
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

AIM327-R

Data preparation tips for highly accurate time-series forecasts

Adarsh Singh

Sr. Software Development Engineer
Amazon Web Services

Charles Laughlin

Principal AI/ML Specialist
Amazon Web Services

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

- Introduction to Amazon Forecast as a fully-managed machine learning (ML)-forecasting service
- How to prepare your input data
- Iterative process of building better models
- Advanced preparation strategies
- Q&A
- Additional resources

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Introduction to Amazon Forecast

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Forecast is a fully managed ML service

CREATE ACCURATE ML-BASED FORECASTS, WITHOUT ANY ML EXPERTISE

Historical data
Sales, inventory, call volume, resource demand

Related data
Price, promotions, in-stock, weather data, custom events

Item metadata
Color, city, category, brand, author, size



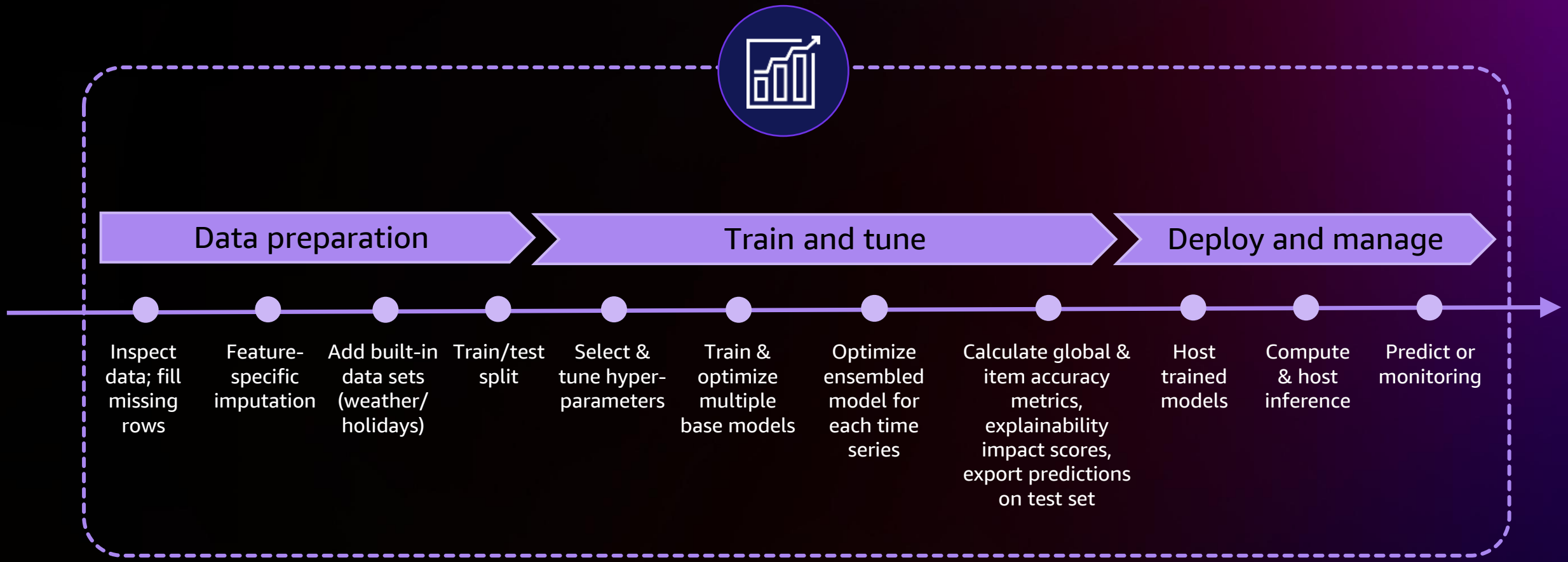
Predictions
Easy to consume predictions in open CSV or parquet format, for each item at every time-step

Built-in AWS curated datasets
(holidays, weather)

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Forecast addresses the entire ML lifecycle

PURPOSE-BUILT FOR TIME-SERIES FORECASTING SO YOU CAN BE MORE PRODUCTIVE



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customers want to know: Where do I start?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Target time series (TTS)

REQUIRED DATA THAT CONTAINS WHAT YOU ARE TRYING TO PREDICT

Optional, up to 10

item_id	Demand	Timestamp	location_id
1001A	177	2022-11-08	US98121
1002A	270	2022-11-08	US98121
1003B	189	2022-11-08	US10001
1004D	54	2022-11-08	US94016
1005C	40	2022-11-08	US88901

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Related time series (RTS)

OPTIONAL DATA THAT HELPS IMPROVE AND EXPLAIN PREDICTIONS

Up to 13 numeric features (x-value multivariate)

Matched TTS dimensions

item_id	Timestamp	checkout_price	regular_price	Email promotion	Homepage promotion	location_id
1001A	2022-01-29	131.95	182.36	0	1	US98121
1002A	2022-10-16	157.14	157.14	0	0	US98121
1003B	2022-10-30	157.17	156.17	1	1	US10001
1004D	2022-12-05	159.08	182.39	0	0	US94016
1005C	2022-03-21	159.08	183.33	1	0	US88901

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Item metadata (IM)

OPTIONAL DATA PROVIDING CATEGORICAL FEATURES

Up to 9 categorical (string)

item_id	Category	Subcategory
1001A	Beverages	Soda/cola
1002A	Beverages	Whole milk
1003B	Fruit	Red apple
1004D	Fruit	Green apple
1005C	Vegetable	Potato

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Feature mapping exercise

CATEGORIZE THE COLUMNS AS TTS, RTS, OR IM DATASET

Item	sale_price	Color	shelf_life	Demand	Promotions	location_id	Department	Timestamp
1	3	Black	120	177	1	A55	Beverages	2022-01-29
2	8	Orange	120	270	0	A55	Beverages	2022-10-16
3	5	Red	120	189	1	A55	Fruits	2022-10-30
4	5	Blue	200	54	0	A55	Fruits	2022-12-05
5	6	Green	300	40	0	A55	Vegetables	2022-03-21

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Iterative flywheel to create a predictor

PREDICTORS ARE SAVED MACHINE LEARNING MODELS USED TO CREATE FORECASTS



Create a predictor

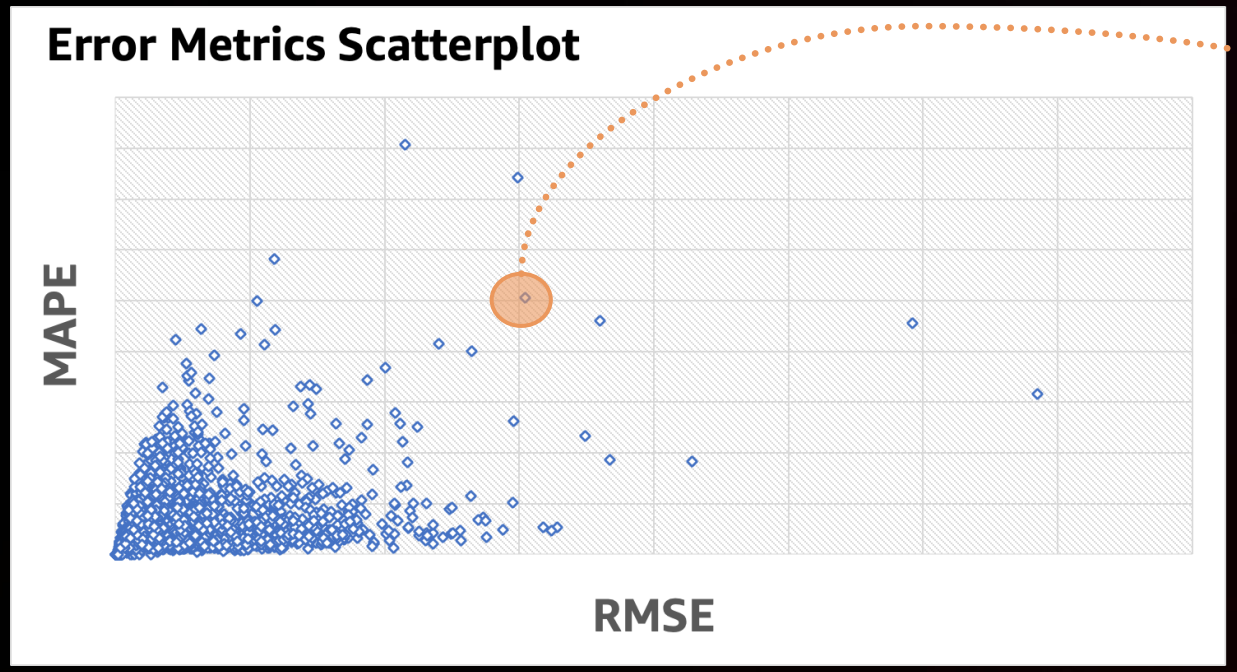
Review predictor metrics

Research, propose & create input data

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Opportunity for additional data signals

DOWNLOAD BACKTEST FORECASTS FOR CUSTOM ACCURACY CALCULATIONS & DATA DISCOVERY



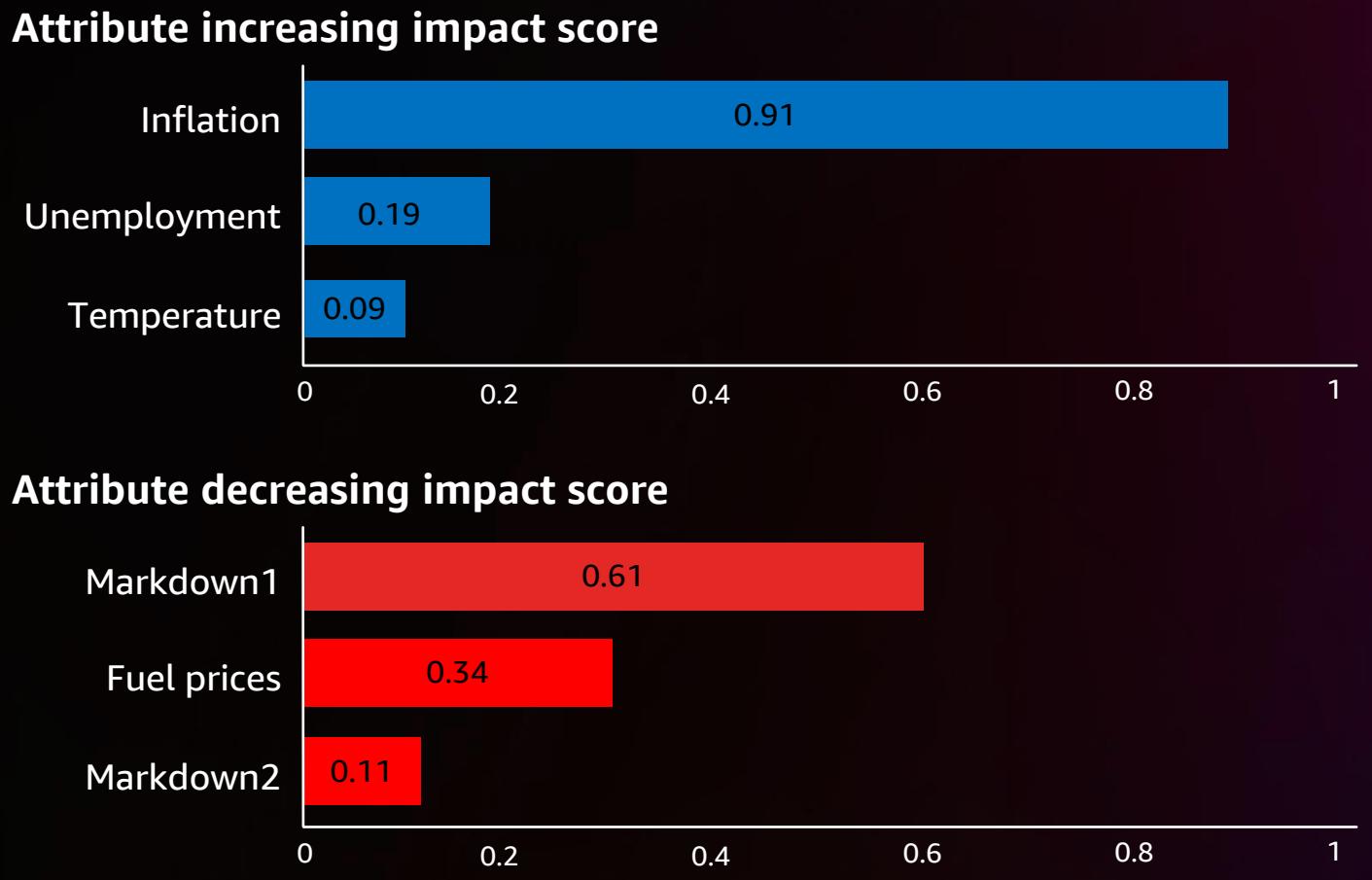
center_id	item_id	date	target	p50	p90
146	1878	6/11/19	350	309.7092	431.9833
146	1878	6/18/19	283	328.7649	449.3195
146	1878	6/25/19	338	338.9323	479.7985
146	1878	7/2/19	339	323.8968	460.5366
146	1878	7/9/19	0	302.3901	439.0211
146	1878	7/16/19	0	292.0136	432.3422
146	1878	7/23/19	285	300.6408	435.9432
146	1878	7/30/19	285	298.8406	438.3807
146	1878	8/6/19	257	300.5123	452.0139
146	1878	8/13/19	312	305.263	438.4845

Use outliers to ask “why?” and identify new input data sources that teach the model

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Improving accuracy overall

USE THE GUIDANCE TO ADJUST INPUT FILES AND CREATE BETTER MODELS



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Advanced data preparation strategies

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

I have missing and sparse data. What do I do?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

My dataset has outliers. How do I handle them?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Learn more about Amazon Forecast

AIM333

How Taco Bell is improving digital availability with ML forecasting

Thursday December 1, 2022, 2:45 PM

Level 3, Chairman’s ballroom, 368, MGM Grand

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Interested in shortening your innovation cycle?



Follow the same no-code MLOps process as Taco Bell and others

Visit our **Amazon Forecast GitHub** to get started today

For additional info, contact Dan Sinnreich
Email: dansinnr@amazon.com

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Adarsh Singh
singhada@amazon.com

Charles Laughlin
czlaugh@amazon.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.