aws
INNOVATE
DATA AND AI/ML EDITION
22 February 2023

aws INNOVATE
DATA AND AI/ML EDITION

Build accurate models combining diverse data types with AutoGluon on Amazon SageMaker

Seema Gupta
Senior Solutions Architect
Amazon Web Services

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

Agenda

- Why AutoML?
- AutoGluon for AutoML
- AutoGluon - architectural approach
- Multimodal learning in machine learning
- Amazon SageMaker
- Demo - example use case with Amazon SageMaker

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

AutoML

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

Machine learning & AutoML

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

AutoGluon for AutoML

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

AutoGluon – AutoML framework

- AutoGluon is an AutoML Toolkit that **democratizes machine learning**
- Automatically builds accurate models for image, text, tabular data & multimodal with **a single line of code**
- Open source & easy to customize
  - pip install AutoGluon

auto.gluon.ai

aws

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

Why AutoGluon ?

	Framework	Wins	Losses	Failures	Champion	Rank	Performance	Time
OpenML	AutoGluon	-	-	**1**	**23**	**1.8438**	**0.8615**	201
	H2O AutoML	4	26	8	2	3.1250	0.7553	220
	TPOT	6	27	5	5	3.3750	0.7966	235
	GCP-Tables	5	20	14	4	3.7500	0.6664	**195**
	auto-sklearn	6	27	6	3	3.8125	0.6803	240
	Auto-WEKA	4	28	6	1	5.0938	0.1999	244
Kaggle	AutoGluon	-	-	**0**	**7**	**1.7143**	**0.7041**	**202**
	GCP-Tables	3	7	1	3	2.2857	0.6281	222
	H2O AutoML	1	7	3	0	3.4286	0.5129	227
	TPOT	1	9	1	0	3.7143	0.4711	380
	auto-sklearn	3	8	**0**	1	3.8571	0.4819	240
	Auto-WEKA	0	10	1	0	6.0000	0.2056	221

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

AutoGluon ranks in tabular + text ML competition leaderboards

Kaggle

- 2nd out of 2742 (Mercari Price Suggestion – 2380 teams with $100K price)
- 2nd out of 172 (California House Prices)
- 20th out of 3537 (PetFinder Pawpularity Contest) – 2022

MachineHack

- 1st out of 1544 (Predict Data Scientist Salary in India)
- 1st out of 314 (Product Sentiment Classification)
- 2nd out of 3922 (Predict the Price of Books)

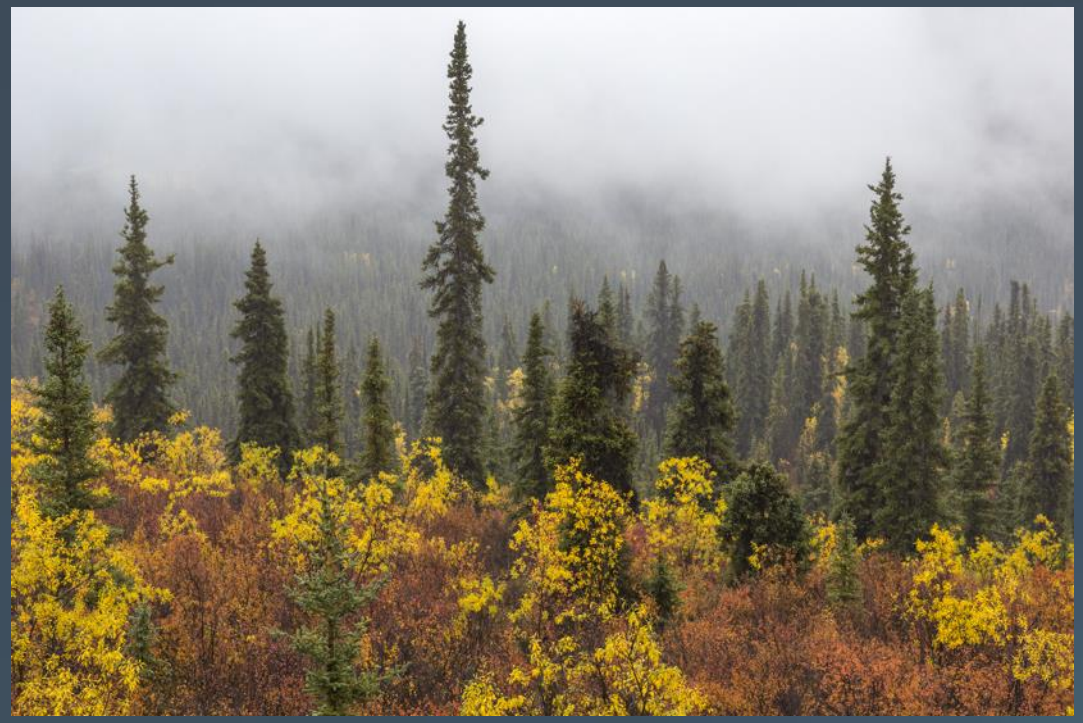
aws

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

AutoGluon – architectural approach

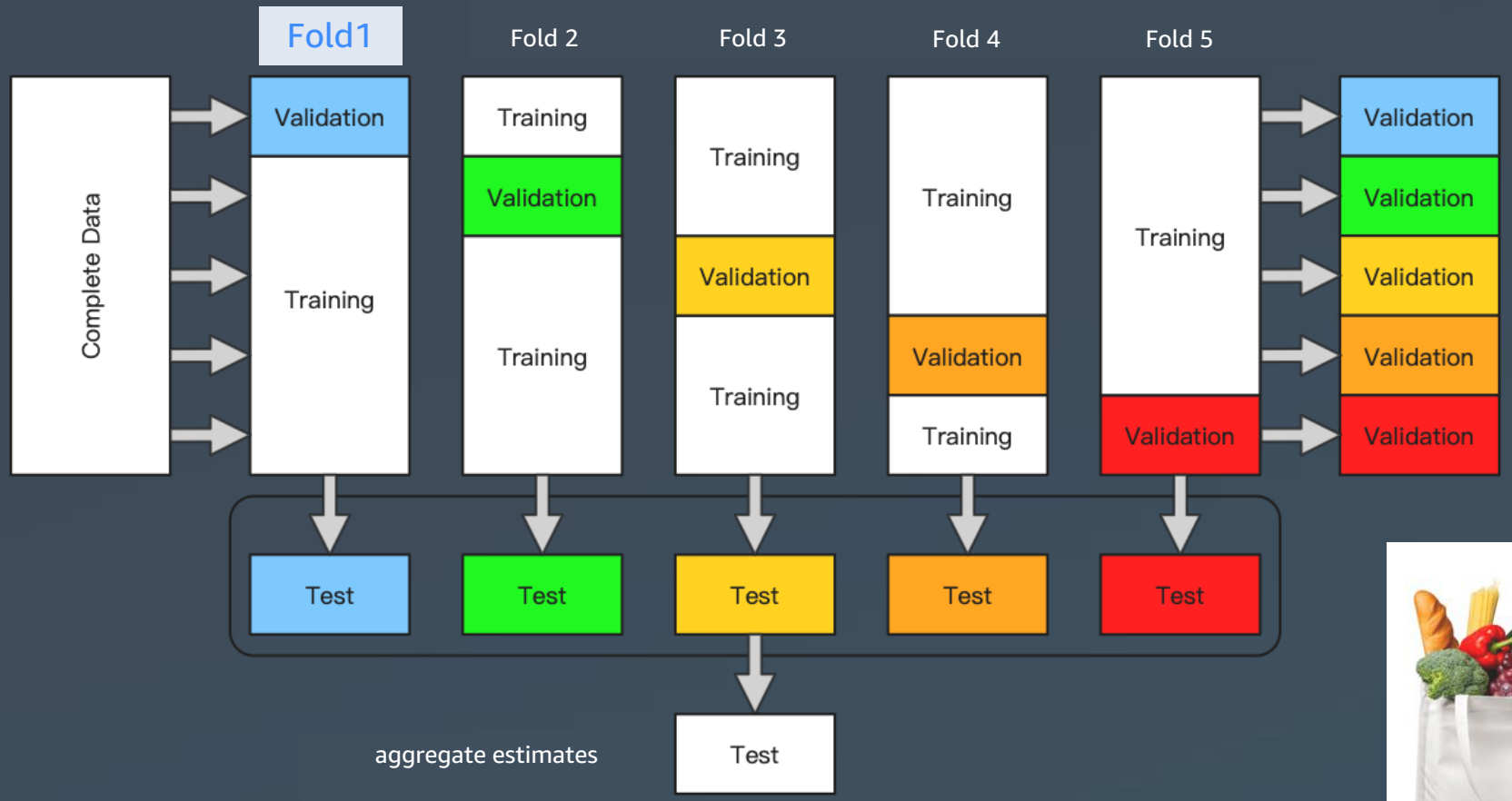
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Decision tree & random forest



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

Bagging



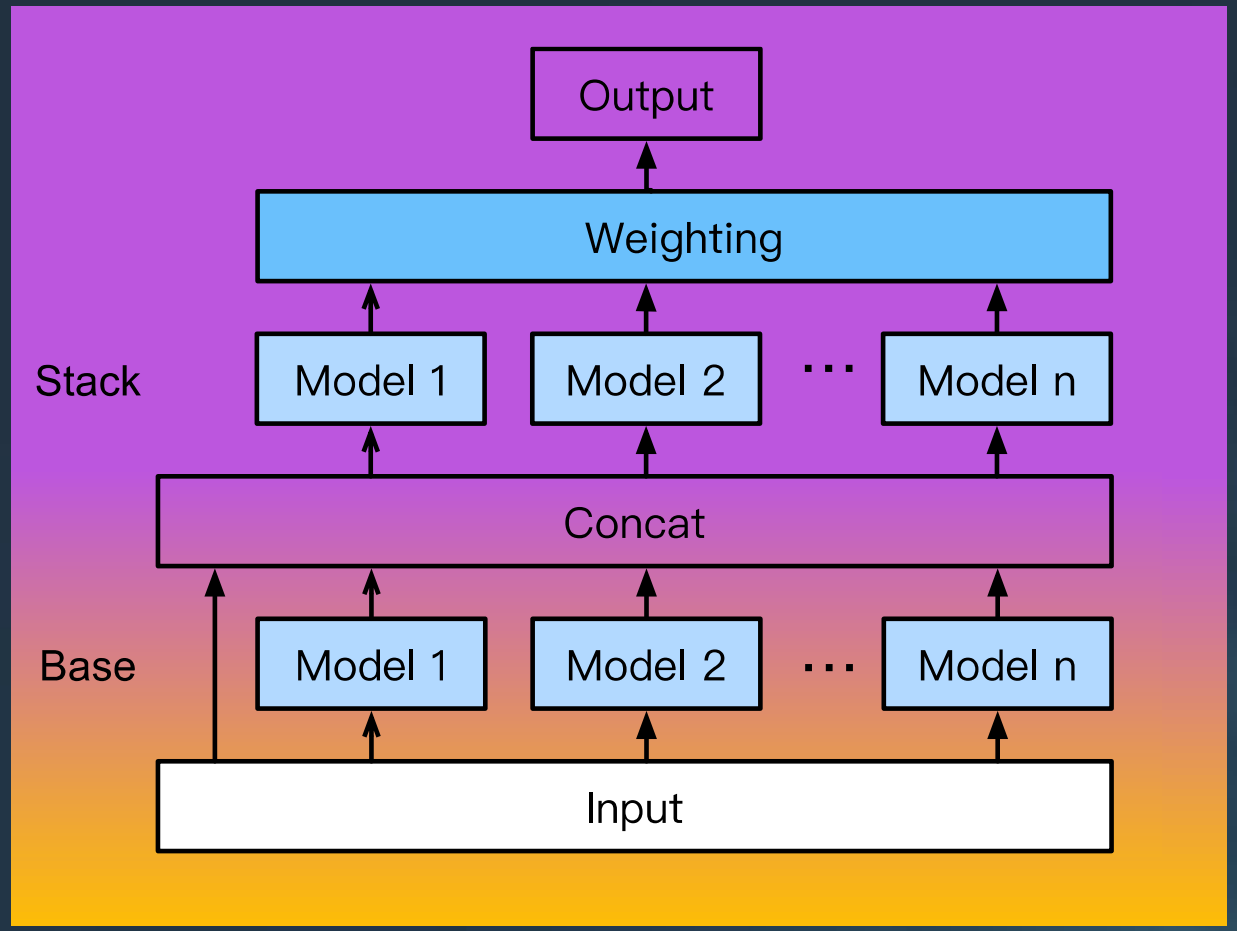
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-layer stack ensembling

Stacker model uses predictions of every base model as extra feature

Layer *L+1* stacker model uses layer *L* predictions as extra feature

For simplicity: stacker model types = base model types



Stacker must be trained with held-out predictions of lower-layer models

aws
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AutoGluon models

Large number of models implemented and optimized out of the box

Powerful Gradient Boosting Models
XGBoost, LightGBM, CatBoost
Custom Tabular Neural Networks
sklearn Random Forest, Extra Trees, KNN
SOTA Image Models: EfficientNet, ResNet, etc.
any TIMM compatible model
SOTA Text Models: BERT, Electra, etc.
any HuggingFace compatible model!

aws
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AutoGluon is customizable

1. Bring your own custom model to be trained, hyperparameter-tuned, and ensembled alongside the default models
2. Add your own custom evaluation metrics
3. Choose the metric to optimize for
4. Define training time
5. Control inference speed

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

Multimodal learning in machine learning

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

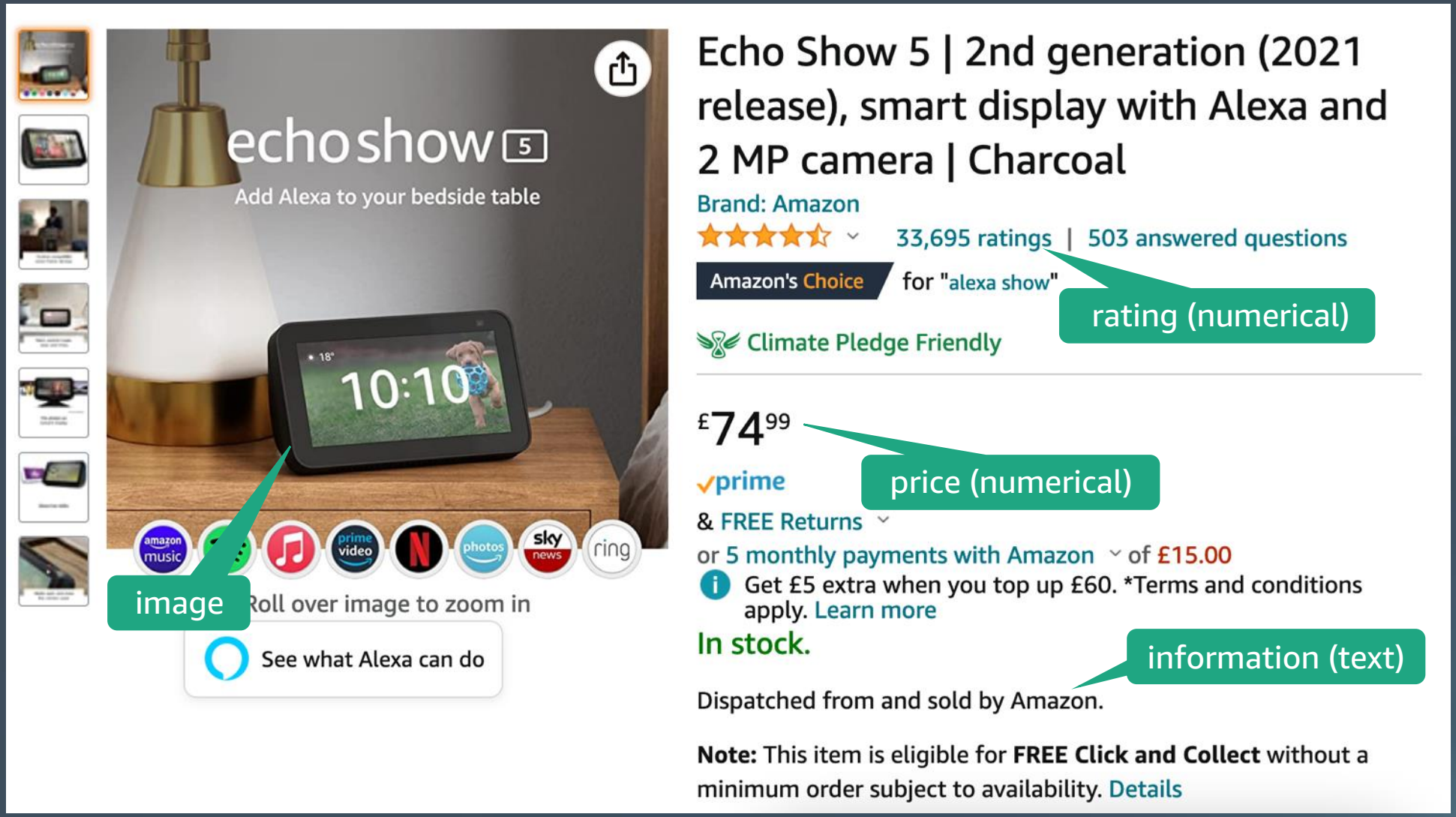
What is multimodal learning?

- Multimodal is the science of heterogenous and interconnected data.

- Modality refers to the way in which something is expressed or perceived.

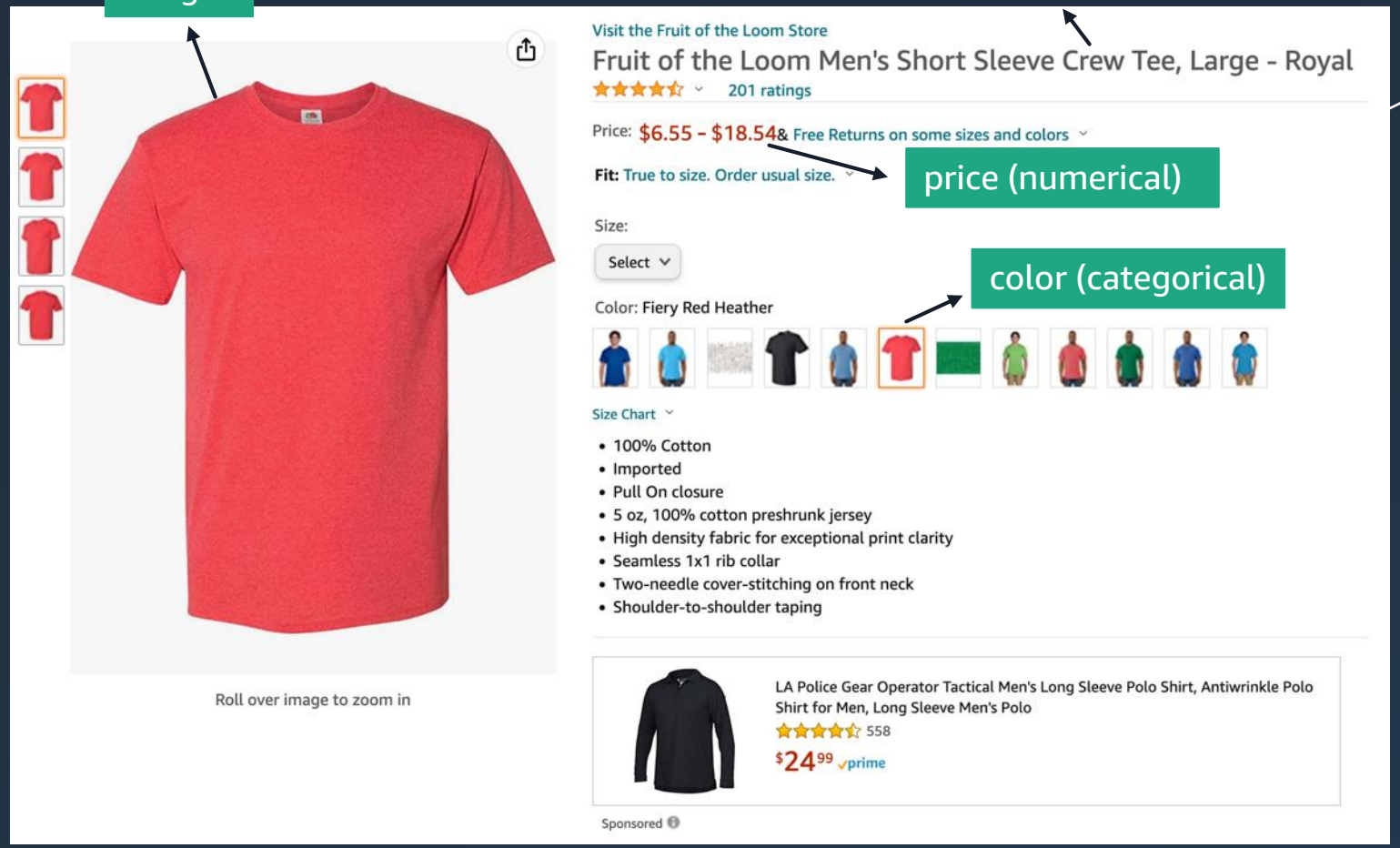
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data as represented in real world



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

Multimodal problems: same product or not?



Not the same product

Uploaded Image

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

Multimodal DataFrame

	Id	Sold Price	Summary	images	Year built	Parking	Bedrooms	Bathrooms	Full bathrooms	Total interior livable area	...	Cooling features	Appliances included	Parking features	Construction materials	Tax assessed value	Annual tax amount	Laundry features
0	19633278	3308000.0	Located in the desirable Highlands neighborhoo...	kitchen.jpg	1959.0	Garage, Garage - Attached, Covered	4	5.0	4.0	2832.0	...	Central	Dishwasher, Garbage disposal	Garage, Garage - Attached, Covered	wood frame	3200000.0	38898.0	Inside
1	19687145	730000.0	Move-in ready second floor unit with vaulted c...	kitchen.jpg	1987.0	Carport, Garage	2	2.0	NaN	871.0	...	Central	Dishwasher, Dryer, Garbage disposal, Microwave...	Carport, Garage	wood frame	590000.0	7470.0	NaN
2	19606535	1450000.0	Great opportunity to make this home uniquely y...	kitchen.jpg;bath.jpg;garage.jpg	1963.0	Garage, Garage - Attached, Covered	Master Bedroom on Ground Floor, More than One ...	2.0	2.0	1466.0	...	None	Dishwasher, Garbage disposal, Range / Oven	Garage, Garage - Attached, Covered	wood frame	1326000.0	18589.0	In Garage, Gas Hookup
3	19807250	883000.0	Remodeled and well-maintained 4 bedrooms/2 bat...	garage.jpg;bath.jpg;living_room.jpg	1979.0	Garage, Garage - Attached, Covered	4	2.0	2.0	1325.0	...	Central	NaN	Garage, Garage - Attached, Covered	wood frame	852000.0	12200.0	In Garage
4	15512590	1750000.0	4BR/2.5BA CUSTOM DESIGNED HOME ALA SEA RANCH D...	garage.jpg	1978.0	Garage, Garage - Attached, Covered	4	3.0	2.0	3100.0	...	None	Dishwasher, Dryer, Garbage disposal, Microwave...	Garage, Garage - Attached, Covered	NaN	1785000.0	20024.0	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
39402	2079493953	165000.0	88 Elliot St, Tubac, AZ 85646 is a vacant land...	living_room.jpg;yard.png	NaN	0 spaces	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN
39403	153172307	1550000.0	936 Washington St, Mountain View, CA 94043 is ...	living_room.jpg	1925.0	Garage, Garage - Attached, Covered	NaN	0.0	NaN	1840.0	...	None	NaN	Garage, Garage - Attached, Covered	NaN	781592.0	9678.0	NaN
39404	19836851	1320000.0	6373 Tamalpais Ave, San Jose, CA 95120 is a si...	kitchen.jpg;bath.jpg	1966.0	Garage, Garage - Attached, Covered	4	2.0	NaN	1862.0	...	NaN	NaN	Garage, Garage - Attached, Covered	wood frame	98605.0	2386.0	NaN
39405	19583409	970000.0	930 Poplar St, Santa Clara, CA 95050 is a sing...	kitchen.jpg	1900.0	Garage, Garage - Attached,	3	1.0	NaN	1602.0	...	None	NaN	Garage, Garage - Attached,	wood frame	67326.0	945.0	NaN

Raw text

Bag of images

Example: house price prediction

Label

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

Finding a home for 'Darling': What is the time to adoption?

Description: Darling was born at the back lane of Jalan Alor and was fostered by a feeder. All his siblings had died of accident. His mother and grandmother had just been spayed. Darling make a great condo/apartment cat. He loves to play a lot. He would make a great companion for someone looking for a cat to love.

Type	2
Name	Darling
Age	4
Breed1	266
Breed2	0
Gender	1
Color1	1
Color2	0
Color3	0
MaturitySize	2
FurLength	1
Vaccinated	2
Dewormed	2
Sterilized	2
Health	1
Quantity	1
Fee	0
State	41401
RescuerID	01f954cdf61526daf3fbeb8a074be742
VideoAmt	0
Description	Darling was born at the back lane of Jalan Alo...
PetID	a0c1384d1
PhotoAmt	5.0
AdoptionSpeed	3
Images	/var/lib/jenkins/workspace/workspace/autogluon...

Source: https://auto.gluon.ai/0.3.1/tutorials/tabular_prediction/tabular-multimodal.html

aws

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

AutoGluon multimodal predictor (AutoMM)

```
from autogluon.multimodal import MultiModalPredictor

predictor = MultiModalPredictor(label=label_col)

predictor.fit( train_data=train_data, time_limit=120, # seconds )
```

Reference : https://auto.gluon.ai/stable/tutorials/multimodal/multimodal_prediction/beginner_multimodal.html

aws

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

Evaluation & prediction

```
scores = predictor.evaluate(test_data, metrics=["roc_auc"])

predictions = predictor.predict(test_data.drop(columns=label_col))

probas = predictor.predict_proba(test_data.drop(columns=label_col))
```

Reference : https://auto.gluon.ai/stable/tutorials/multimodal/multimodal_prediction/beginner_multimodal.html

aws

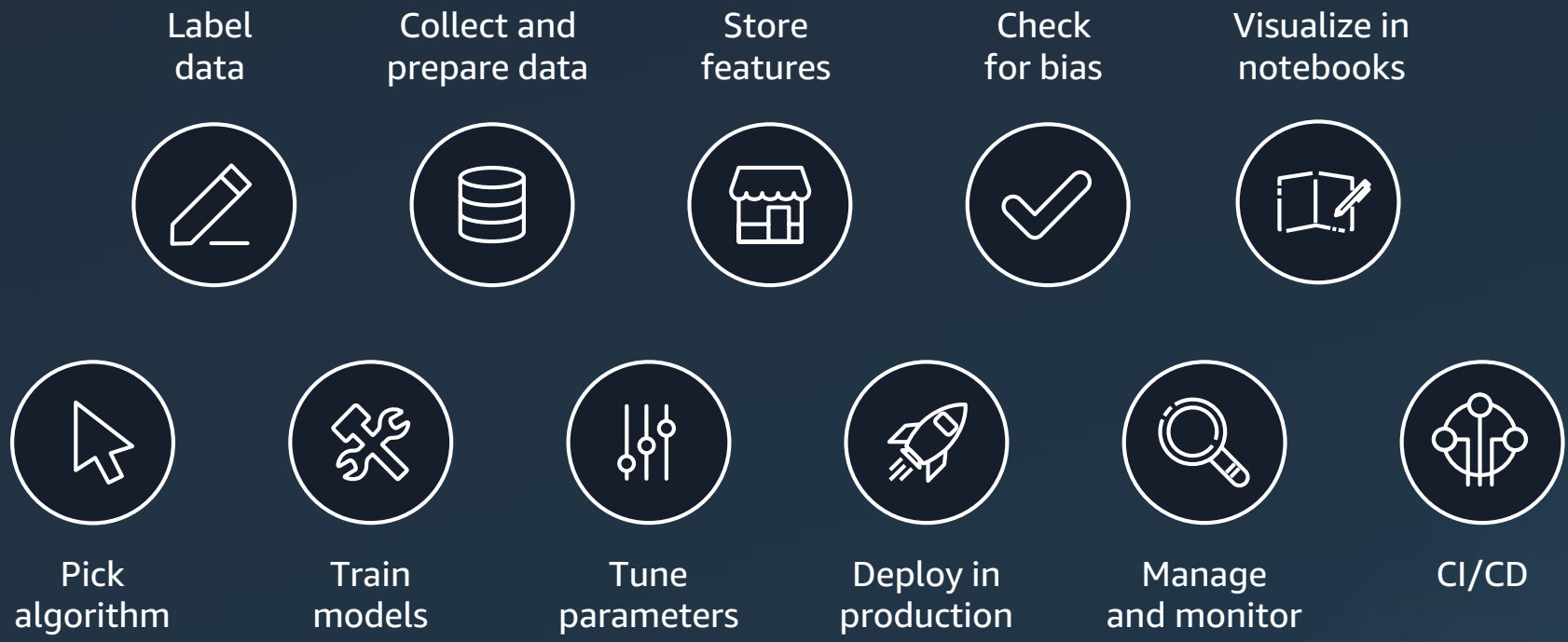
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker

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

Amazon SageMaker: makes ML more accessible

Amazon SageMaker Studio IDE



Model management for edge devices

aws

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

Running AutoMM on Amazon SageMaker demo

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

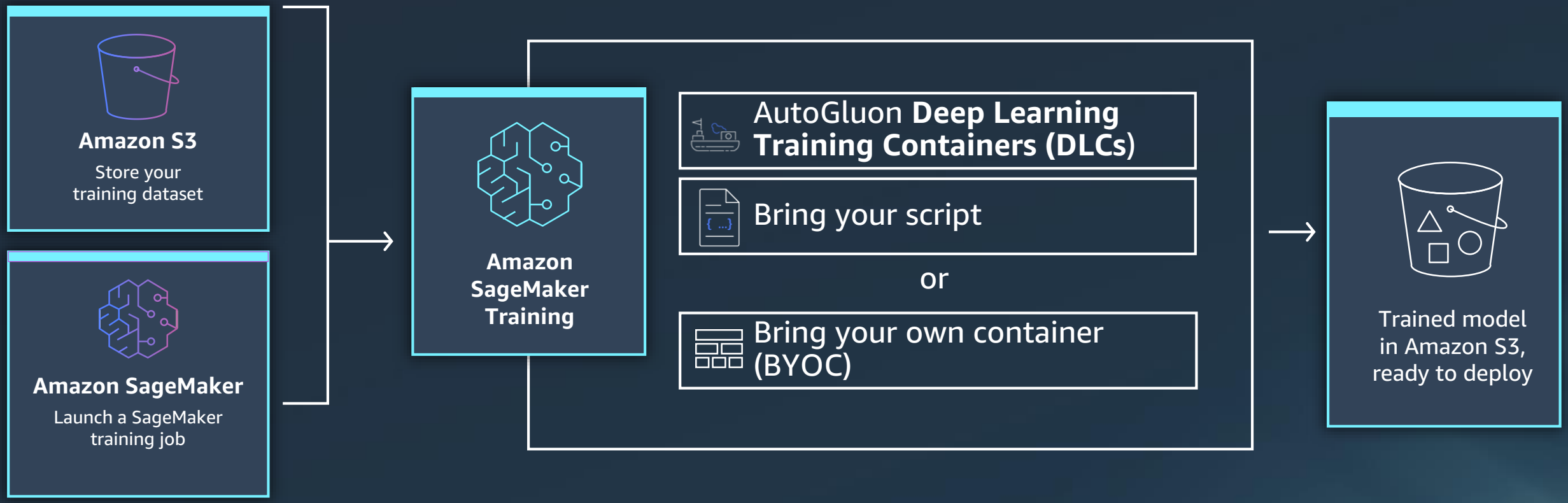
Login to Amazon SageMaker & explore dataset

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

Demo here - Video Playing here

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

Training with AutoGluon on Amazon SageMaker



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

Training using AutoMM with Amazon SageMaker

In this segment of demo we will run training using the following steps:

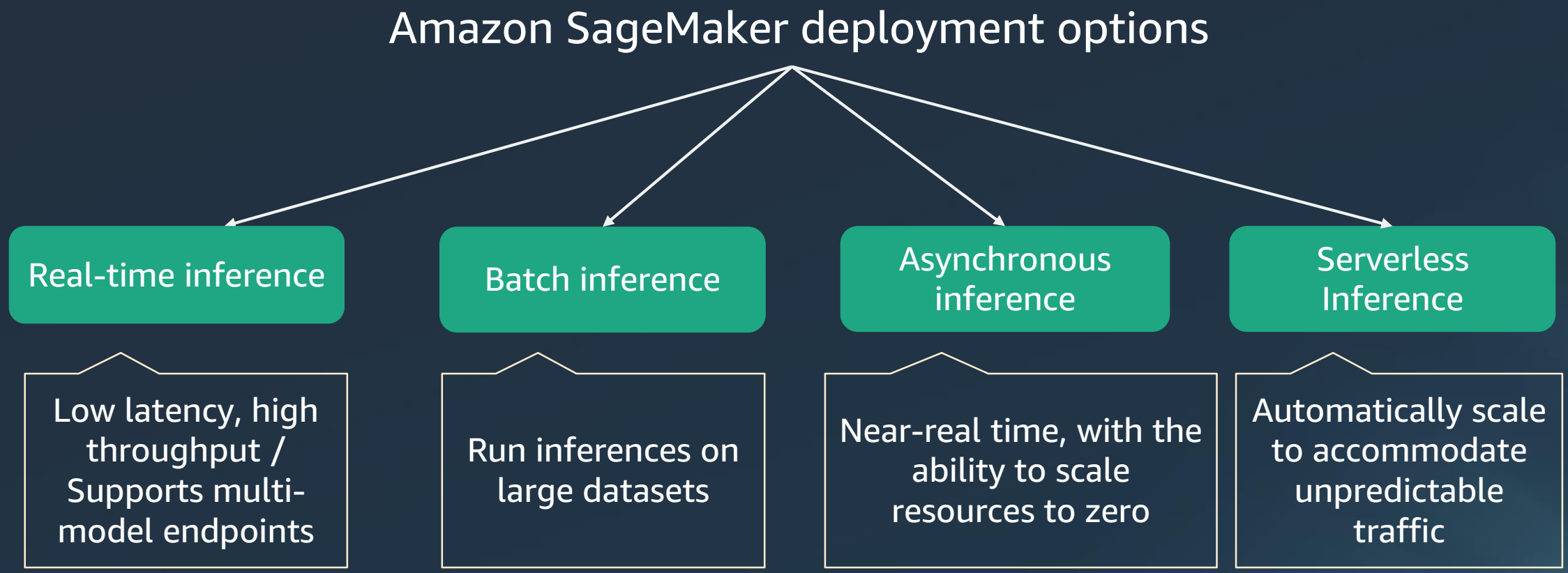
- Choose the right machine
- Load datasets
- Retrieve AWS managed AutoGluon container
- Create an estimator & call fit

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

Demo here - Training using AutoMM with Amazon Sagemaker Video Running

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

Host your model on Amazon SageMaker



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

Demo here - Amazon SageMaker Inferencing Video Running

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

Recap

- AWS Deep Learning Containers let you deploy custom ML environments quickly
- Fast track development of your ML platform & solutions with Amazon managed ML service SageMaker
- No docker container skills needed to bring your ML training/deployment code to Amazon SageMaker with script mode

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

Additional resources

- Amazon SageMaker onboarding video
- AutoGluon multimodal tutorial
- Tune in to other Amazon SageMaker sessions:
    - Get started with Amazon SageMaker in minutes
    - Train ML models quickly & cost-effectively with Amazon SageMaker
    - Building text-to-image model with Amazon SageMaker

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

Visit the Data & AI/ML resource hub

Dive deeper into these resources, get inspired and learn how you can use AI and machine learning to accelerate your business outcomes.

- 6 steps to machine learning success e-book
- 7 leading machine learning use cases e-book
- Machine learning at scale e-book
- Achieving transformative business results with machine learning e-book
- Tackling our world's hardest problems with machine learning e-book
- Accelerating machine learning innovation through security e-book
- … and more!

https://bityl.co/FqdC

Visit resource hub

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

AWS Training and Certification

Access the AI & ML learning plan courses built by AWS experts on AWS Skill Builder

- Get started with digital self-paced, on-demand training and ramp-up guides to help you grow your technical skills
- Learn how to apply machine learning, artificial intelligence, and deep learning to unlock new insights and value in your role
- Take the steps today, towards validating your expertise with an AWS Certified Machine Learning – Specialty Certification

https://bit.ly/3FnxDH7

Learn your way explore.skillbuilder.aws »

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

Thank you for attending AWS Innovate – Data & AI/ML Edition

We hope you found it interesting! A kind reminder to **complete the survey.**
Let us know what you thought of today's event and how we can improve the event experience for you in the future.

aws-apj-marketing@amazon.com

twitter.com/AWSCloud

facebook.com/AmazonWebServices

youtube.com/user/AmazonWebServices

slideshare.net/AmazonWebServices

twitch.tv/aws

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

Thank you!

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