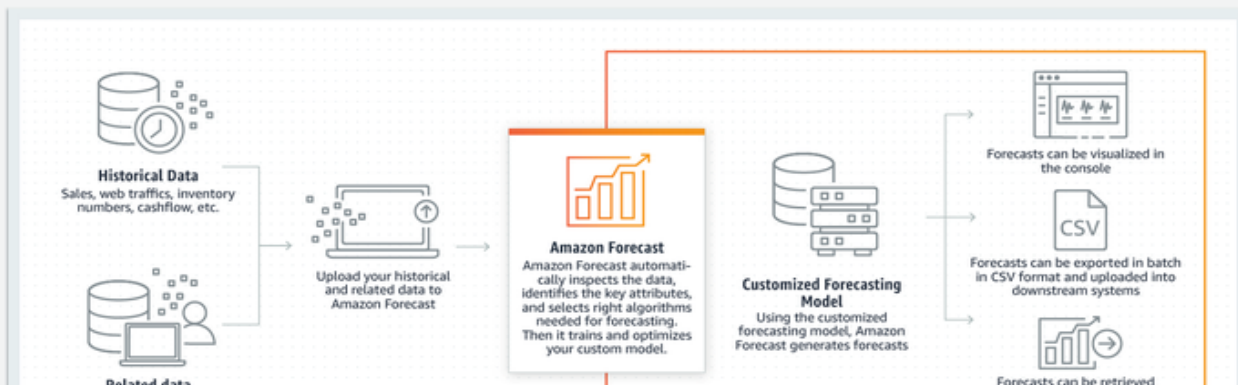
Amazon Forecast is a fully managed service that uses machine learning to deliver highly accurate forecasts. Based on the same machine learning forecasting technology used by Amazon.com.



New feature launch

Amazon Forecast is now up to 40% more accurate, 50% faster to re-train and helps you understand attributes that influence forecasts.

How it works



Create new dataset group

Dataset groups are containers for your datasets, predictors, and forecasts. To create a dataset group, choose Create dataset group.

[Create dataset group](#)

Pricing (US)

Pay only for what you use. There are no minimum fees and no upfront commitments. Different types of charges apply depending on the features of Amazon Forecast that you are using.

[View pricing details](#)

More resources

[Documentation](#)[FAQ](#)

Create dataset group [Info](#)

Dataset group details

Dataset group name

The name can help you distinguish this dataset group from other dataset groups on the dataset groups dashboard.

walmart_sales_dataset_group

The dataset group name must have 1 to 63 characters. Valid characters: a-z, A-Z, 0-9, and _

Forecasting domain [Info](#)

A forecasting domain defines a forecasting use case. You can choose a predefined domain, or you can create your own domain.

Retail

This is a predefined domain for forecasting demand for a retailer.

► Tags - optional [Info](#)

A tag is an administrative label that you assign to AWS resources to make it easier to manage them. Each tag consists of a key and an optional value. Use tags to search and filter your resources or track your AWS costs.

Cancel

Next

1 단계: 데이터셋 검증과 가져오기

AMAZON FORECAST는 3 종류의 데이터셋을 받아들입니다.

타깃 시계열
(과거 수요)

타임스탬프	품목 Id	위치	타깃 값
Jan 1	1111	US_98121	50
Jan 1	2121	US_98121	89
Jan 2	1111	US_98003	35
Jan 2	3434	US_98003	73
Jan 3	1111	US_98003	45
Jan 3	5000	US_43505	13

수요, 매출, 통화량, 에너지 사용량,
클라우드 스토리지 사용량

품목 메타데이터

품목 Id	색상	카테고리
1111	Yellow	Outdoor
2121	Red	Outdoor
3434	Green	Indoor
5000	Blue	Indoor

카테고리, 부문, 색상, 질감

연관 시계열

타임스탬프	품목 Id	가격
Jan 1	1111	\$5
Jan 1	2121	\$20
Jan 2	1111	\$5
Jan 2	3434	\$13
Jan 3	1111	\$5
Jan 3	5000	\$4

가격, 프로모션, 이벤트, 영업시간, 경쟁사 평균가

팁:

1. 데이터 소스를 식별하고 데이터가 시계열인지 확인합니다.
2. 품목 ID와 과거 데이터 포인트를 대량으로 수집합니다
(예: 3년 치의 일일 데이터).
3. 타임스탬프와 빈도가 일관되게 데이터를 정리하여 준비합니다.

1 단계: 연관 시계열과 품목 메타데이터 예시

모델 정확도 향상을 위해 연관 데이터를 추가하고 변환하는 작업은 반복 실험으로 이루어집니다.

사용 사례	연관 시계열	품목 메타데이터
퀵서비스 레스토랑	영업시간	메뉴 분류 (아침, 점심, 저녁)
식료품 소매	가격, 영업시간, 휴일, 프로모션	분야, 매장 규모, 유통 기한, 매장 유형, 주요 소비자 군, 경쟁 매장, 매장 처리량.
제조업	Bloomberg의 경제 지표, 백로그 데이터	마케팅 채널, 분야
하이퍼 로컬 주문형 서비스	공휴일, 날씨	도시의 특징

참고:

- 대부분의 예측력은 과거 수요 데이터에서 나옵니다.
- 연관 데이터를 추가한다고 해서 항상 예측 정확도가 높아지는 것은 아닙니다.
- 먼저 타깃 시계열만으로 모델을 만들고 연관 데이터를 추가하기 전에 정확도 향상의 정도를 평가합니다.
- 연관 데이터를 추가하여 이점을 얻는 것은 피쳐 엔지니어링에 달려 있습니다.

1 단계: 타깃, 연관 시계열과 품목 메타데이터 형상

target_time_series			related_time_series															item_metadata		
item_id	timestamp	demand	item_id	timestamp	Store	Dept	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday	item_id	store_type	store_size	
1_1	2010-02-05	24924.5	1_1	2010-02-05	1	1	42.31	2.572						211.0963582	8.106	0	1_1	A	151315	
1_1	2010-02-12	46039.49	1_1	2010-02-12	1	1	38.51	2.548						211.2421698	8.106	1	1_2	A	151315	
1_1	2010-02-19	41595.55	1_1	2010-02-19	1	1	39.93	2.514						211.2891429	8.106	0	1_3	A	151315	
1_1	2010-02-26	19403.54	1_1	2010-02-26	1	1	46.63	2.561						211.3196429	8.106	0	1_4	A	151315	
1_1	2010-03-05	21827.9	1_1	2010-03-05	1	1	46.5	2.625						211.3501429	8.106	0	1_5	A	151315	
1_1	2010-03-12	21043.39	1_1	2010-03-12	1	1	57.79	2.667						211.3806429	8.106	0	1_6	A	151315	
1_1	2010-03-19	22136.64	1_1	2010-03-19	1	1	54.58	2.72						211.215635	8.106	0	1_7	A	151315	
1_1	2010-03-26	26229.21	1_1	2010-03-26	1	1	51.45	2.732						211.0180424	8.106	0	1_8	A	151315	
1_1	2010-04-02	57258.43	1_1	2010-04-02	1	1	62.27	2.719						210.8204499	7.808	0	1_9	A	151315	
1_1	2010-04-09	42960.91	1_1	2010-04-09	1	1	65.86	2.77						210.6228574	7.808	0	1_10	A	151315	










raw_data/

Copy S3 URI




Objects | Properties

Objects (3)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

  Copy S3 URI  Copy URL  Download  Open  Delete  Actions  Create folder  Upload

 Find objects by prefix

<input type="checkbox"/>	Name ▲	Type ▼	Last modified ▼	Size ▼	Storage class ▼
<input type="checkbox"/>	 item_metadata.csv	csv	January 7, 2022, 20:56:13 (UTC+09:00)	2.1 KB	Standard
<input type="checkbox"/>	 related_time_series.csv	csv	January 7, 2022, 20:56:12 (UTC+09:00)	1.3 MB	Standard
<input type="checkbox"/>	 target_time_series.csv	csv	January 7, 2022, 20:56:12 (UTC+09:00)	483.4 KB	Standard

Create target time series dataset [Info](#)

Dataset details

Dataset name

The name can help you distinguish this dataset from other datasets on your Datasets dashboard.

walmart_sales_target_time_series

The dataset name must have 1 to 63 characters. Valid characters: a-z, A-Z, 0-9, and _

Frequency of your data

This is the frequency at which entries are registered into your data file.

Your data entries have a time interval of

1

▼

day

▼

Data schema [Info](#)

Use the data schema section to specify the attribute types for each column in your dataset. You can specify the schema in two ways:

- ☒ Schema builder
Specify your Attribute Name, Attribute Type, and attribute order in the text boxes provided.
- ☐ JSON schema
Specify AttributeName and AttributeType in the JSON format.

Schema Builder [Info](#)

The attributes below are required for your chosen domain. You may add additional attributes. All attributes displayed must exist in your CSV file and must be ordered in the same order that they appear in your CSV file. To reorder the attributes, simply drag and drop each attribute to the correct position.

Column

1	<div><div><div>⋮</div><div>Attribute Name</div><div>Item_id</div></div><div><div>Attribute Type</div><div>string</div><div>▼</div></div></div>
2	<div><div><div>⋮</div><div>Attribute Name</div><div>timestamp</div></div><div><div>Attribute Type</div><div>timestamp</div><div>▼</div></div><div><div>Timestamp Format Info</div><div>yyyy-MM-dd</div><div>▼</div></div></div>
3	<div><div><div>⋮</div><div>Attribute Name</div><div>demand</div></div><div><div>Attribute Type</div><div>float</div><div>▼</div></div></div>

3



demand

float

**Add attribute**

You can add up to 10 attributes.

Dataset import details

Dataset import name

The name can help you distinguish this dataset import from other imports on your dataset detail page.

walmart_sales_target_time_series_train

The dataset import name must have 1 to 63 characters. Valid characters: a-z, A-Z, 0-9, and _

Select time zone [Info](#)

Select a time zone for your dataset.

Do not use time zone



Data location [Info](#)

The location is the path to the file in your S3 bucket that contains your data.

s3://amazon-forecast-ad0bdfa1-2057-4585-96f4-e934ddc028b6/raw_

View **Browse S3**

Your files must be in CSV format.

IAM role [Info](#)

Dataset groups require permissions from IAM to read your dataset files in S3. Choose or create a role using this control.

AmazonForecast-ExecutionRole-1641470097629



Cancel

Previous

Start



Dashboard [Info](#)

Overview



Import your data

Datasets are required to train predictors, which are then used to generate forecasts.

Target time series data

Create pending data

Item metadata data

[Import](#)

Related time series data

[Import](#)

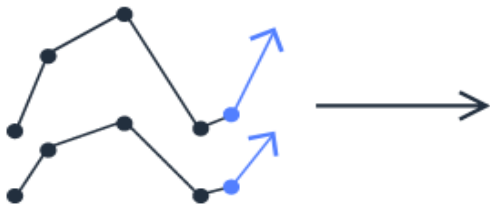


Train a predictor

Train a predictor, a custom model with underlying infrastructure that Amazon Forecast trains on your datasets.

Predictor training

[Start](#)



Generate forecasts

Generate forecasts using your trained predictors.

Forecast generation

[Start](#)



Explore insights

Explore and identify the attributes impacting your forecasts. Insights provide explainability to your forecasts.

Insight generation

[Explore insights](#)

Dataset details

Dataset name

The name can help you distinguish this dataset from other datasets on your Datasets dashboard.

walmart_sales_item_metadata

The dataset name must have 1 to 63 characters. Valid characters: a-z, A-Z, 0-9, and _

Data schema [Info](#)

Use the data schema section to specify the attribute types for each column in your dataset. You can specify the schema in two ways:

☐ Schema builder
Specify your Attribute Name, Attribute Type, and attribute order in the text boxes provided.

☒ JSON schema
Specify AttributeName and AttributeType in the JSON format.

To help Amazon Forecast understand the fields in your data, you must define the schema. Specify the attributes in the same order as they appear in your CSV file. The pre-populated attributes below are required for the chosen domain. You may specify up to 13 attributes.

```
1 {
2   "Attributes": [
3     {
4       "AttributeName": "item_id",
5       "AttributeType": "string"
6     },
7     {
8       "AttributeName": "store_type",
9       "AttributeType": "string"
10    },
11    {
12      "AttributeName": "store_size",
13      "AttributeType": "integer"
14    }
15  ]
16 }
```

Datasets (3) [Info](#)

View details

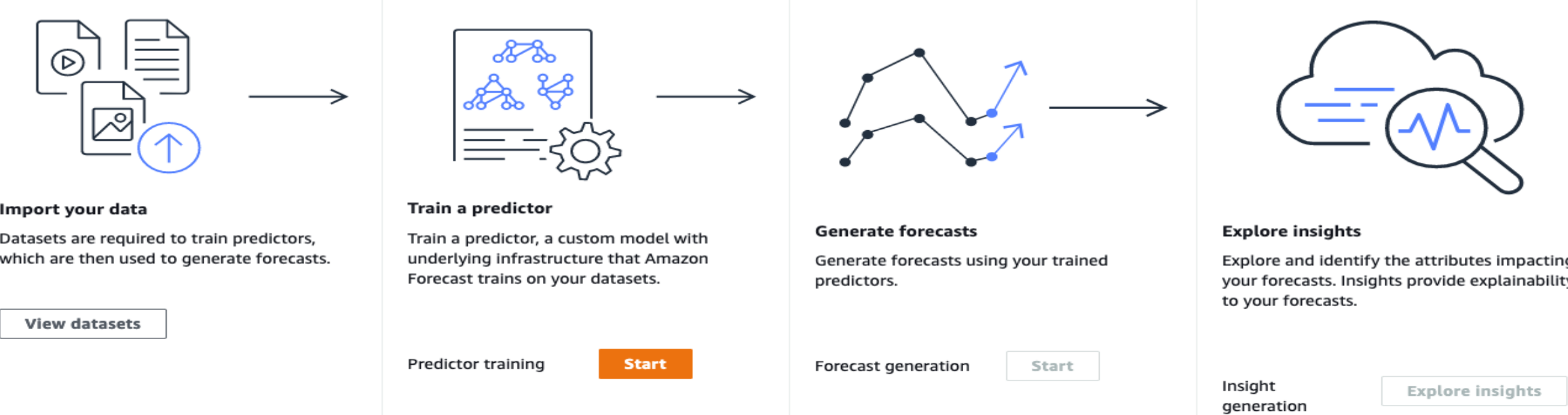
Delete

Upload dataset

	Dataset name ▾	Dataset type ▾	Status ▾	Latest import status	Domain ▾	Date created ▾
<input type="radio"/>	walmart_sales_related_time_series	RELATED_TIME_SERIES	✔ Active	✔ Active	RETAIL	Fri, 07 Jan 2022 12:32:25 GMT
<input type="radio"/>	walmart_sales_item_metadata	ITEM_METADATA	✔ Active	✔ Active	RETAIL	Fri, 07 Jan 2022 12:30:53 GMT
<input type="radio"/>	walmart_sales_target_time_series	TARGET_TIME_SERIES	✔ Active	✔ Active	RETAIL	Fri, 07 Jan 2022 12:22:39 GMT

Dashboard [Info](#)

Overview



Predictor settings

Predictor name

The name can help you distinguish this predictor from your other predictors.

walmart_sales_predictor

The predictor name must have 1 to 63 characters. Valid characters: a-z, A-Z, 0-9, and _

Forecast configuration

Forecast frequency

This is the frequency at which your forecasts are generated.

Your forecast frequency is

1

▼

day

▼

Forecast horizon [Info](#)

This number tells Amazon Forecast how far into the future to predict your data at the specified forecast frequency.

30

⬆

⬇

⬆

Forecast dimensions - optional

Item id is used in training by default. Select additional keys you would like to use to generate a forecast. These keys are fields in your dataset.

Choose a forecast dimension

▼

Forecast quantiles - optional [Info](#)

Specify the quantiles used to create forecasts and evaluate predictors. Choose up to 5 quantiles between 0.01 and 0.99 (by increments of 0.01). You can also include the mean forecast with 'mean'.

Forecast quantiles

	Value	
Forecast quantile 1	0.10	Remove
Forecast quantile 2	0.50	Remove
Forecast quantile 3	0.90	Remove

Add new forecast quantile

You can add up to 2 more forecast quantiles.



Predictor settings

AutoPredictor [Info](#)

AutoPredictor creates predictors by ensembling the optimal combination of algorithms for each time series in your dataset.

☒ Enable AutoPredictor

Optimization metric - optional [Info](#)

Use a specific accuracy metric to optimize your predictor.

WAPE

▼

Enable explainability [Info](#)

☒ Enable explainability

▼ Input data configuration - optional

Weather [Info](#)

Include historical and projected weather information in predictor training to improve forecast accuracy.

☐ Enable weather index

Holidays [Info](#)

Include holidays in predictor training to improve forecast accuracy.

☐ Enable holidays

▼ Advanced configurations [Info](#)

Set advanced configurations for your predictor and forecasts.

Featurizations

```
1  [
2    {
3      "AttributeName": "demand",
4      "Transformations": {
5        "aggregation": "sum",
6        "frontfill": "none",
7        "middlefill": "zero",
8        "backfill": "zero"
9      }
10   }
11 ]
```

bars generally increase forecast values, while attributes with red bars generally decrease forecast values.

For the model level, impact scores measure the effect attributes have on all time series and throughout the entire forecast window.

Learn more [↗](#)

[Forecast Explainability](#)



2 단계: 예측 모델 훈련하기

AUTOML을 통해 데이터에 가장 적합한 모델을 학습합니다. AMAZON FORECAST는 사용자 맞춤형으로 다양한 알고리즘을 제공합니다.

	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)	✓	✓	✗	✗	✗	✗
Accepts the Weather Index built-in featurization	✓	✓	✓	✗	✗	✗
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)	✓	✓	✗	✗	✗	✗

Predictor settings

AutoPredictor [Info](#)

AutoPredictor creates predictors by ensembling the optimal combination of algorithms for each time series in your dataset.

☒ Enable AutoPredictor

Optimization metric - optional [Info](#)

Use a specific accuracy metric to optimize your predictor.

WAPE

▼

Enable explainability [Info](#)

☒ Enable explainability

▼ Input data configuration - optional

Weather [Info](#)

Include historical and projected weather information in predictor training to improve forecast accuracy.

☐ Enable weather index

Holidays [Info](#)

Include holidays in predictor training to improve forecast accuracy.

☐ Enable holidays

▼ Advanced configurations [Info](#)

Set advanced configurations for your predictor and forecasts.

Featurizations

```
1  [
2    {
3      "AttributeName": "demand",
4      "Transformations": {
5        "aggregation": "sum",
6        "frontfill": "none",
7        "middlefill": "zero",
8        "backfill": "zero"
9      }
10   }
11 ]
```

bars generally increase forecast values, while attributes with red bars generally decrease forecast values.

For the model level, impact scores measure the effect attributes have on all time series and throughout the entire forecast window.

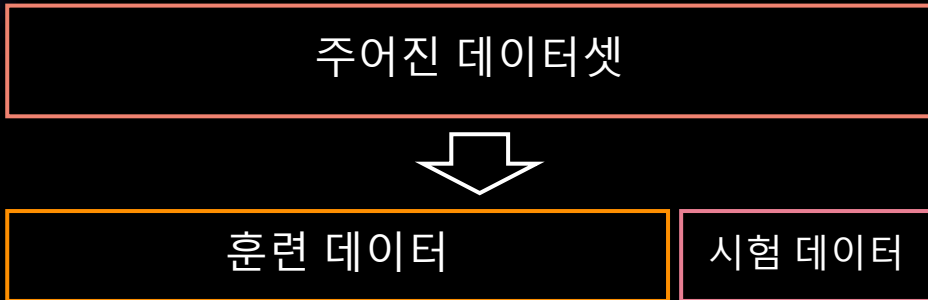
Learn more [↗](#)

[Forecast Explainability](#)



백테스팅을 통한 정확도 자동 측정

AMAZON FORECAST는 훈련 및 테스트 셋 데이터로 자동 분할합니다.



• Forecast는 다양한 모델 정확도 측정 항목을 제공합니다.

1. 가중 분위수 손실(wQL) 메트릭 - 분포의 여러 지점에서 계산
2. 가중 절대 백분율 오차(WAPE) - 평균으로 계산
3. RMSE(제공 평균 제공근 오차) - 평균에서 계산됨

참고: 값이 낮을수록 오차가 작으므로 모델이 더 정확함을 나타냅니다.

사용 사례	wQL	WAPE	RMSE
상이한 영향도를 가질 수 있는 과소 또는 과대 예측에 대해 최적화하는 경우	0		
인기 상품이나 수요가 높은 상품이 수요가 적은 상품보다 중요한 경우	0	0	
예측 오류의 큰 편차와 관련된 비즈니스 비용이 강조되는 경우			0
과거 데이터 포인트에서 대부분의 품목에 대한 수요가 0인 희소 데이터셋을 평가하는 경우	0	0	

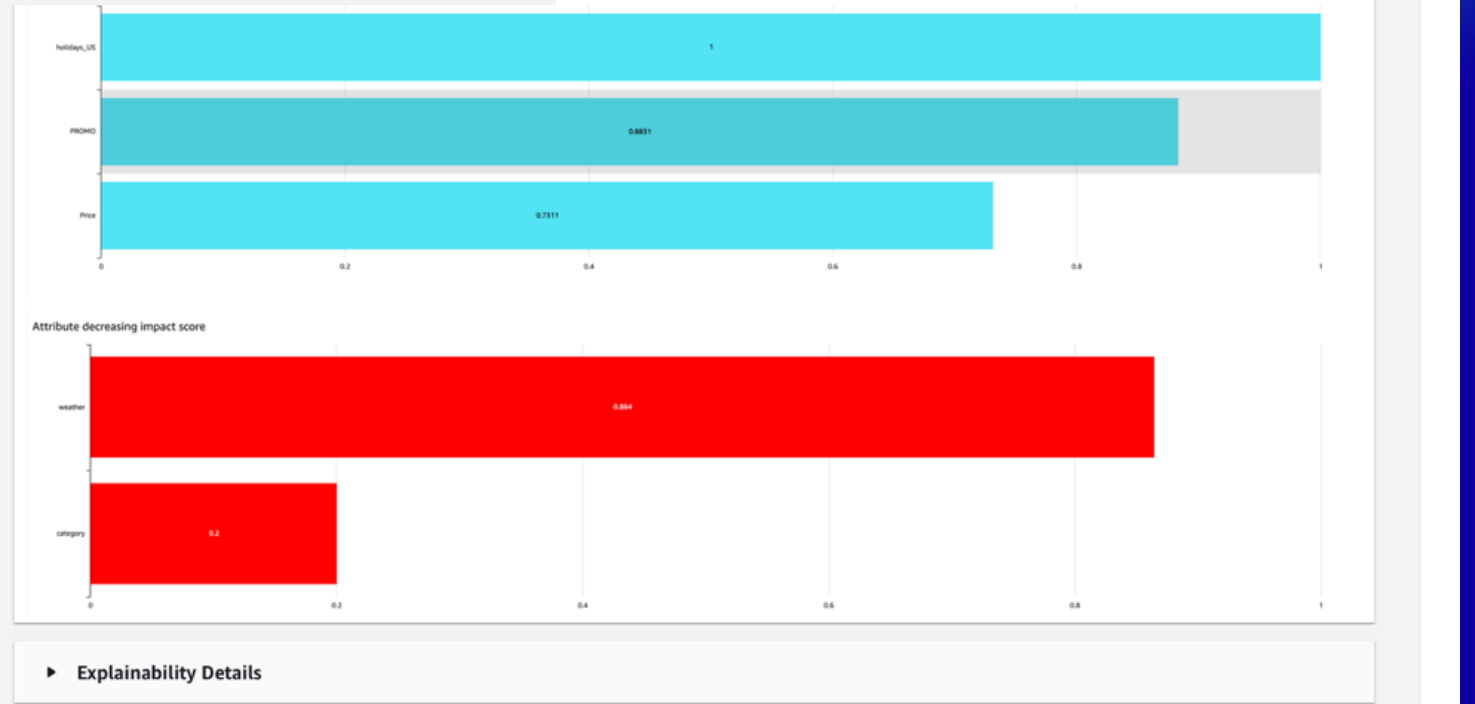
신규 Amazon Forecast API를 통한 속성 별 영향도 점수

Amazon Forecast > Dataset groups > walmart_sales_dataset_group > Predictors

Predictors (1) [Info](#) [Manage notifications](#) [Stop](#) [Retrain](#) [Delete](#) [Create new forecast](#) [Train new predictor](#)

This page gives you an overview of your predictors, and provides information about their status and performance.

	Predictor name	Training status	Forecast types	WAPE	RMSE	AutoPredictor Info	Date created
<input type="radio"/>	walmart_sales_predictor	Active	0.1 0.5 0.9	0.0863	1422.7232	True	Fri, 07 Jan 2022 13:55:59 GMT



Overview



Import your data

Datasets are required to train predictors, which are then used to generate forecasts.

[View datasets](#)

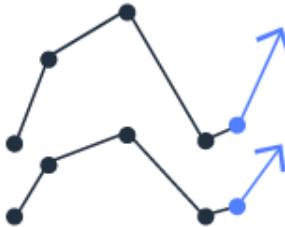


Train a predictor

Train a predictor, a custom model with underlying infrastructure that Amazon Forecast trains on your datasets.

[View predictors](#)

[Train predictor](#)



Generate forecasts

Generate forecasts using your trained predictors.

Forecast generation

[Start](#)



Explore insights

Explore and identify the attributes impacting your forecasts. Insights provide explainability to your forecasts.

Insight generation

[Explore insights](#)

Forecast details

Forecast name

The name can help you distinguish this forecast from your other forecasts.

walmart_sales_forecast

The forecast name must have 1 to 63 characters. Valid characters: a-z, A-Z, 0-9, and _

Predictor [Info](#)

The predictor that you want to use to create forecasts.

walmart_sales_predictor

Forecast quantiles - optional [Info](#)

Enter up to 5 quantile values between 0.01 to 0.99 including 'mean'. By default, Amazon Forecast will generate forecasts for 0.10, 0.50 and 0.90 quantiles.

0.10, 0.50, 0.90, 0.99, mean

Separate forecast types with commas.

► Tags - optional [Info](#)

A tag is an administrative label that you assign to AWS resources to make it easier to manage them. Each tag consists of a key and an optional value. Use tags to search and filter your resources or track your AWS costs.

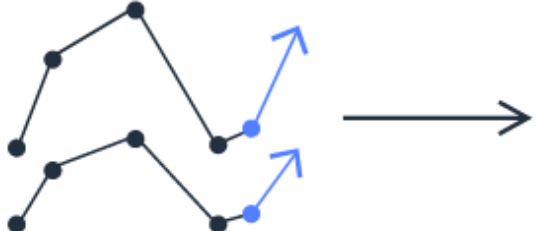
Cancel

Start



3 단계: 모델 배포 및 예측 값 생성하기

예측 값을 손쉽게 생성하고 CSV 파일로 내보냅니다. AWS 서비스를 이용하여 배포합니다.

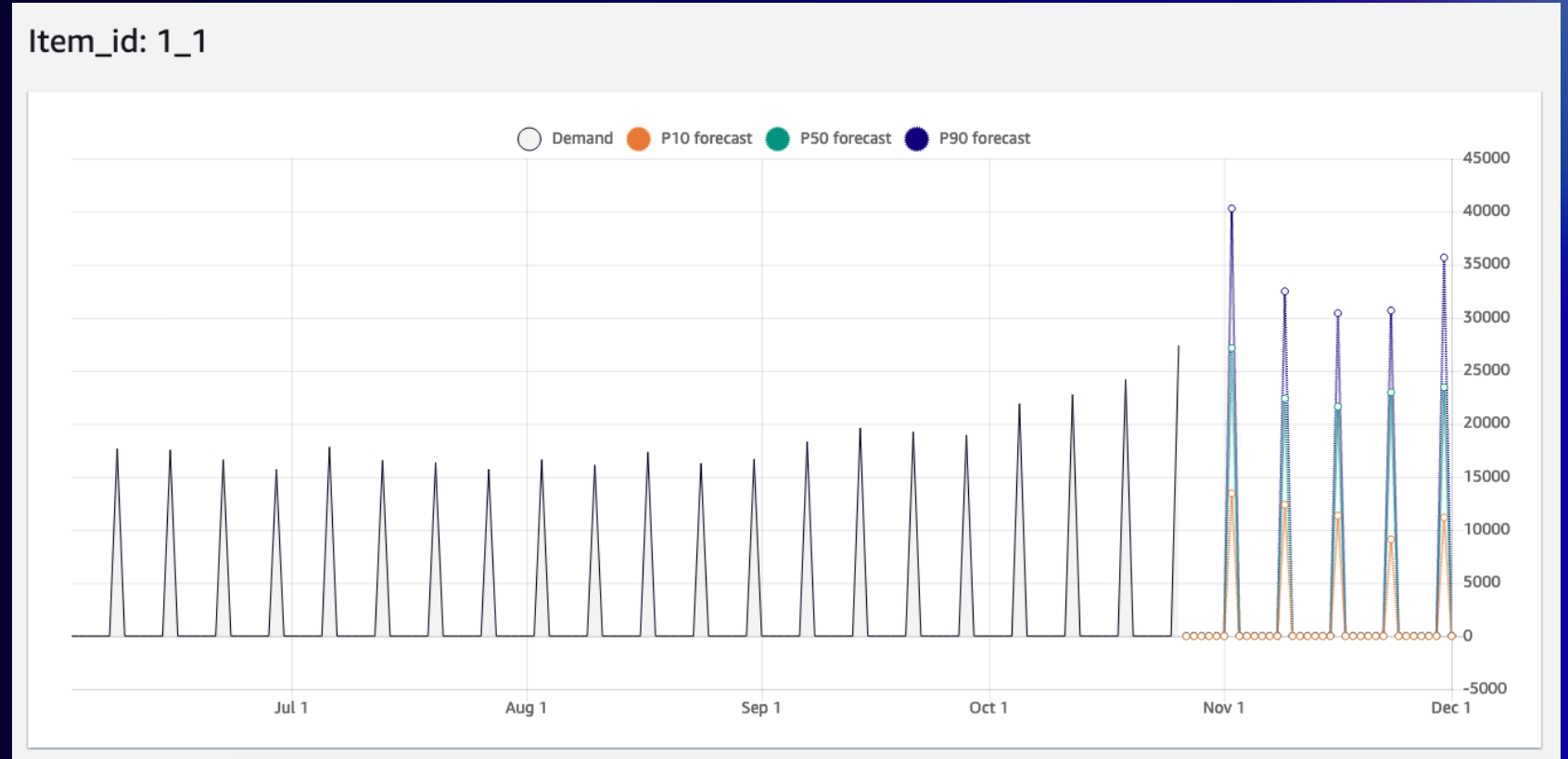


Generate forecasts

Generate forecasts using your trained predictors.

Create a forecast

Query forecast



Forecast API를 사용하는 노트북 예제

GITHUB 리포지토리에 작업 시작을 위한 많은 노트북 예제가 있습니다.

<https://github.com/aws-samples/amazon-forecast-samples>

1. 기본 튜토리얼
2. AutoML 시작하기
3. 연관 시계열 데이터 통합하기
4. 품목 메타 데이터 통합하기
5. 품목 수준에서 정확도 평가하기
6. Amazon Forecast 날씨 지표 사용하기
7. 콜드 스타트 품목 예측하기
8. What-if 분석 수행하기
9. 다양한 모델 비교해보기

AI & ML 리소스 허브

AWS가 제공하는 AI 및 ML에 관한 다양한 자료들을 통해 더욱 심층적으로 학습해보세요!

- 기계 학습 여정 가이드
- 기계 학습의 7가지 주요 사용 사례
- 데이터, 분석 및 기계 학습을 위한 전략 플레이북
- 올바른 클라우드 서비스 및 인프라를 통한 기계 학습 혁신 가속화 전략 가이드
- 기계 학습에 적합한 컴퓨팅 인프라 선택 가이드
- 컨택트 센터의 서비스 개선 및 비용 절감 방법
- + 외의 다양한 동영상 학습 자료 및 기술 학습 자료!



<https://bit.ly/3yUk0Kx>

리소스 허브 방문하기

AWS Innovate - AI/ML 특집에 참석해주셔서 대단히 감사합니다.

저희가 준비한 강연, 어떻게 보셨나요?
더 나은 세미나를 위하여 **설문을 꼭 작성해 주세요!**



aws-korea-marketing@amazon.com



twitter.com/AWSKorea



facebook.com/amazonwebservices.ko



youtube.com/user/AWSKorea



linkedin.com/company/amazon-web-services



twitch.tv/aws

Thank you!