機械学習のための統合開発環境(IDE) Amazon SageMaker Studioの概要とデモンストレーション

アマゾンウェブサービスジャパン株式会社
機械学習ソリューションアーキテクト
伊藤 芳幸

© 2020, Amazon Web Services, Inc. or its Affiliates.

自己紹介

伊藤 芳幸

- 機械学習ソリューションアーキテクト
- 前々職はSIer
- 前職は事業会社にてデータ活用推進
- 好きなサービス
    - Amazon SageMaker
    - Amazon Forecast
    - Amazon SageMaker Studio(?)

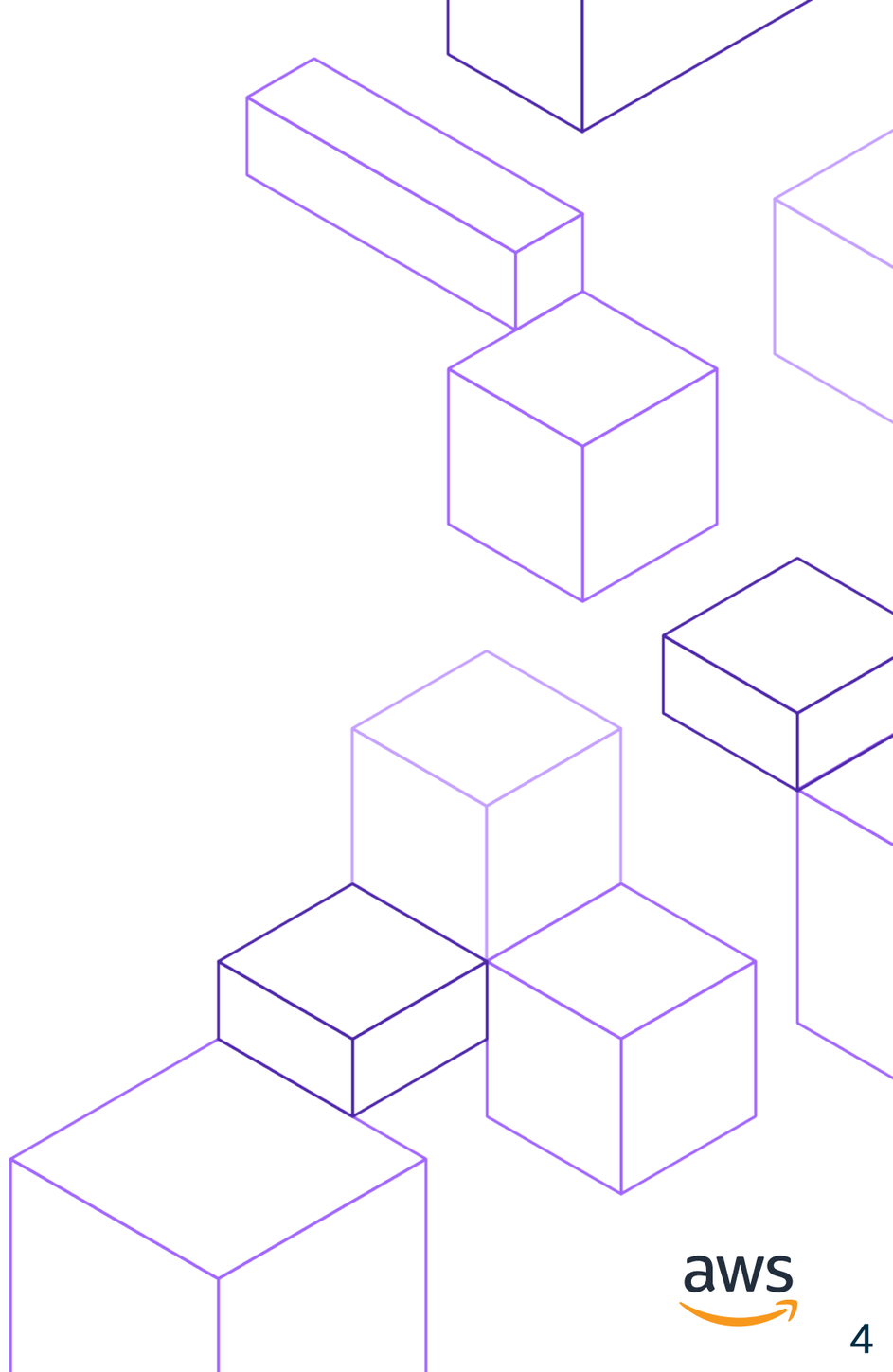
© 2020, Amazon Web Services, Inc. or its Affiliates.

アジェンダ

- Amazon SageMaker とは
- Amazon SageMaker Studio の機能紹介
    - Amazon SageMaker Notebooks
    - Amazon SageMaker Experiments
    - Amazon SageMaker Debugger
    - Amazon SageMaker Model Monitor
    - Amazon SageMaker Autopilot
- デモンストレーション

© 2020, Amazon Web Services, Inc. or its Affiliates.

Amazon SageMakerとは



© 2020, Amazon Web Services, Inc. or its Affiliates.

AWS の ML スタック

AI サービス: 機械学習の深い知識なしに利用可能

VISION	SPEECH		TEXT			SEARCH NEW	CHATBOTS	PERSONALIZATION	FORECASTING	FRAUD NEW	DEVELOPMENT NEW	CONTACT CENTERS NEW
Amazon Rekognitions *+Custom labels* NEW	Amazon Polly	Amazon Transcribe *+Medical* NEW	Amazon Comprehend *+Medical*	Amazon Translate	Amazon Textract	Amazon Kendra	Amazon Lex	Amazon Personalize	Amazon Forecast	Amazon Fraud Detector	Amazon CodeGuru	Contact Lens *For Amazon Connect*

ML サービス: 機械学習のプロセス全体を効率化するマネージドサービス

Amazon SageMaker

Ground Truth	NEW Augmented AI	SageMaker Studio IDE NEW							SageMaker Neo
		Built-in algorithms	SageMaker Notebooks NEW	SageMaker Experiments NEW	Model tuning	SageMaker Debugger NEW	SageMaker Autopilot NEW	Model hosting	SageMaker Model Monitor NEW	

ML フレームワークとインフラストラクチャ: 機械学習の環境を自在に構築して利用

TensorFlow, mxnet, Chainer, PYTORCH, GLUON, Keras, Horovod

Deep Learning AMIs & Containers | GPUs & CPUs | Elastic Inference | Inferentia | FPGA

© 2020, Amazon Web Services, Inc. or its Affiliates.

AWS の ML スタック

AI サービス: 機械学習の深い知識なしに利用可能

VISION | SPEECH | TEXT | SEARCH NEW | CHATBOTS | PERSONALIZATION | FORECASTING | FRAUD NEW | DEVELOPMENT NEW | CONTACT CENTERS NEW

Amazon Rekognitions *+Custom labels* | Amazon Polly | Amazon Transcribe *+Medical NEW* | Amazon Comprehend *+Medical* | Amazon Translate | Amazon Textract | Amazon Kendra | Amazon Lex | Amazon Personalize | Amazon Forecast | Amazon Fraud Detector | Amazon CodeGuru | Contact Lens *For Amazon Connect*

Amazon SageMaker

ML サービス: 機械学習のプロセス全体を効率化するマネージドサービス

Amazon SageMaker

Ground Truth | Augmented AI NEW

SageMaker Studio IDE NEW

Built-in algorithms | SageMaker Notebooks NEW | SageMaker Experiments NEW | Model tuning | SageMaker Debugger NEW | SageMaker Autopilot NEW | Model hosting | SageMaker Model Monitor NEW

SageMaker Neo

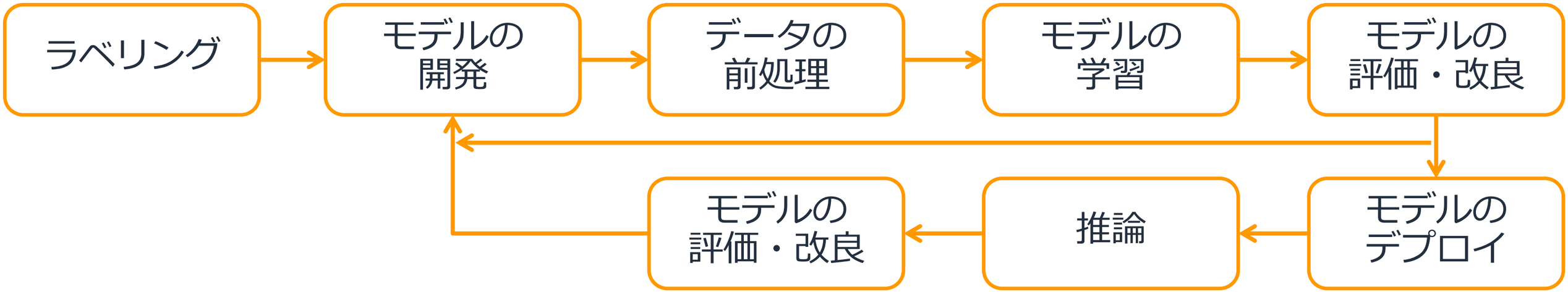
ML フレームワークとインフラストラクチャ: 機械学習の環境を自在に構築して利用

TensorFlow | mxnet | GLUON | Keras | Chainer | PYTORCH | Horovod

Deep Learning AMIs & Containers | GPUs & CPUs | Elastic Inference | Inferentia | FPGA

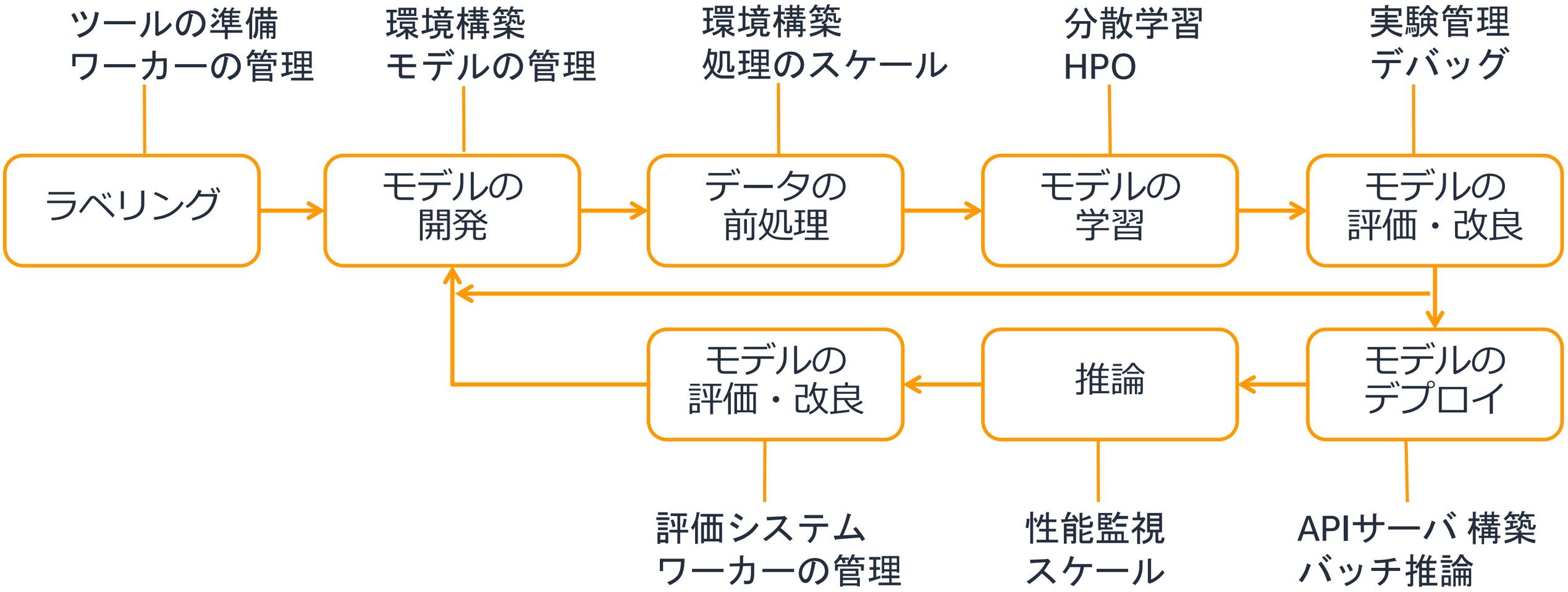
© 2020, Amazon Web Services, Inc. or its Affiliates.

機械学習の一般的なワークフロー



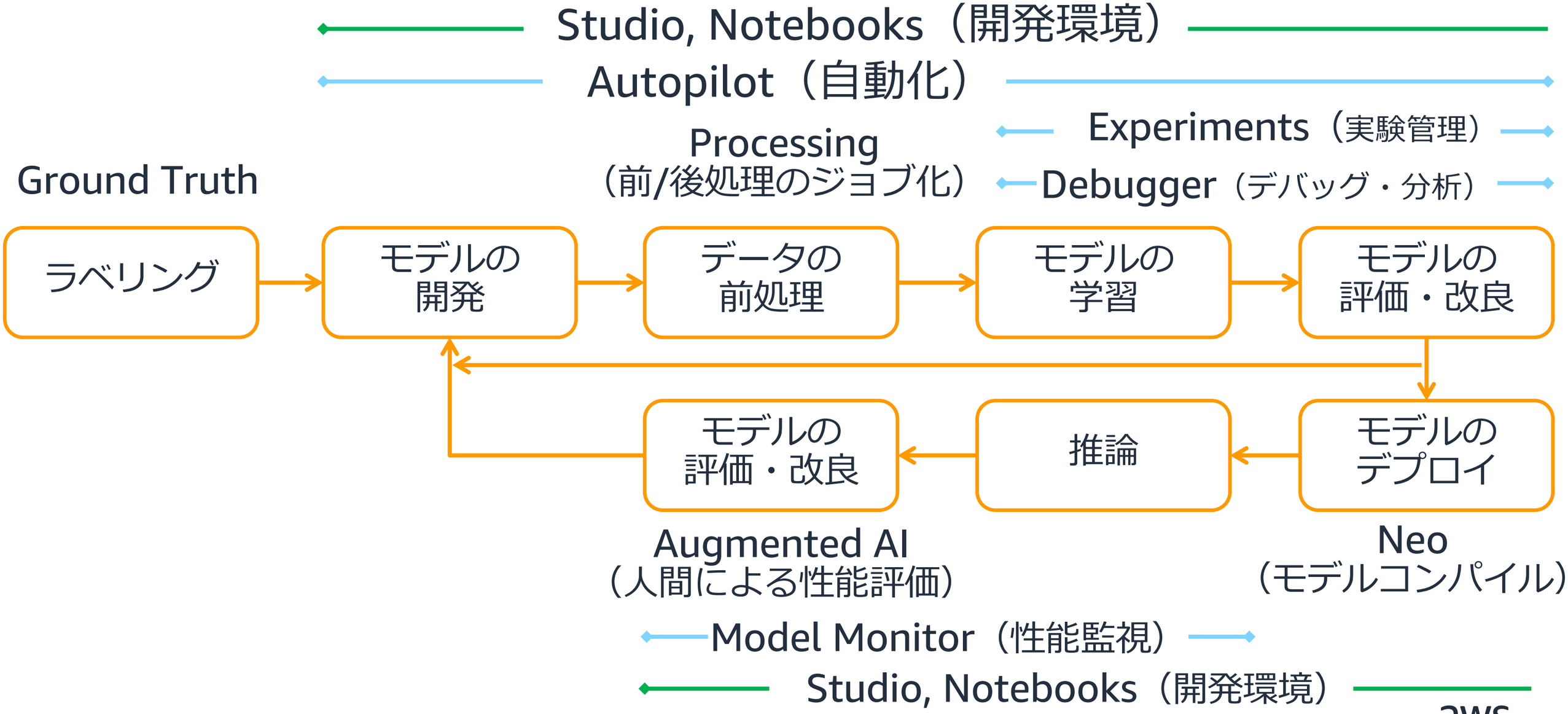
© 2020, Amazon Web Services, Inc. or its Affiliates.

各工程における課題の例



© 2020, Amazon Web Services, Inc. or its Affiliates.

Amazon SageMaker とは



© 2020, Amazon Web Services, Inc. or its Affiliates.

Amazon SageMaker の基本機能

開発

- 数分で開発環境を構築可能
- Jupyter notebookを使用可能
- コンテナイメージによって 同じ環境をいくつでもすぐに 構築可能

学習

- 状況に合わせて学習用インスタンスタイプの切り替えが容易
- 分散学習が容易
- 学習したモデルやソースコードは自動的に保存

推論

- APIエンドポイント（推論環境）の立ち上げが容易
- APIエンドポイントは 負荷に応じてオートスケール が可能
- リアルタイム推論とバッチ推論に対応

© 2020, Amazon Web Services, Inc. or its Affiliates.

開発・学習・推論は個別に利用可能

例 1: 部署全員に対して マネージドノートブック環境 を提供したい場合

管理不要のノートブックをホストする環境として SageMaker を利用

例 2: プロダクション環境がオンプレミス にすでにある場合

スケーラブルな学習環境としてのみ SageMaker を利用可能

例 3: GPU クラスタなどの 学習環境がオンプレミス にすでにある場合

オンプレミスで学習済のモデルを AWS 上のプロダクション環境にデプロイ

© 2020, Amazon Web Services, Inc. or its Affiliates.

Amazon SageMaker を使うメリット

コスト

- いま必要な分のインスタンス（プロビジョニング不要）
- 処理終了後、自動的にインスタンスが停止
- セキュリティ・コンプライアンスへのサポート

スピード

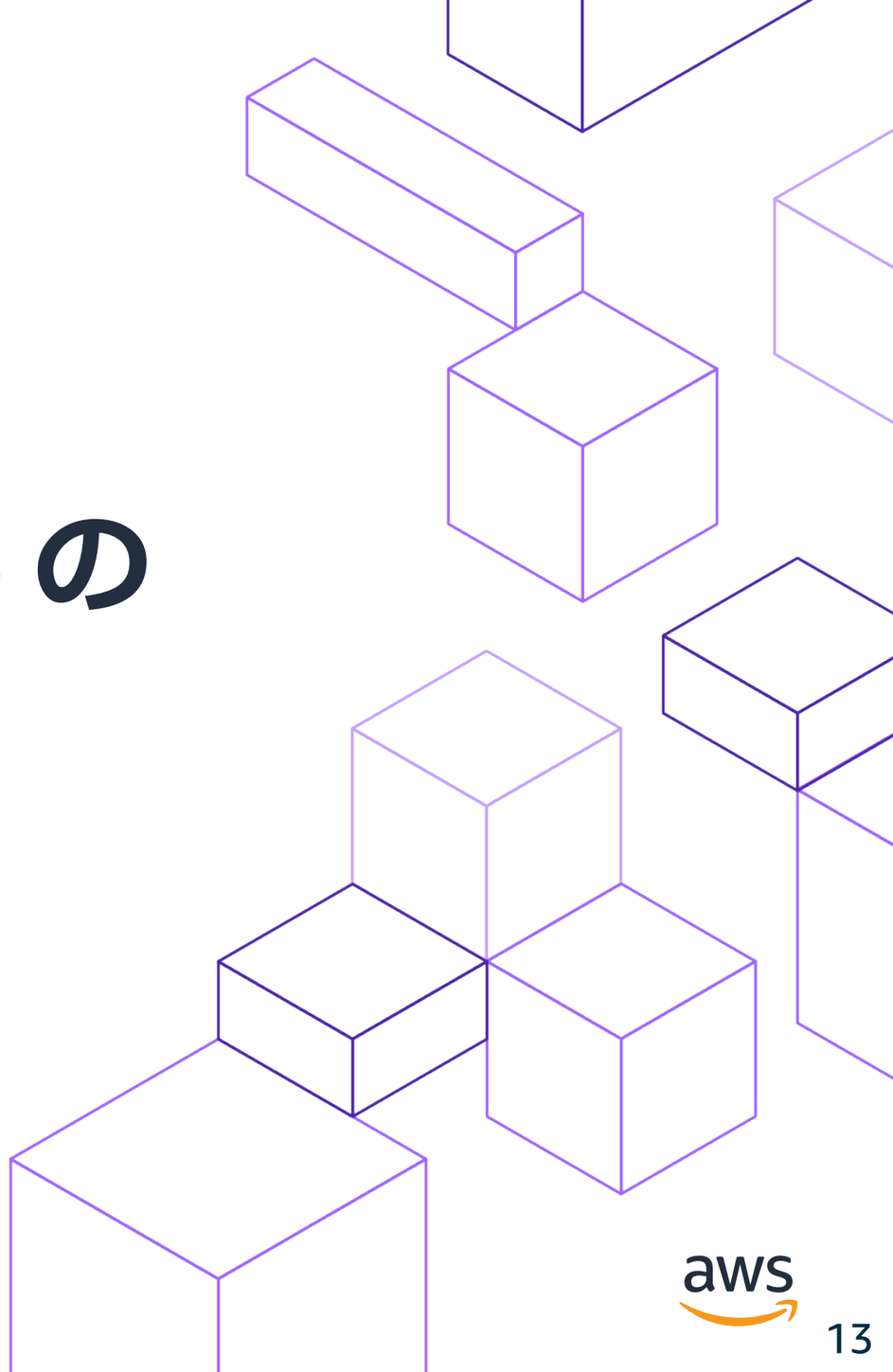
- 数分で環境構築
- API で推論環境構築
- 分散学習で学習時間の短縮

開発負荷

- 機械学習ワークフロー全体をカバー
- AWS の他のサービスとの連携が容易
- 独自のコンテナも使用可能

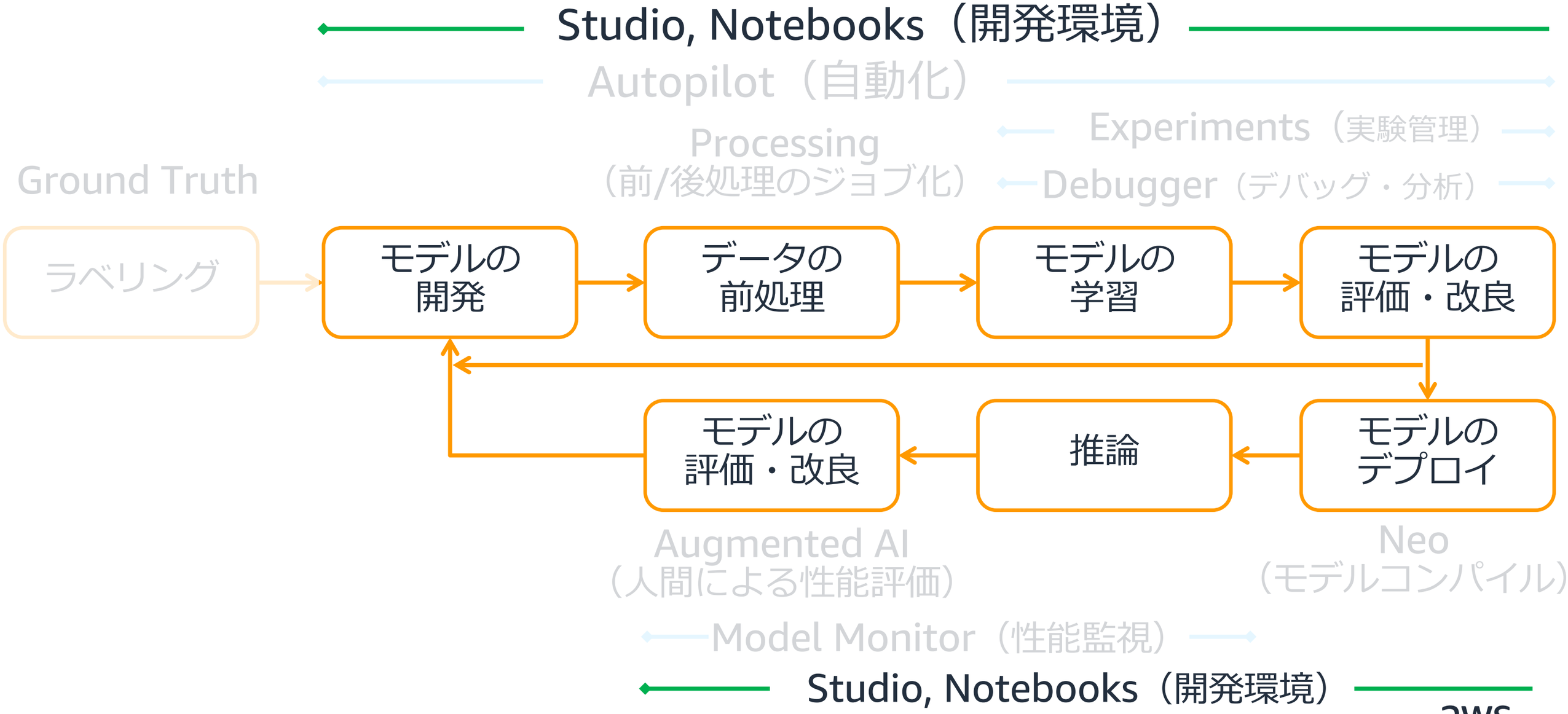
© 2020, Amazon Web Services, Inc. or its Affiliates.

Amazon SageMaker Studio の機能紹介



© 2020, Amazon Web Services, Inc. or its Affiliates.

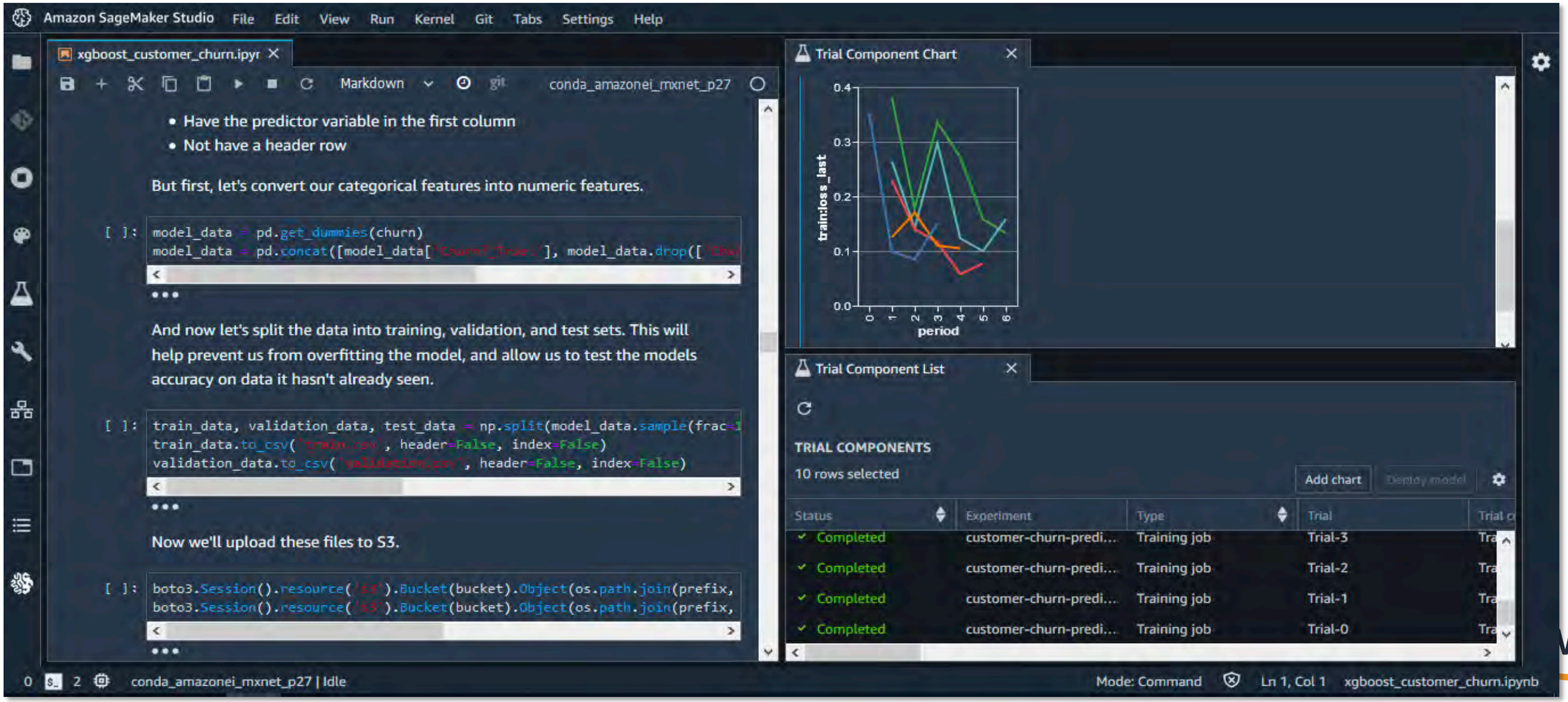
Amazon SageMaker とは



© 2020, Amazon Web Services, Inc. or its Affiliates.

Amazon SageMaker Studio

- SageMaker の各機能を呼び出すことができるweb上の統合開発環境
- AWS Single Sign-On (AWS SSO) , AWS IAM でアクセス可能
- 対応リージョン: オハイオ

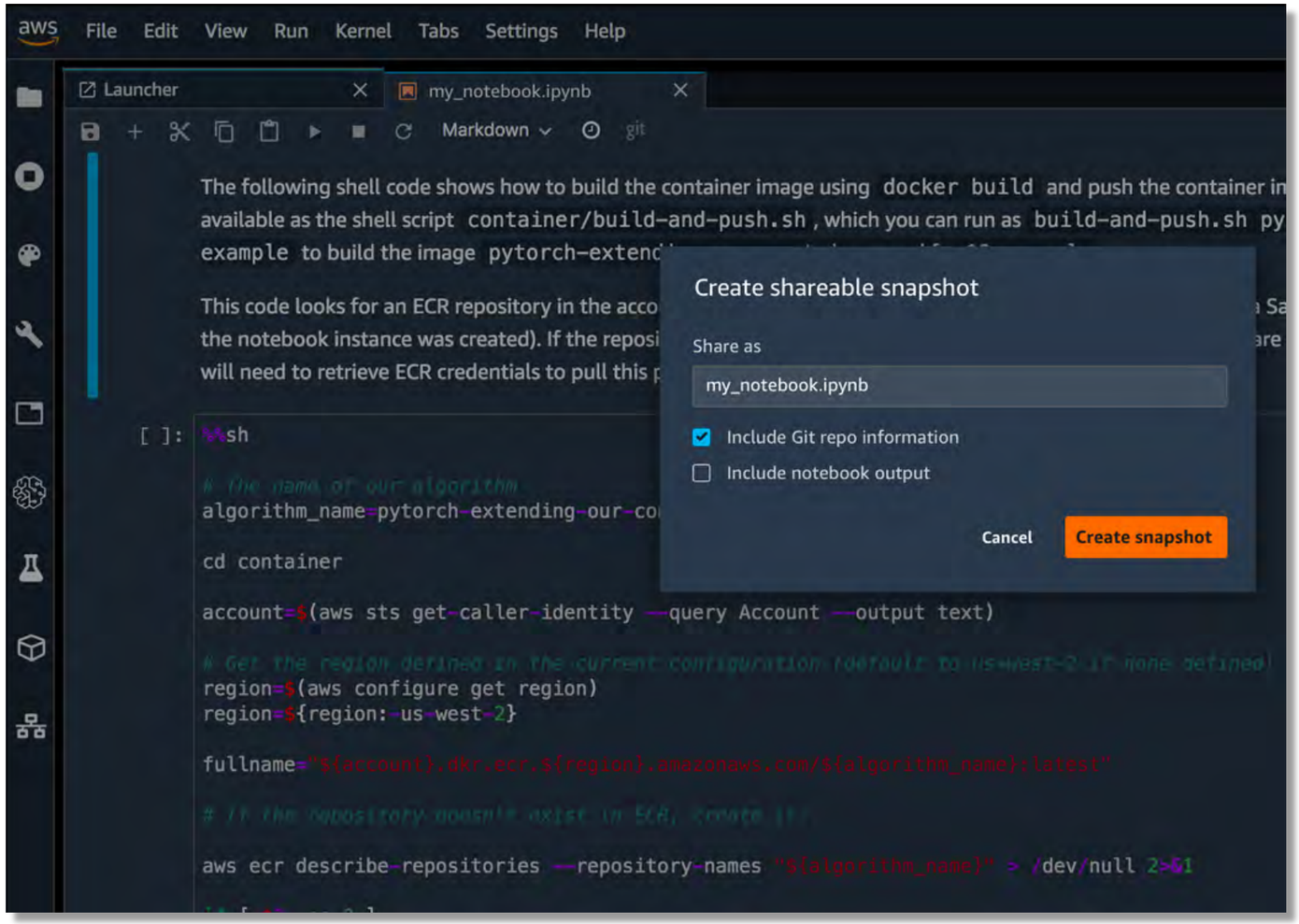


© 2020,

Amazon SageMaker Notebooks

Preview

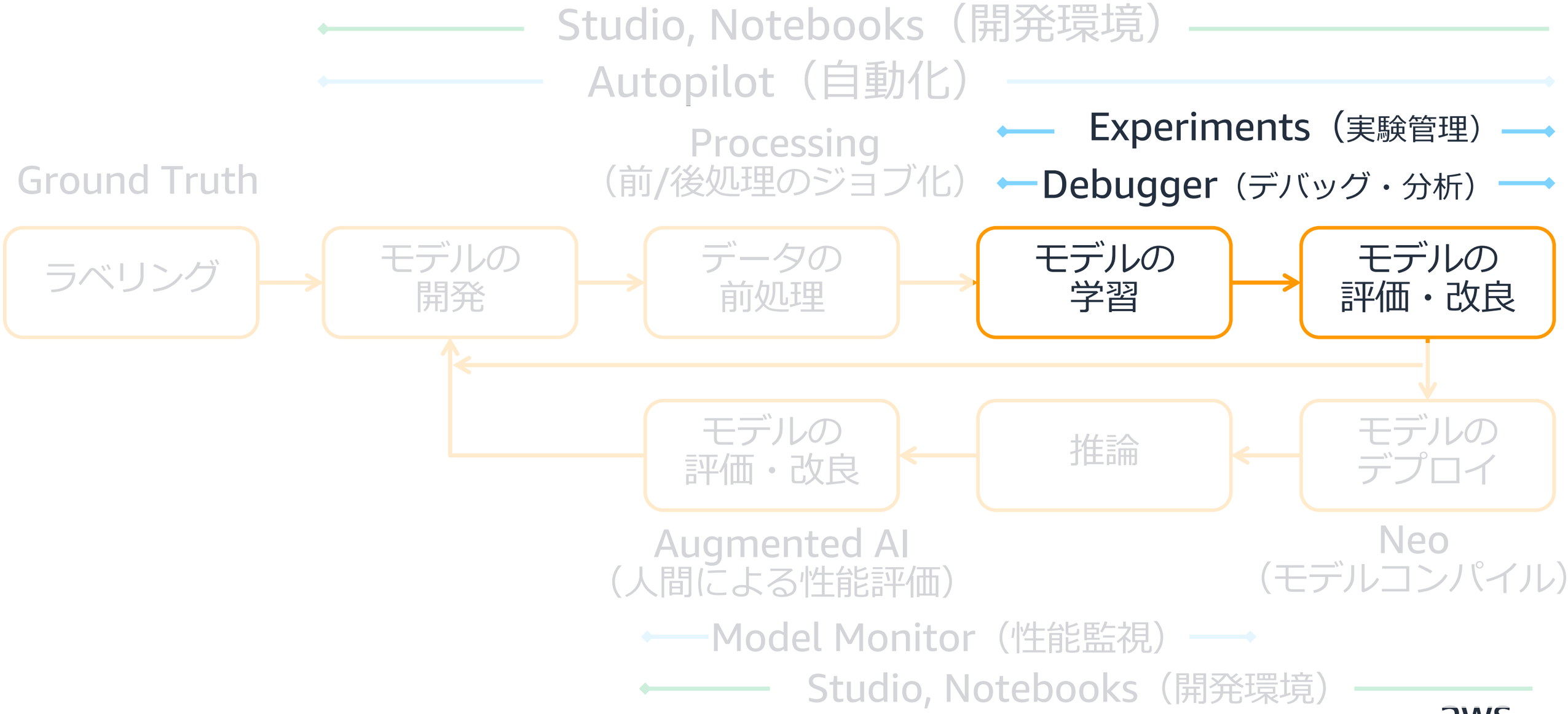
- AWS SSO でのアクセス が可能な ノートブック
- 待ち時間なくワンクリックで ノートブックを起動
- リソースの拡大・縮小は バックエンドで自動的に 行われる
- SSOのアプリケーションに ユーザーを登録し共同作業が可能
- オハイオリージョンで preview



https://aws.amazon.com/about-aws/whats-new/2019/12/introducing-the-new-amazon-sagemaker-notebook-experience-now-in-preview/

© 2020, Amazon Web Services, Inc. or its Affiliates.

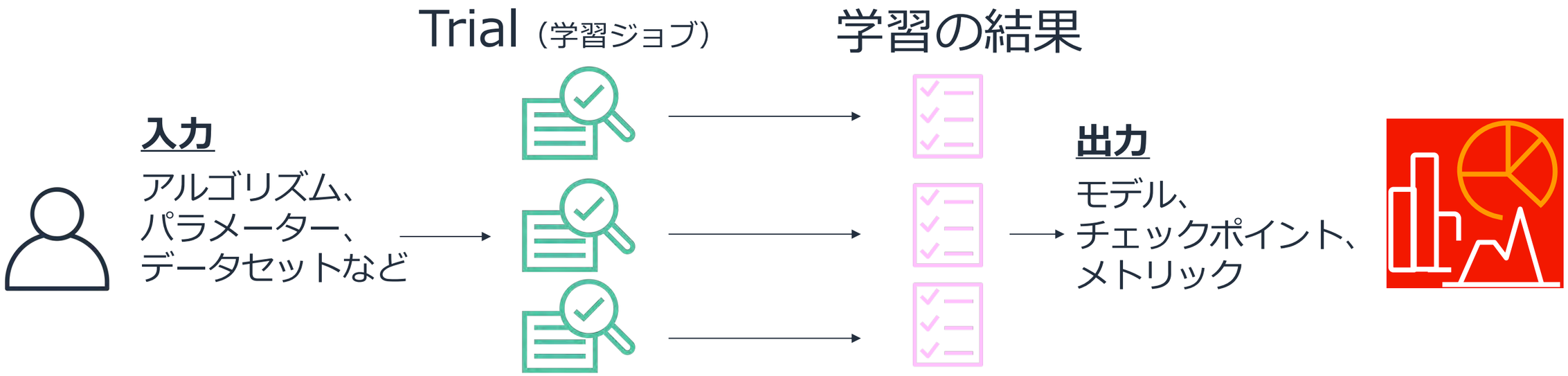
Amazon SageMaker とは



© 2020, Amazon Web Services, Inc. or its Affiliates.

Amazon SageMaker Experiments

- SageMaker Experiments SDK
- 学習を改善するための試行錯誤 (実験) の結果を管理する機能
- 複数の異なる学習を1つの実験と紐付け、メトリクスなどを一元管理
- 実験の結果を、pandas や matplotlib を使って確認できる



https://aws.amazon.com/jp/blogs/news/amazon-sagemaker-experiments-organize-track-and-compare-your-machine-learning-trainings/

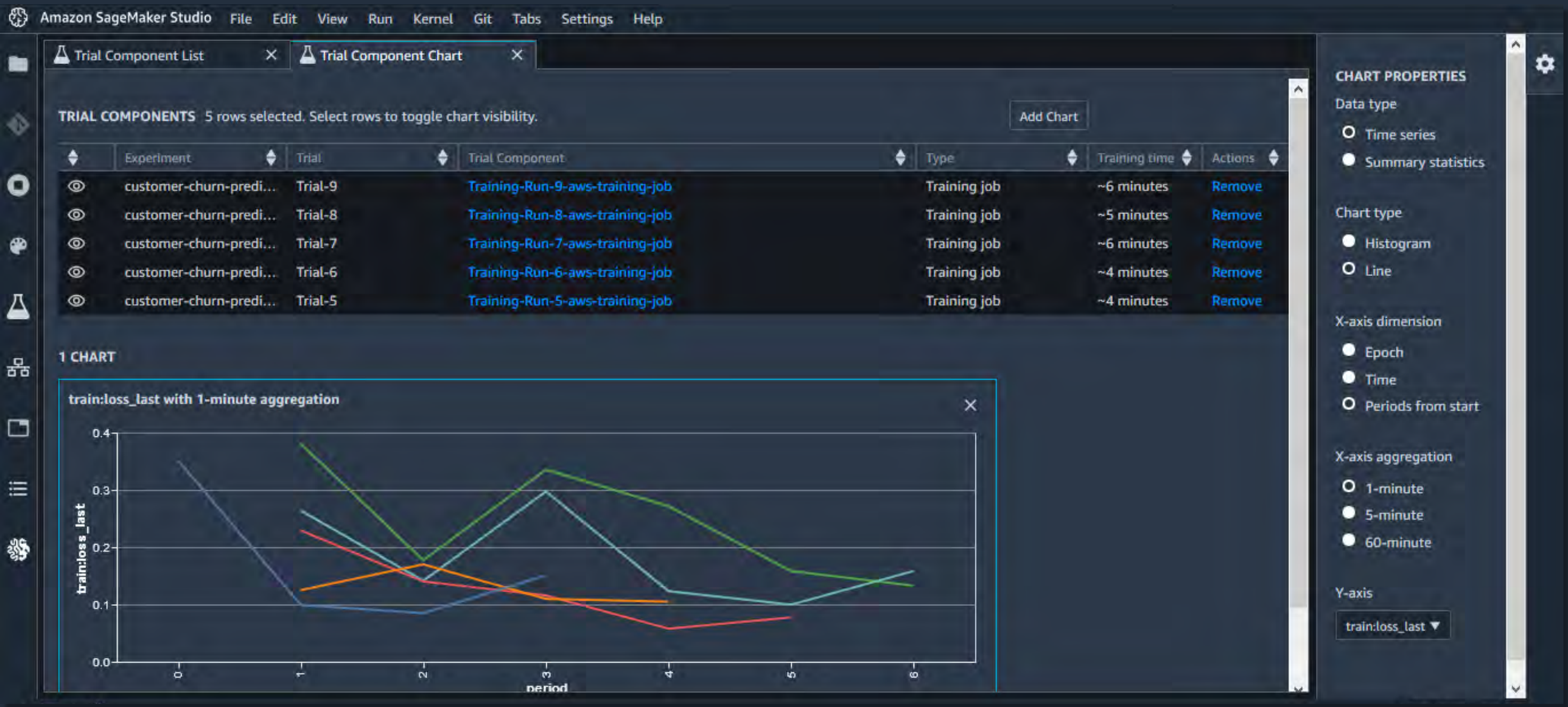
© 2020, Amazon Web Services, Inc. or its Affiliates.

サンプルノートブックからの抜粋

一連の Trial の結果を DataFrame にして一覧表示

	TrialComponentName	DisplayName	SourceArn	dropout	epochs	hidden_channels	optimizer	test:accuracy - Min	test:accuracy - Max	test:accuracy - Avg	test:acc - S
0	cnn-training-job-1576391108-aws-training-job	Training	[redacted]	0.2	2.0	32.0	"sgd"	95.0	97.0	96.0	1.4
1	cnn-training-job-1576390823-aws-training-job	Training		0.2	2.0	20.0	"sgd"	96.0	97.0	96.5	0.7
2	cnn-training-job-1576390538-aws-training-job	Training		0.2	2.0	10.0	"sgd"	95.0	95.0	95.0	0.0
3	cnn-training-job-1576390029-aws-training-job	Training		0.2	2.0	2.0	"sgd"	95.0	95.0	95.0	0.0
4	cnn-training-job-1576390254-aws-training-job	Training		0.2	2.0	5.0	"sgd"	94.0	94.0	94.0	0.0

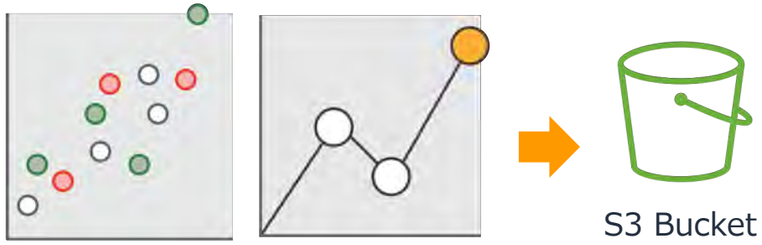
© 2020, Amazon Web Services, Inc. or its Affiliates.



© 2020, Amazon Web Services, Inc. or its Affiliates.

Amazon SageMaker Debugger

- Amazon SageMaker Debugger SDK
- 学習時の異常出力等の問題を検出
- 問題検出のルールがあらかじめ与えられ（追加も可能）、 S3 に出力される学習の途中計算結果（テンソル）からデバッグ
- TensorFlow, Keras, Apache MXNet, PyTorch, XGBoost に対応
- SageMakerが使える全てのリージョンで提供




```
Python
description = client.describe_training_job(TrainingJobN
print(description['DebugRuleEvaluationStatuses'][0]['Ru
print(description['DebugRuleEvaluationStatuses'][0]['Ru

ExplodingTensor
IssuesFound
```

```
Python
tensor = 'gradients/multiply/MatMul_1_gra
for s in list(trial.tensor(tensor).steps(
print("Value: ", trial.tensor(tensor).ste

Value: [[1.1508383e+23] [1.0809098e+23]]
Value: [[1.0278440e+23] [1.1347468e+23]]
Value: [[nan] [nan]]
Value: [[nan] [nan]]
Value: [[nan] [nan]]
Value: [[nan] [nan]]
```

https://aws.amazon.com/jp/blogs/news/amazon-sagemaker-debugger-debug-your-machine-learning-models/

© 2020, Amazon Web Services, Inc. or its Affiliates.

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

SMDEbugger-CloudWatch-Lo

Markdown git conda_tensorflow_p36

Using SageMaker Rules

In this example we'll demonstrate how to use SageMaker rules to be evaluated against your training. You can find the list of SageMaker rules and the configurations best suited for using them here.

We specify a few rules that check for overfitting, descrease in loss across epochs and for saturated activations.

[8]:
```
estimator = TensorFlow(
    role=sagemaker.get_execution_role(),
    base_job_name='mnist-tensorflow-example',
    train_instance_count=1,
    train_instance_type='ml.p3.2xlarge',
    image_name=cpu_training_image,
    entry_point=entrypoint_script,
    framework_version='1.15',
    py_version='py3',
    train_max_run=3600,
    script_mode=True,
    sagemaker_session=sess,
    ## New parameter
    rules = [ Rule.sagemaker(rule_configs.vanishing_gradient()),
              Rule.custom(name='Overfitting', # used to identify the rule
                          image_uri='[redacted].dkr.ecr.us-west-2.amazonaws.com
                          instance_type='ml.c4.xlarge', # instance type to run th
                          source='my_custom_rule.py', # path to the rule source f
                          rule_to_invoke='CustomGradientRule', # name of the clas
                          volume_size_in_gb=400, # EBS volume size required to be
                          collections_to_save=[CollectionConfig(name='losses')],
                          rule_parameters={
                              "threshold": "20.0" # this will be used to intializ
                          }) ],
    hyperparameters = {'num_epochs' : 100 }
)
```

Note that Sagemaker-Debugger is only supported for py_version='py3' currently.

Let's start the training by calling `fit()` on the MXNet estimator

[9]:
```
# After calling fit, SageMaker will spin off 1 training job and 1 rule job for y
# The rule evaluation status(es) will be visible in the training logs
# at regular intervals

estimator.fit(wait=False)
```

Result

Describe Trial Component

Experiment: Unassigned

Trial: Unassigned

Trial stages

Charts Metrics Parameters Artifacts AWS Settings Debugger

mnist-tensorflow-example-2019-12-02-09-52-13-126-aws-training-job
Created
15 minutes ago

Status	Last modified	Rule name	Job ARN
Issues Found	4 minutes ago	VanishingGradient	arn:aws:sagemaker:us-west-2:3
Issues Found	4 minutes ago	Overfitting	arn:aws:sagemaker:us-west-2:3

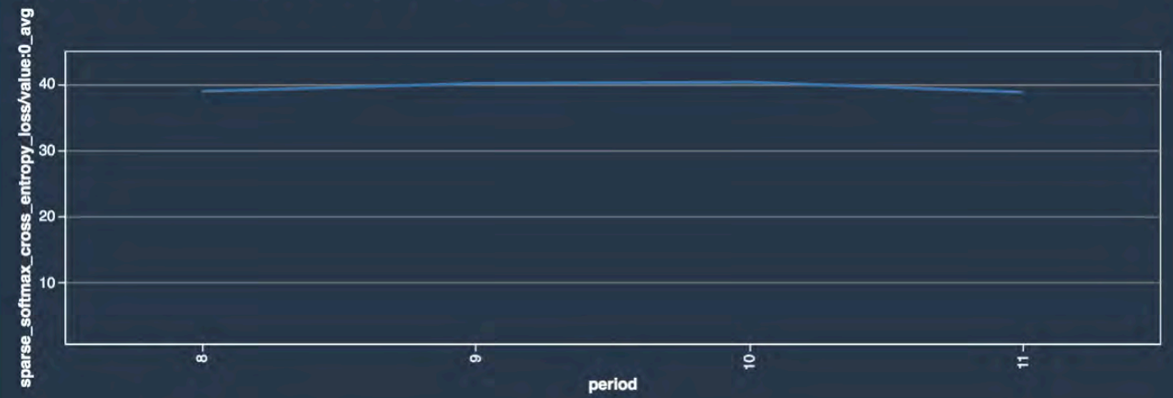
Trial Component Chart

TRIAL COMPONENTS 1 rows selected. Select rows to toggle chart visibility. Add Chart

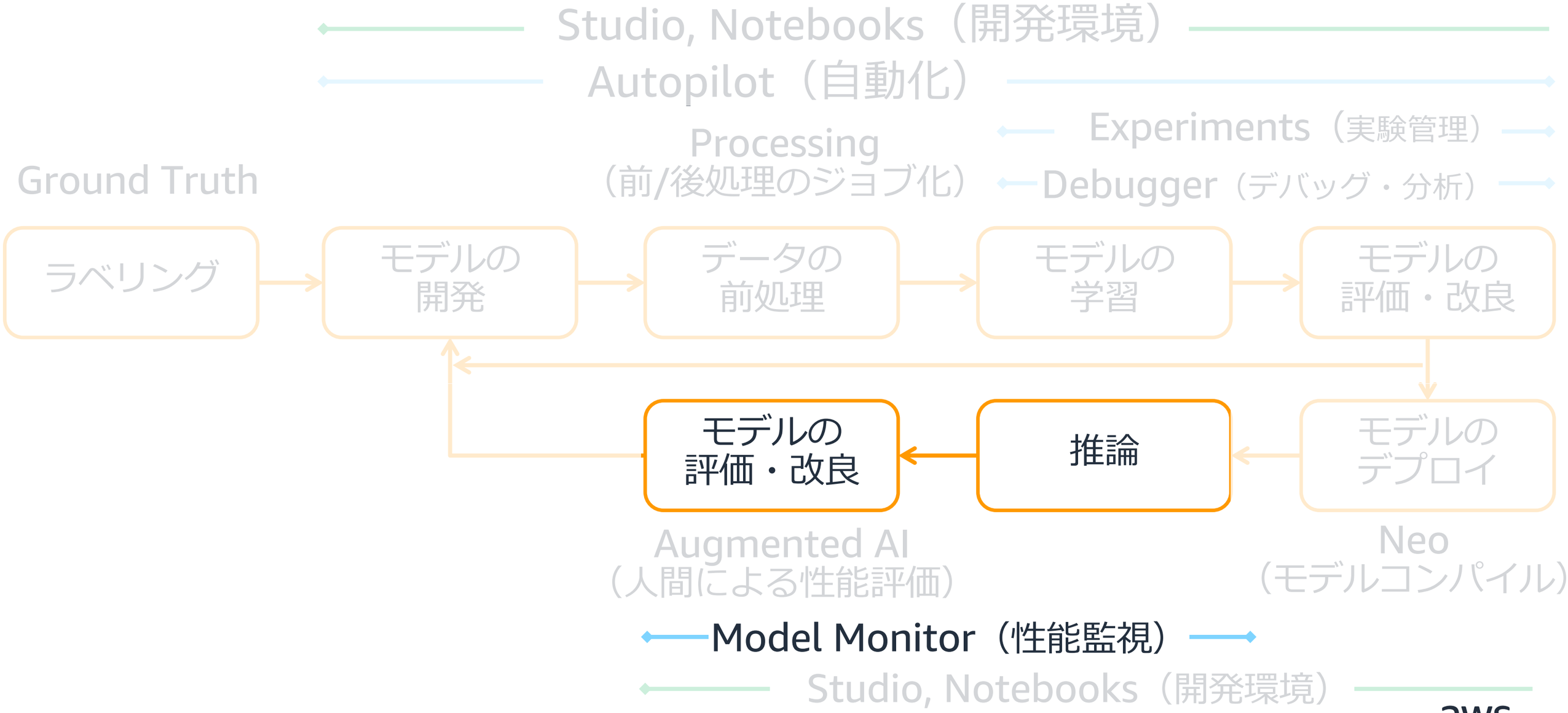
	Experiment	Trial	Trial Component	Type	Train
👁	N/A	N/A	mnist-tensorflow-example-2019-12-02-09-52-13-126-aws-trainin...	Training job	~10

2 CHARTS

sparse_softmax_cross_entropy_loss/value:0_avg with 1-minute aggregation



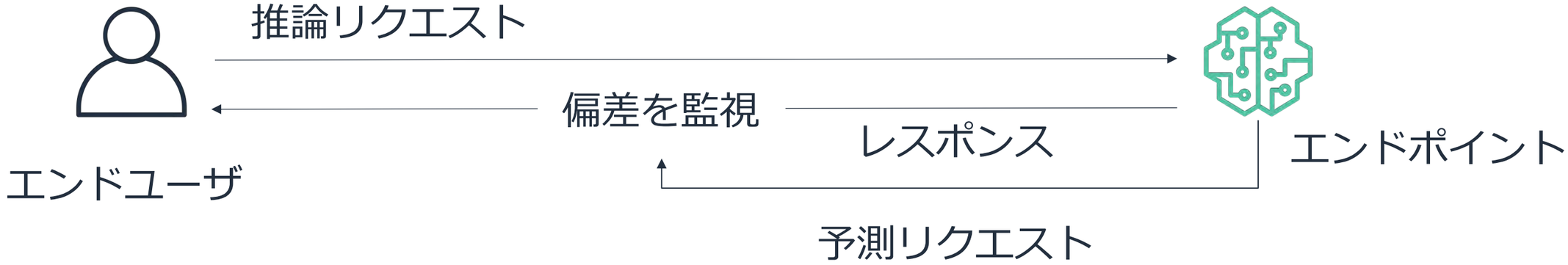
Amazon SageMaker とは



© 2020, Amazon Web Services, Inc. or its Affiliates.

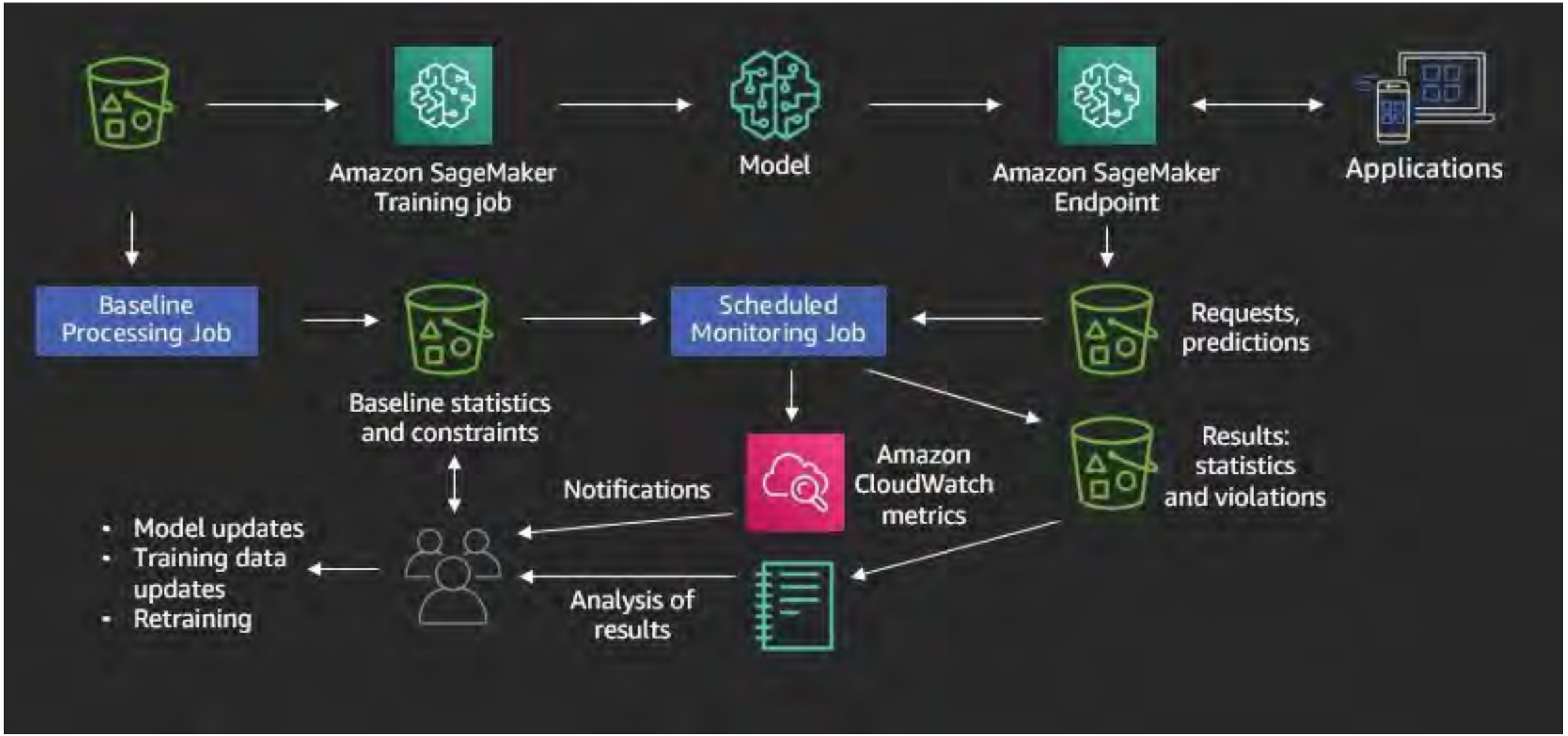
Amazon SageMaker Model Monitor

- MLモデルの品質を維持する機能
- SageMaker の推論エンドポイントの入出力を継続的に監視し、学習時に用いたデータをベースにした予測リクエストと実際のリクエストに乖離がないかを評価する
- 乖離がある場合は、モデルが正しく推論できていない可能性があり、最新のデータに合わせて再学習を行う必要がある

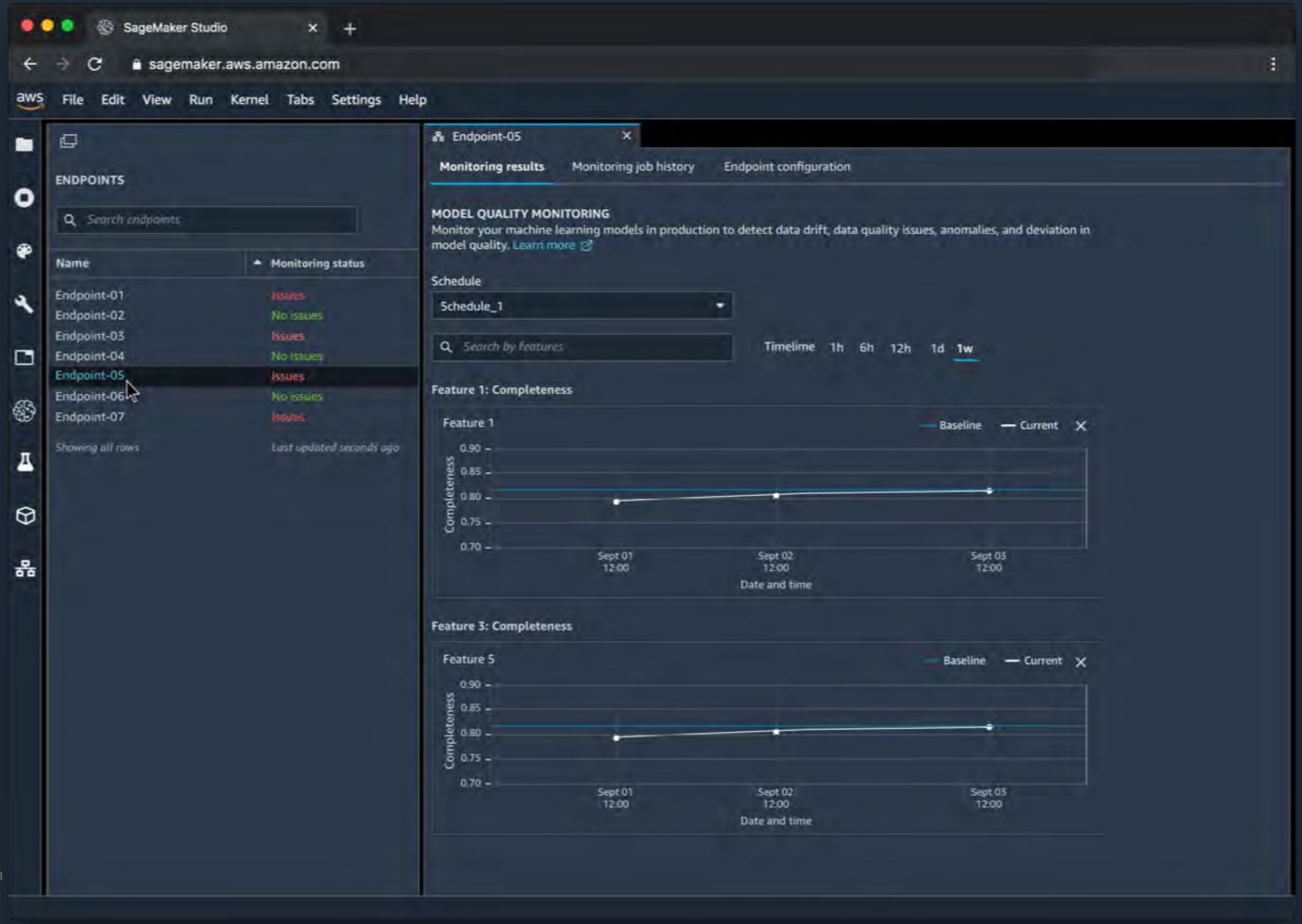


https://aws.amazon.com/jp/blogs/news/amazon-sagemaker-model-monitor-fully-managed-automatic-monitoring-for-your-machine-learning-models/

Amazon SageMaker Model Monitor

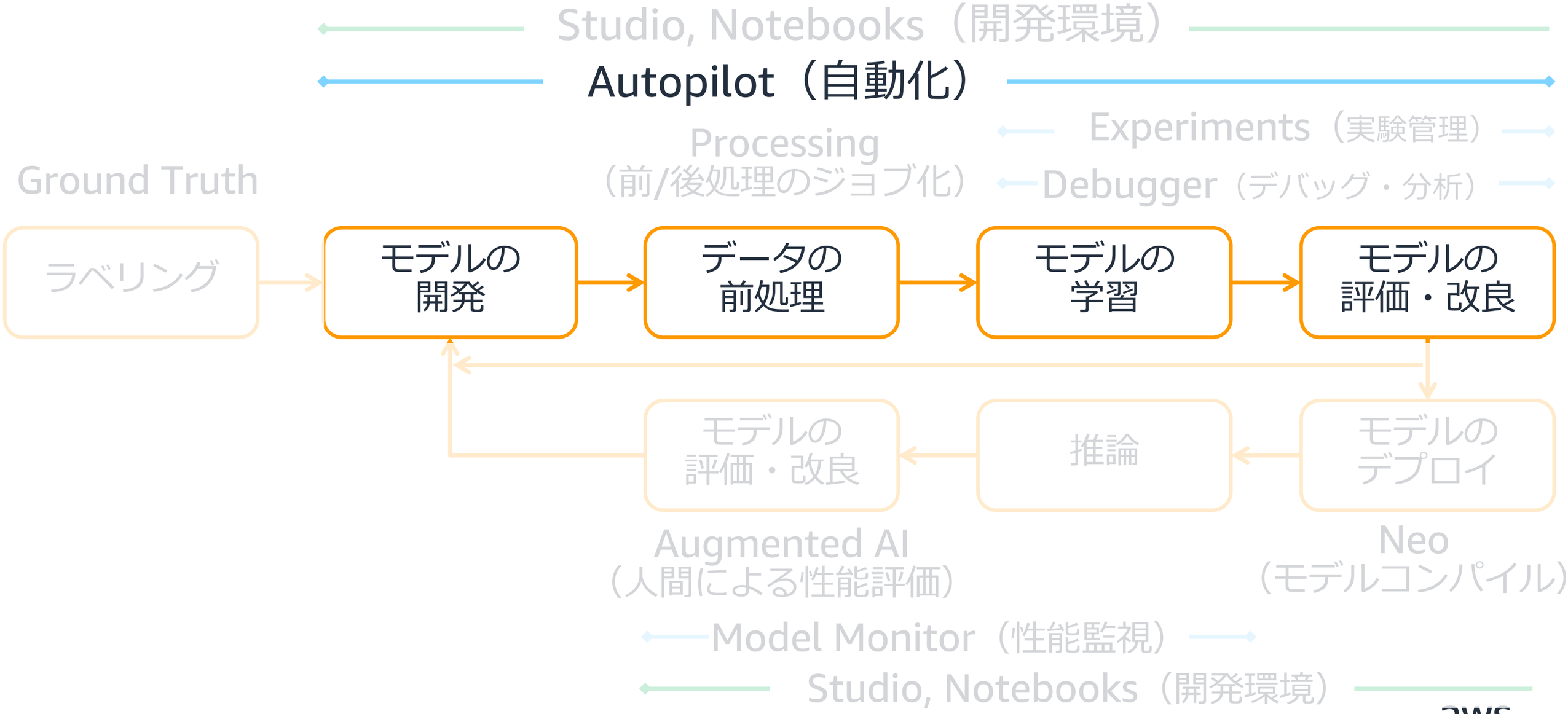


© 2020, Amazon Web Services, Inc. or its Affiliates.


SageMaker Studio
sagemaker.aws.amazon.com
aws
File
Edit
View
Run
Kernel
Tabs
Settings
Help
ENDPOINTS
Search endpoints
Name
Monitoring status
Endpoint-01
Issues
Endpoint-02
No issues
Endpoint-03
Issues
Endpoint-04
No issues
Endpoint-05
Issues
Endpoint-06
No issues
Endpoint-07
Issues
Showing all rows
Last updated seconds ago
Endpoint-05
Monitoring results
Monitoring job history
Endpoint configuration
MODEL QUALITY MONITORING
Monitor your machine learning models in production to detect data drift, data quality issues, anomalies, and deviation in model quality. Learn more
Schedule
Schedule_1
Search by features
Timelime
1h
6h
12h
1d
1w
Feature 1: Completeness
Feature 1
Baseline
Current
Completeness
0.90
0.85
0.80
0.75
0.70
Sept 01 12:00
Sept 02 12:00
Sept 03 12:00
Date and time
Feature 3: Completeness
Feature 5
Baseline
Current
Completeness
0.90
0.85
0.80
0.75
0.70
Sept 01 12:00
Sept 02 12:00
Sept 03 12:00
Date and time

© 2020, Amazon

Amazon SageMaker とは

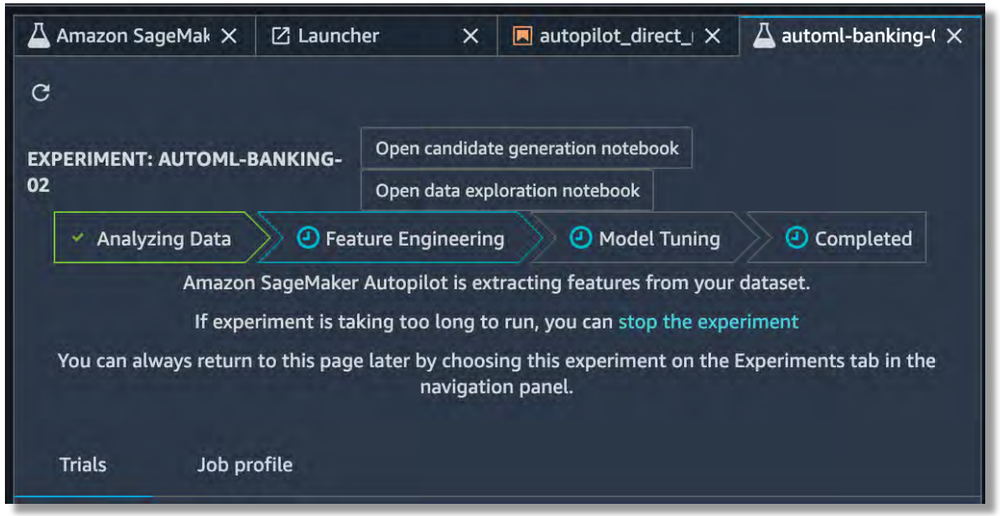


© 2020, Amazon Web Services, Inc. or its Affiliates.

Amazon SageMaker Autopilot

- テーブルデータに対する予測と分類の機械学習のAutoML機能
- 不正検知、ユーザ分類（離反予測）、レコメンデーションなど多数のユースケースに適用できる
- データソースと予測対象を指定
- データの前処理、アルゴリズムの選択、インスタンスタイプ選択なども自動化
- US4リージョン、東京など16リージョンで利用可能

SageMaker 上でAutoMLを簡単に実行



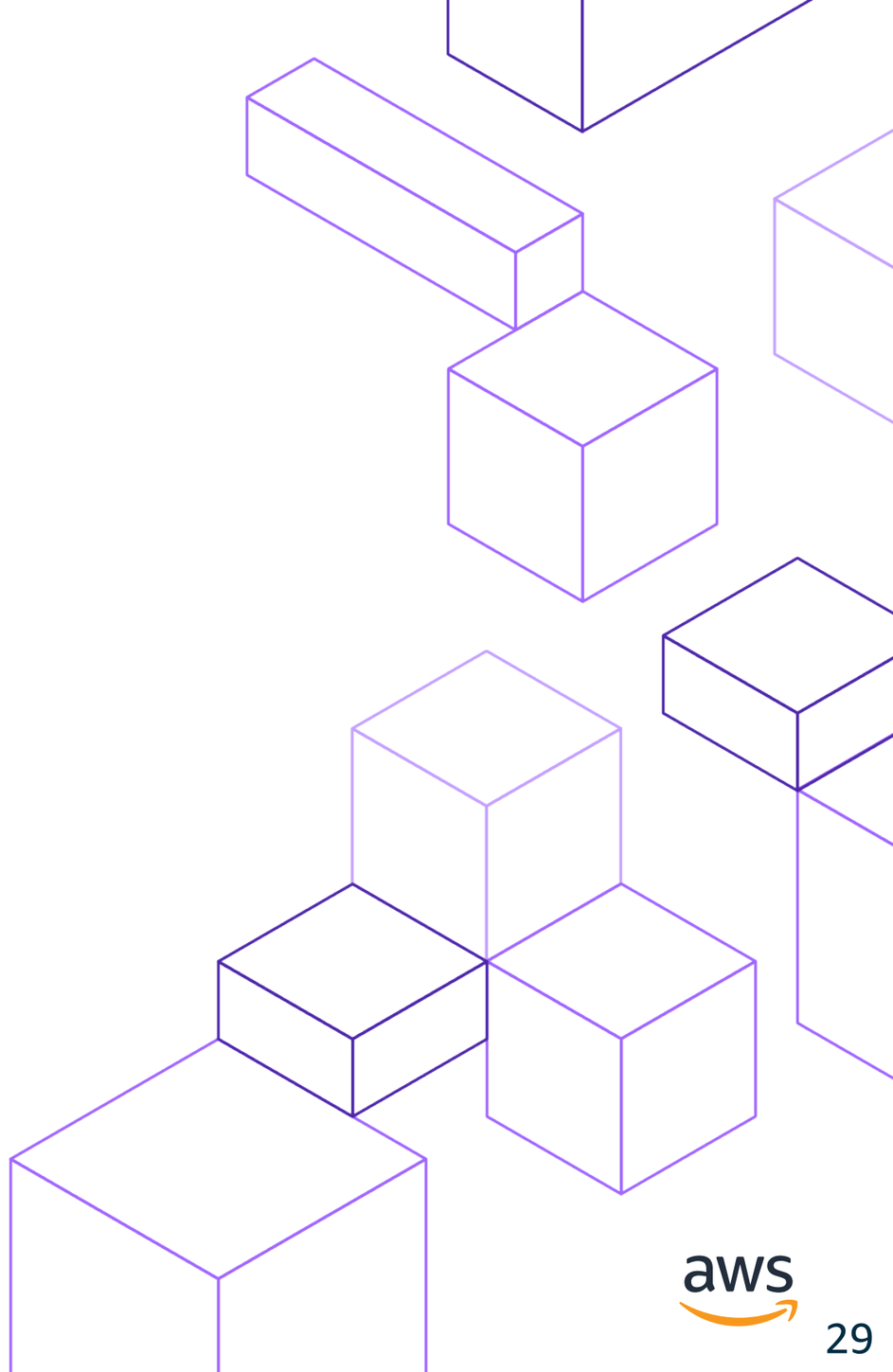
生データの検査や特徴量抽出

	age	job	marital	education	default	housing	loan
0	56	housemaid	married	basic.4y	no	no	no
1	57	services	married	high.school	unknown	no	no
2	37	services	married	high.school	no	yes	no
3	40	admin.	married	basic.6y	no	no	no
4	56	services	married	high.school	no	no	yes
5	45	services	married	basic.9y	unknown	no	no

https://aws.amazon.com/blogs/aws/amazon-sagemaker-autopilot-fully-managed-automatic-machine-learning/

© 2020, Amazon Web Services, Inc. or its Affiliates.

デモンストレーション



© 2020, Amazon Web Services, Inc. or its Affiliates.

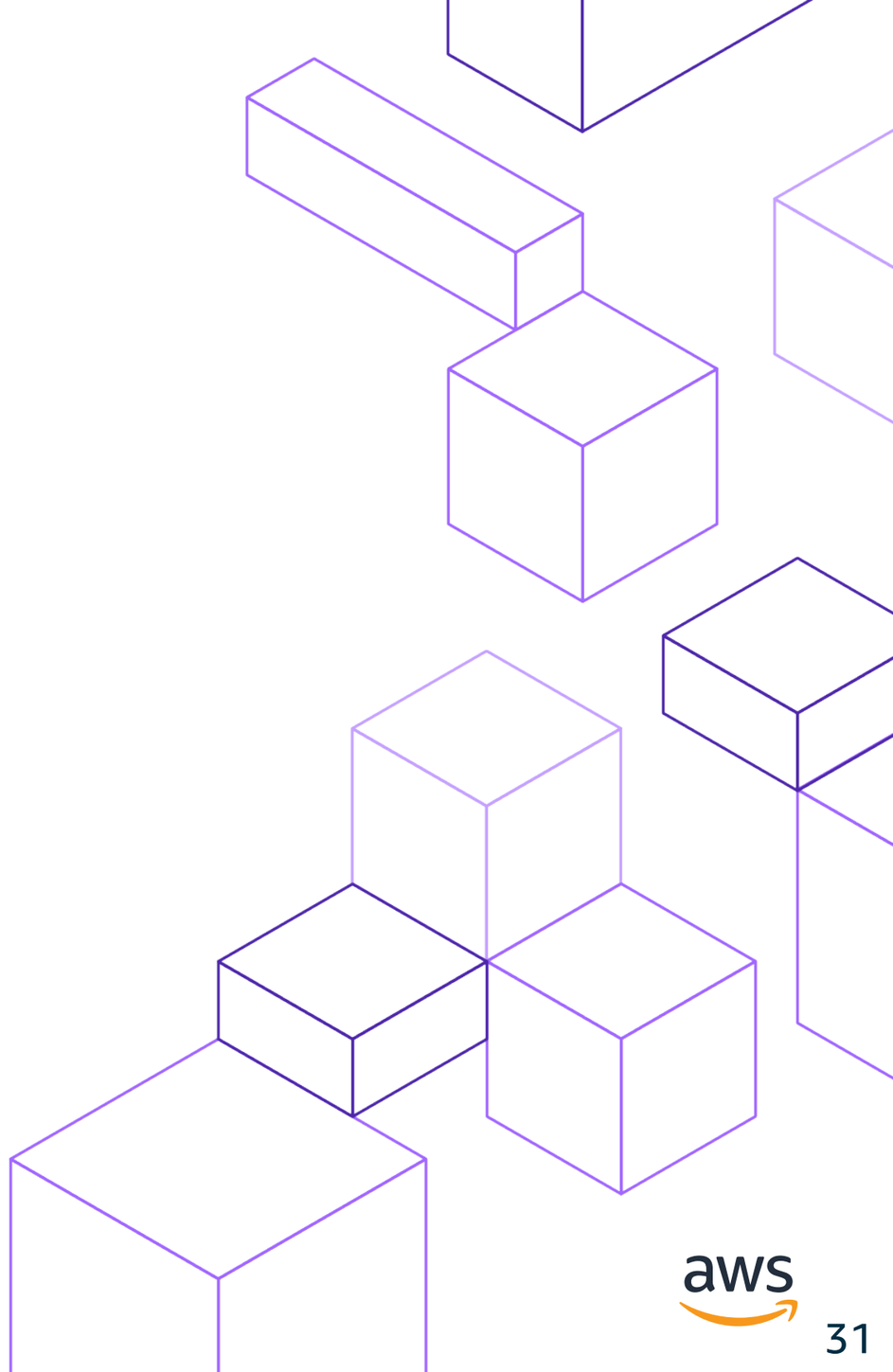
デモンストレーションの流れ

- Studioのドメイン構築(SSO方式)からスタートし、Studio上でAutopilotを試します。
- 本デモンストレーションは以下のYoutubeの流れを実行したものです。
    - Amazon SageMaker Studio - AutoML with Amazon SageMaker AutoPilot (part 1) - YouTube
        - https://www.youtube.com/watch?v=qMEtqJPhqpA
    - Amazon SageMaker Studio - AutoML with Amazon SageMaker AutoPilot (part 2) - YouTube
        - https://www.youtube.com/watch?v=WsfRAeGzgm8
    - Amazon SageMaker Studio - AutoML with Amazon SageMaker AutoPilot (part 3) - YouTube
        - https://www.youtube.com/watch?v=KZSTsWrDGXs
    - Amazon SageMaker Studio - AutoML with Amazon SageMaker AutoPilot (part 4) - YouTube
        - https://www.youtube.com/watch?v=vRHyX3kDstI
- 後片付け：ドメイン削除手順（デモ内でも削除いたします）
    - https://docs.aws.amazon.com/ja_jp/sagemaker/latest/dg/gs-studio-delete-domain.html

※途中待ち時間が発生するため、時間を飛ばしています。ご了承ください。

© 2020, Amazon Web Services, Inc. or its Affiliates.

まとめ



© 2020, Amazon Web Services, Inc. or its Affiliates.

SageMakerのメリット

- 機械学習システムの各工程の課題に対応。中長期的な視点でシステムの効率的な運用に貢献。

- 統合開発環境である Amazon SageMaker Studio により、ブラウザからお手軽に機械学習環境にアクセス可能。

- 各機能はこれからも継続的に改善されていく。

© 2020, Amazon Web Services, Inc. or its Affiliates.

Thank you!

© 2020, Amazon Web Services, Inc. or its Affiliates.