

Japan Wrap-up Session: AI/ML

アマゾン ウェブ サービス ジャパン株式会社
ソリューションアーキテクト
鮫島 正樹

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 Rekognition	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	Augmented AI ^{NEW}	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	 GLUON	 Keras	Deep Learning AMIs & Containers		GPUs & CPUs	Elastic Inference	Inferentia	FPGA
 Chainer	 PYTORCH								

AWS の ML スタック

New

AI サービス: 機械学習の深い知識なしに利用可能

VISION  Amazon Rekognition	SPEECH  Amazon Polly	 Amazon Transcribe +Medical NEW	TEXT  Amazon Comprehend +Medical	 Amazon Translate	 Amazon Textract	SEARCH NEW  Amazon Kendra	CHATBOTS  Amazon Lex	PERSONALIZATION  Amazon Personalize	FORECASTING  Amazon Forecast	FRAUD NEW  Amazon Fraud Detector	DEVELOPMENT NEW  Amazon CodeGuru	CONTACT CENTERS NEW  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

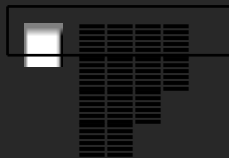
 Amazon SageMaker	Ground Truth	SageMaker Studio IDE NEW								SageMaker Neo
		Augmented AI NEW	Built-in algorithms	SageMaker Notebooks NEW	SageMaker Experiments NEW	Model tuning	SageMaker Debugger NEW	SageMaker Autopilot NEW	Model hosting	

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	Deep Learning AMIs & Containers	GPUs & CPUs	Elastic Inference	Inferentia	FPGA
 Chainer	 PYTORCH	 HOROVOD						

AI/ML 関連の発表概要

- 新しい AI サービス・機能を多数発表: 文書検索、不正検知、コードレビュー、AIコールセンターなど
- ML サービスでは、統合開発環境 **Amazon SageMaker Studio** によって、デバッグ、モデル監視、実験管理など、機械学習に必要な機能を簡単に実行
- ML フレームワークとインフラストラクチャでは、ML 推論のための推論用チップ Inferentia を搭載した **inf1** インスタンスを発表
- 機械学習をより楽しんで学ぶために、**AWS DeepRacer** は2台でレース可能に、音楽を機械学習でアレンジする **AWS DeepComposer** も発表

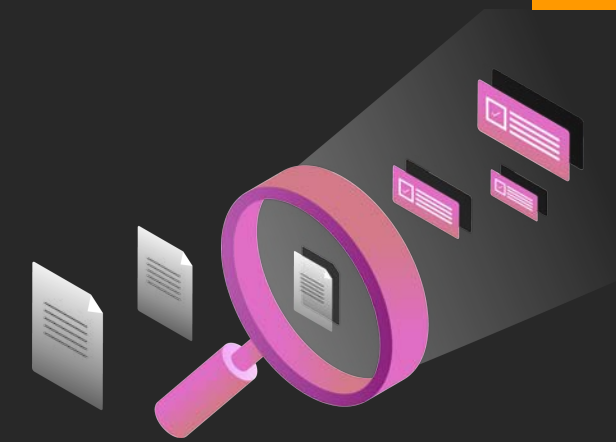


AI サービスアップデート

Amazon Kendra

Preview

- 機械学習を利用した高精度な文書検索
- 自然文の質問に対応する回答の検索
- 関連文書をランキング表示
- コネクタを利用して様々なデータソースを取り込むことが可能
- 学習なしで、IT、金融、保険など、各業種の文章に対応（追加学習も対応予定）
- 対応言語: 米国英語のみ
- リージョン: バージニア、オレゴン、アイランド



What is Amazon Kendra?

Kendra's suggested answer

Amazon Kendra is a highly accurate and easy to use enterprise search service that's powered by machine learning. Kendra delivers powerful natural language search capabilities to your websites and applications so your end users can more easily find the information they need within the vast amount of content spread across your company.

Key Benefits of Kendra

Ask natural language questions, get immediate answers

Use natural language questions instead of just simple keywords to get the answers you're looking for, whether that is a precise answer, an FAQ, or an entire document. Say goodbye to sifting through long lists of links and hoping one has the information you need.

Bring all of your data together in a few clicks

Get rid of information silos. Kendra lets you easily add content from file systems, SharePoint, intranet sites, file sharing services, and more, into a centralized location so you can quickly search all of your information to find the best answer.

Constantly improving search results

Your results get better over time search, because Kendra's machine learning algorithms learn which results your users find most valuable. You also have the option to fine-tune results by manually adjusting the importance of certain data sources or document freshness.

<https://aws.amazon.com/kendra/>

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Amazon Fraud Detector

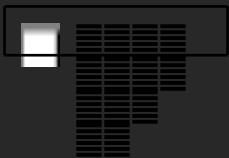
Preview

- 機械学習による不正検知サービス
- 学習なしで利用可能で、以下の既存テンプレートから利用を開始できる
 - アカウント作成
 - オンライン決済、など
- Fraud Detector の予測結果に対して、ルールベースで最終的な不正検知が可能
- リージョン: バージニア



<https://aws.amazon.com/fraud-detector/>

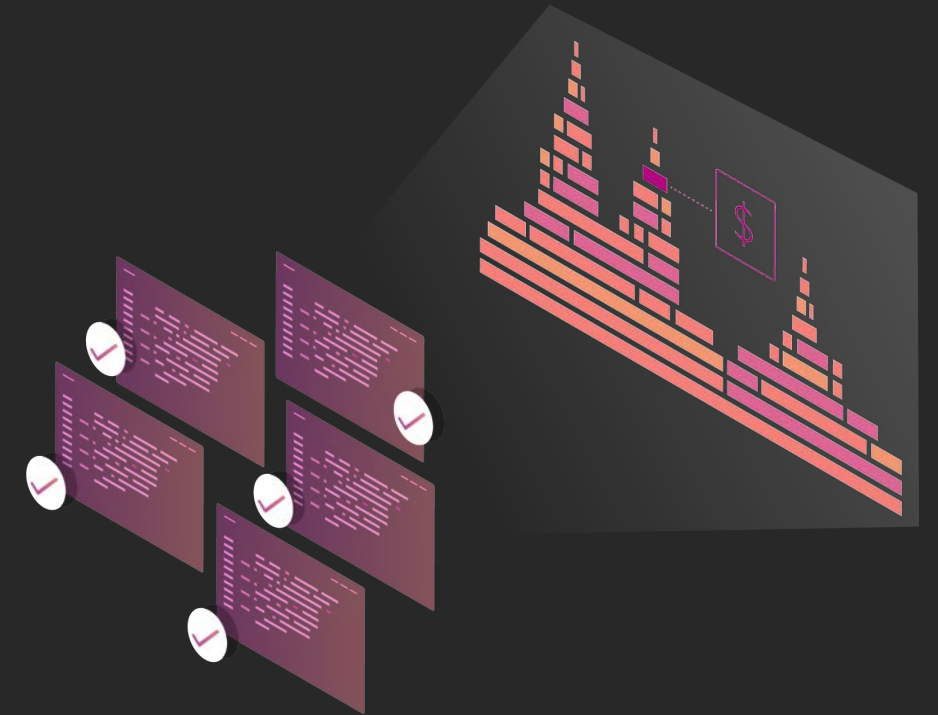
© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Amazon CodeGuru

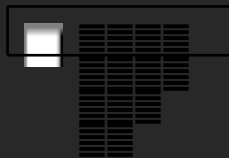
Preview

- コードレビューの自動化と性能改善のためのガイドを行う開発者向けのサービス
- CodeGuru Reviewer
スレッドセーフでないなどの様々な問題を検出する
- CodeGuru Profiler
本番環境で継続的に稼働し、アプリケーションパフォーマンスの最適化を図る
- リージョン: バージニア、オレゴン、オハイオ、アイルランド、シドニー



<https://aws.amazon.com/codeguru/>

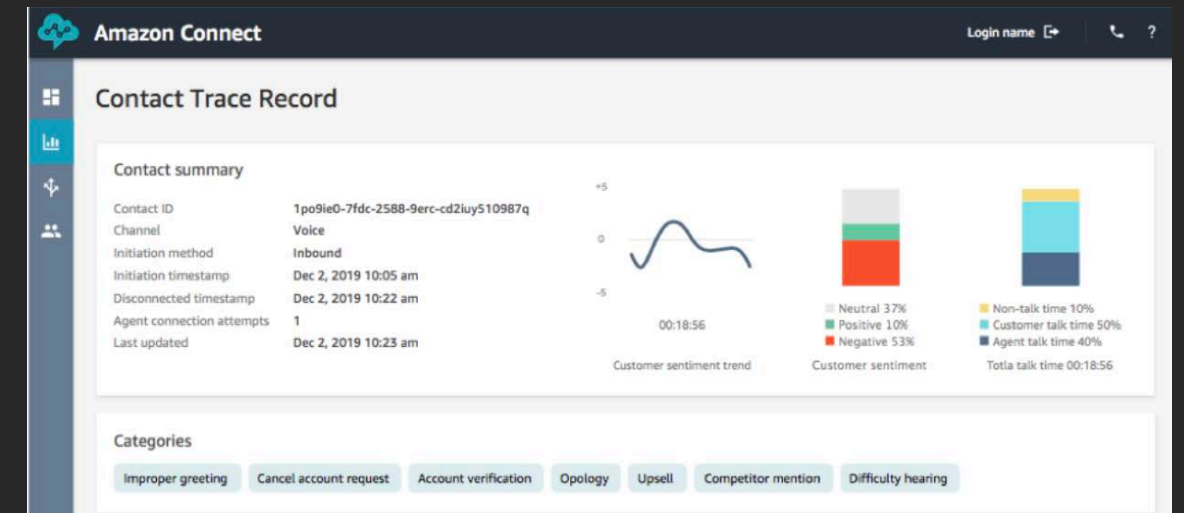
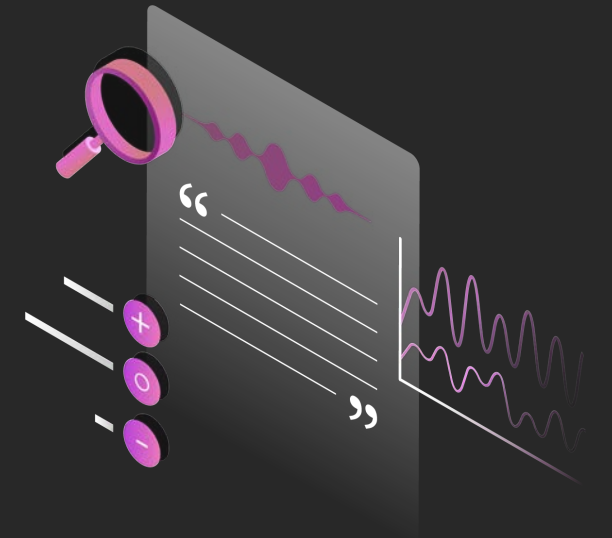
© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Contact Lens for Amazon Connect

Preview

- コンタクトセンターにおける業務を AI で効率化するためのサービス
- 通話やチャットのセンチメント分析
- 問い合わせ内容の自動分類
- 会話が途絶えている状況の自動検出、など
- 分析のための UI を提供
- 対応言語: 米国英語



<https://aws.amazon.com/connect/contact-lens/>

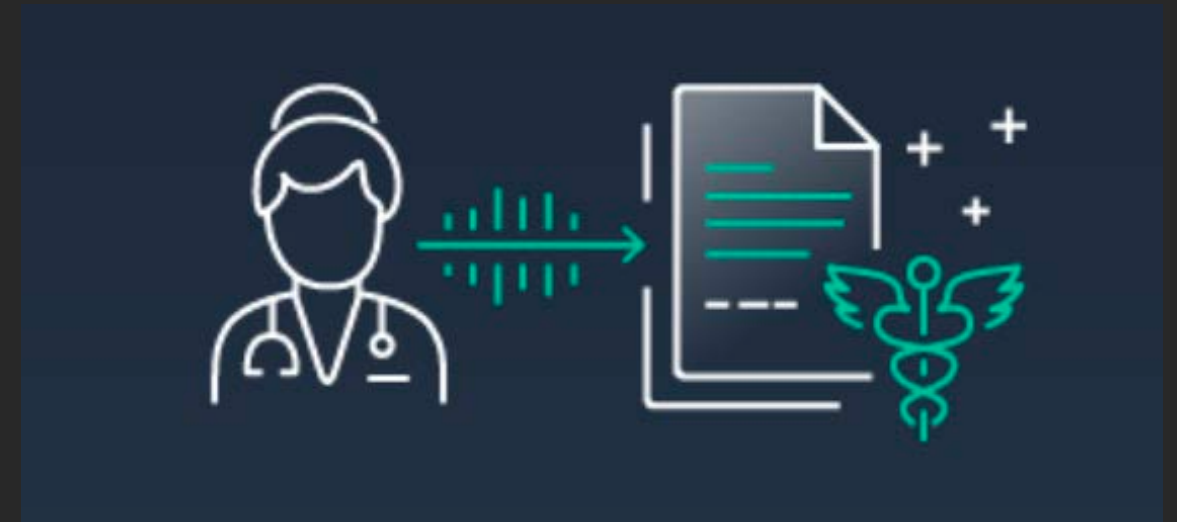
© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Amazon Transcribe Medical

GA

- 医療向けの高精度な音声文字起こしサービス
- 医者と患者の会話 (2人) と医者の読み上げ (1人) それぞれの文字起こしに対応
- Amazon Comprehend Medical と組み合わせて会話内容の分析が可能
- 対応言語: 米国英語
- リージョン: バージニア、オハイオ、オレゴン、シドニーなど



<https://aws.amazon.com/transcribe/medical/>

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

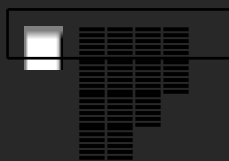


- The screenshot shows the Label Studio interface. The main image is a photograph of a torque converter, a mechanical component used in automatic transmissions. A green bounding box is drawn around the torque converter. The label 'torque converter' is entered in the text field on the right. The interface includes a top bar with 'img_ttc_2.jpg (1/1)', navigation buttons 'Previous (shift + <)' and 'Next (shift + >)', and a 'Done (shift + d)' button. The bottom bar contains icons for 'Box', 'Delete', 'Undo', 'Redo', 'Zoom in', 'Zoom out', 'Move', and 'Fit image'.

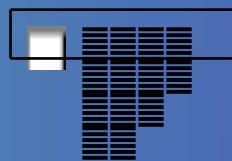


Amazon Rekognition

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



ML サービスアップデート Amazon SageMaker



© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



これまでの Amazon SageMaker

機械学習のエンジニアは、**これまでの開発・学習・デプロイの方式を大きく変えることなく、機械学習プロセス全体を効率化できる**

- 開発・学習・デプロイのインフラ管理は不要
- 既存のコード・コンテナなどを持ち込んで学習・推論可能
- ラベリング・モデル変換などもサポート



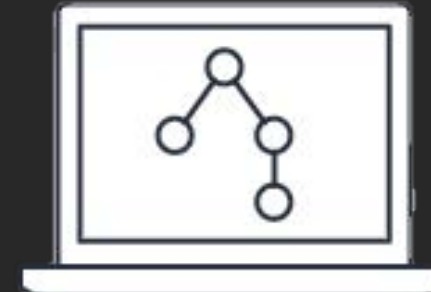
ラベリング



開発



学習



モデル変換・デプロイ



機械学習で残されている課題

開発

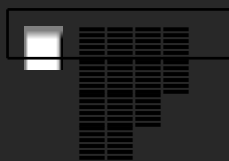
- 各ユーザの環境で開発が実施され、**共同開発が難しい**

学習

- 前処理やアルゴリズム選択など**必要な処理を手動で組む必要**がある
- 良い結果を得るための、学習の繰り返し、**結果の管理・分析が煩雑**
- 悪い結果が出た場合の、学習過程を解析する**デバッガー**がない

デプロイ

- デプロイして終わりではなく、デプロイ後の**モデルの監視**が必要
- 機械学習の予測結果を検証するため、**人間の介入**が求められる場合も



Amazon SageMaker の Update 一覧

開発

学習

デプロイ

Amazon SageMaker Studio
機械学習のための統合開発環境

手動で組んでいた作業の自動化

Amazon SageMaker Autopilot
モデルを自動で構築・学習

Amazon SageMaker Processing
学習と推論テストのための前処理・後処理

共同開発・柔軟
な環境変更

**Amazon SageMaker
Notebooks**
柔軟なノートブック

結果の管理・分析

**Amazon SageMaker
Experiments**
実験管理 (学習などの管理)

デバッガー

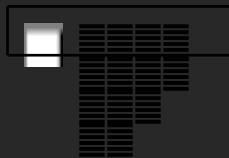
AWS SageMaker Debugger
学習のデバッグ

モデルの監視

**Amazon SageMaker
Model Monitor**
モデルの性能監視

人間の介入

Amazon Augmented AI
推論結果を人間が評価

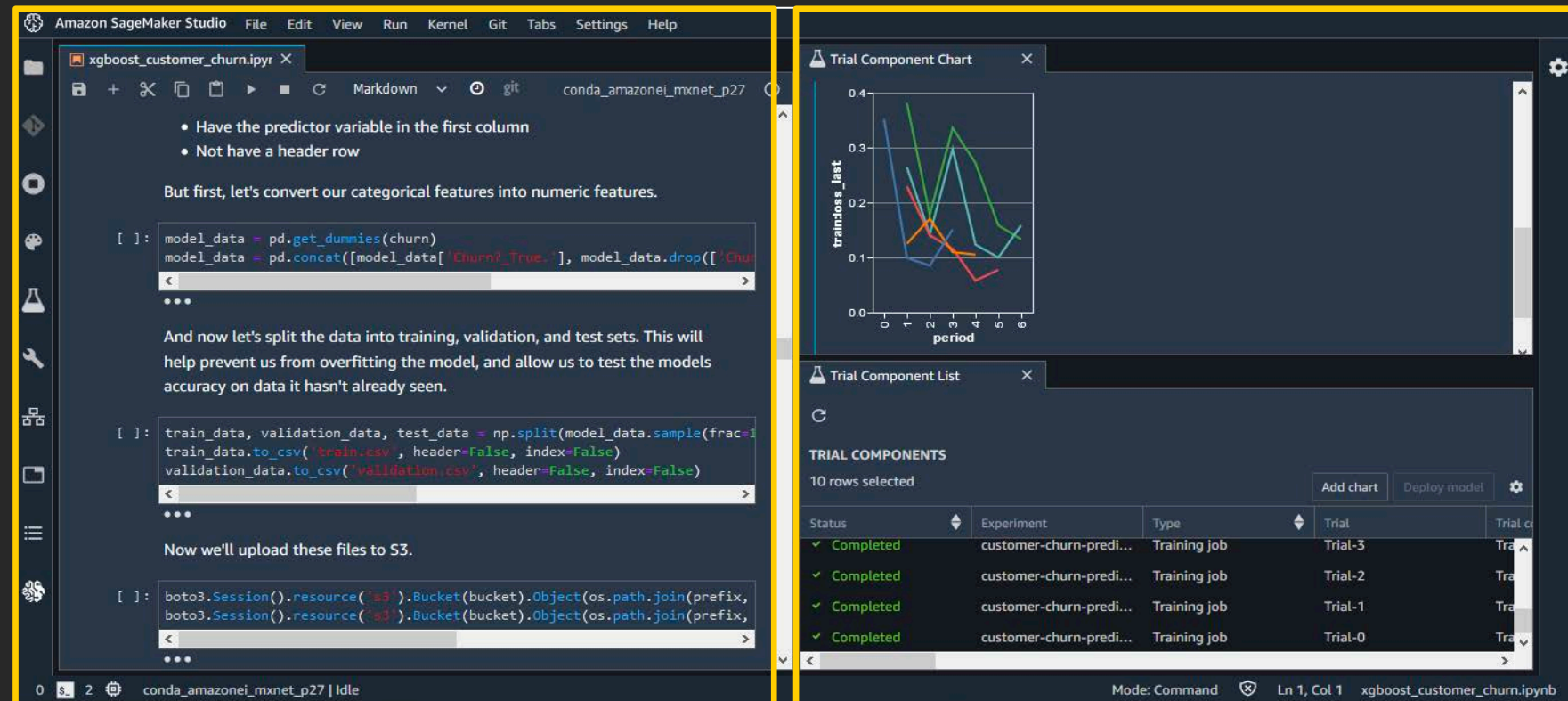


Amazon SageMaker Studio

GA

- SageMaker の各機能呼び出すことができる統合開発環境
- AWS IAM 以外に AWS Single Sign-On (AWS SSO) でもアクセス可能
- リージョン: オハイオ

Amazon
SageMaker
Notebooks



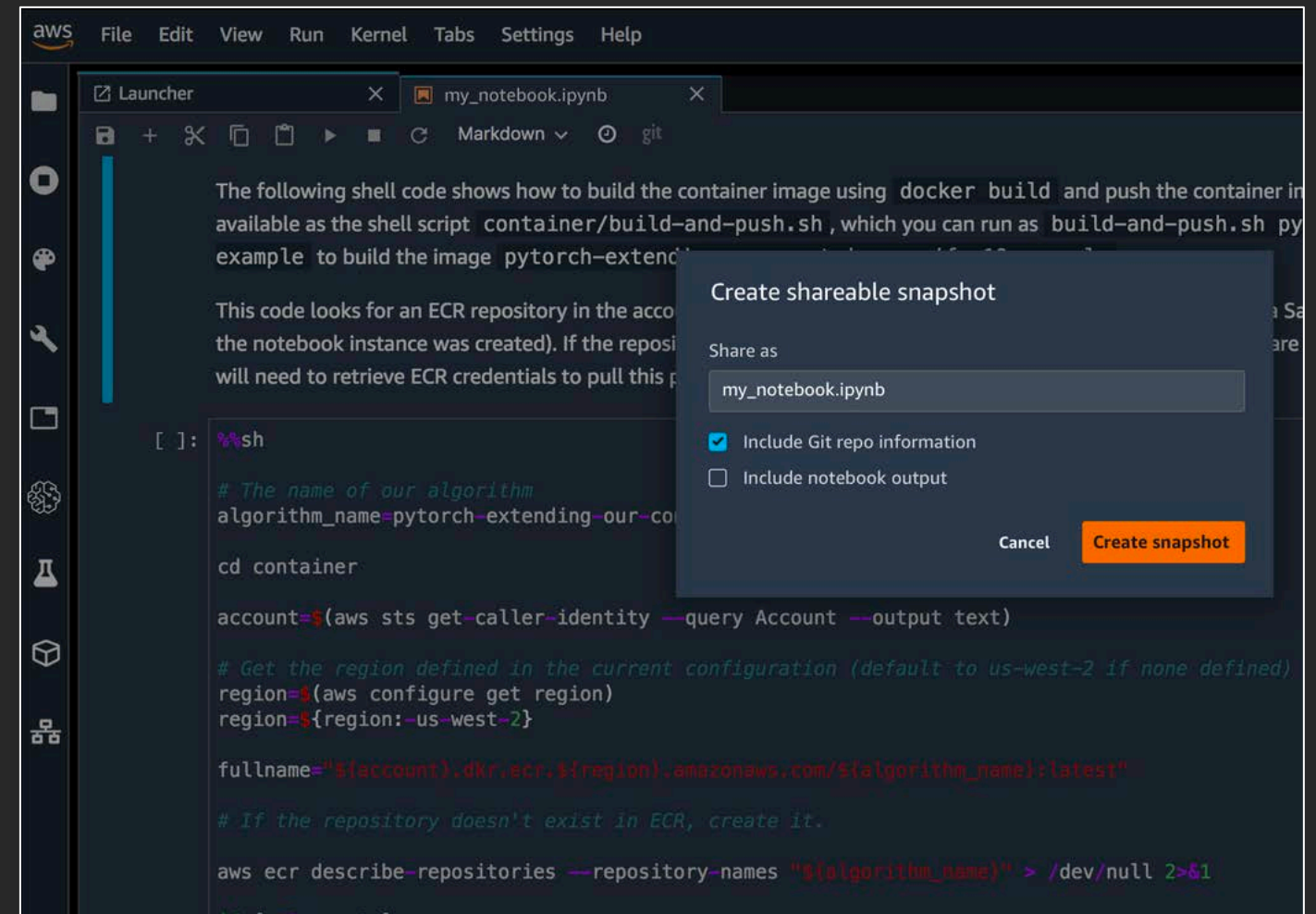
Amazon
SageMaker
Experiments

<https://aws.amazon.com/jp/blogs/news/amazon-sagemaker-studio-the-first-fully-integrated-development-environment-for-machine-learning/>

Amazon SageMaker Notebooks

Preview

- AWS SSO でのアクセスが可能なノートブック
- リソースの拡大・縮小を再起動無しで行うことが出来る
- ワンクリックでノートブックをシェア可能



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

Amazon SageMaker Autopilot

GA

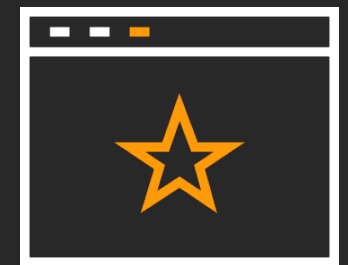
- 表形式のデータに対する予測と分類の機械学習を自動化
- 不正検知、ユーザ分類（離反予測）、リコメンデーションなど多数のユースケース
- 最低限、データと予測対象を指定
- データの前処理、アルゴリズムの選択、インスタンスタイプの選択なども自動化

欠損や非数値を含む表形式のデータ

	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



自動でモデル作成



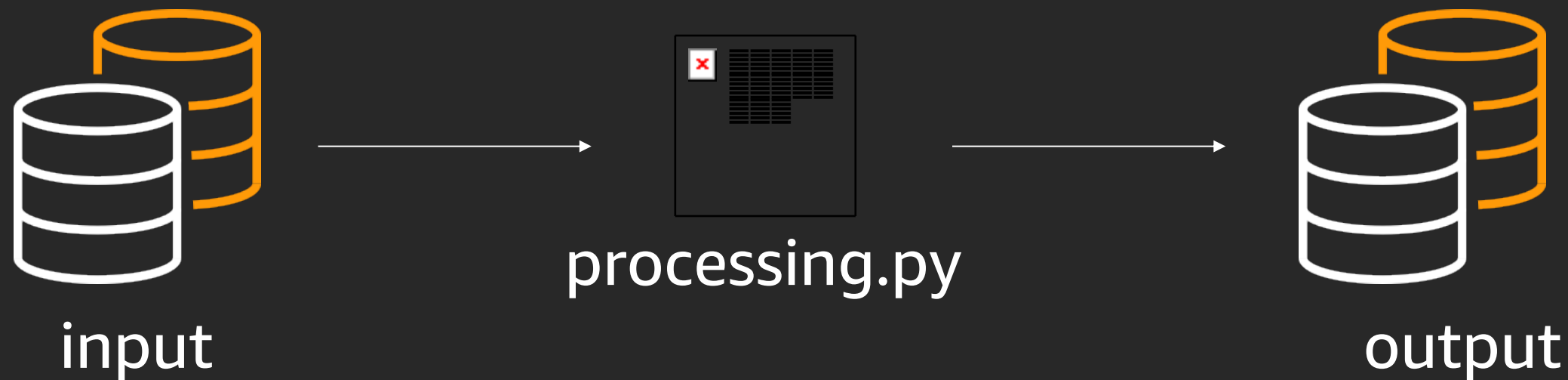
リーダーボード機能

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

Amazon SageMaker Processing

GA

- 機械学習で必要となるデータの前処理・後処理をバッチで行う
- 学習・テストデータの一括整形、推論結果からのフィルタリングなど
- scikit-learn コンテナを利用して処理可能 (コンテナ持ち込みも可能)
- 処理開始時にインスタンスが起動、処理終了時にインスタンスも終了



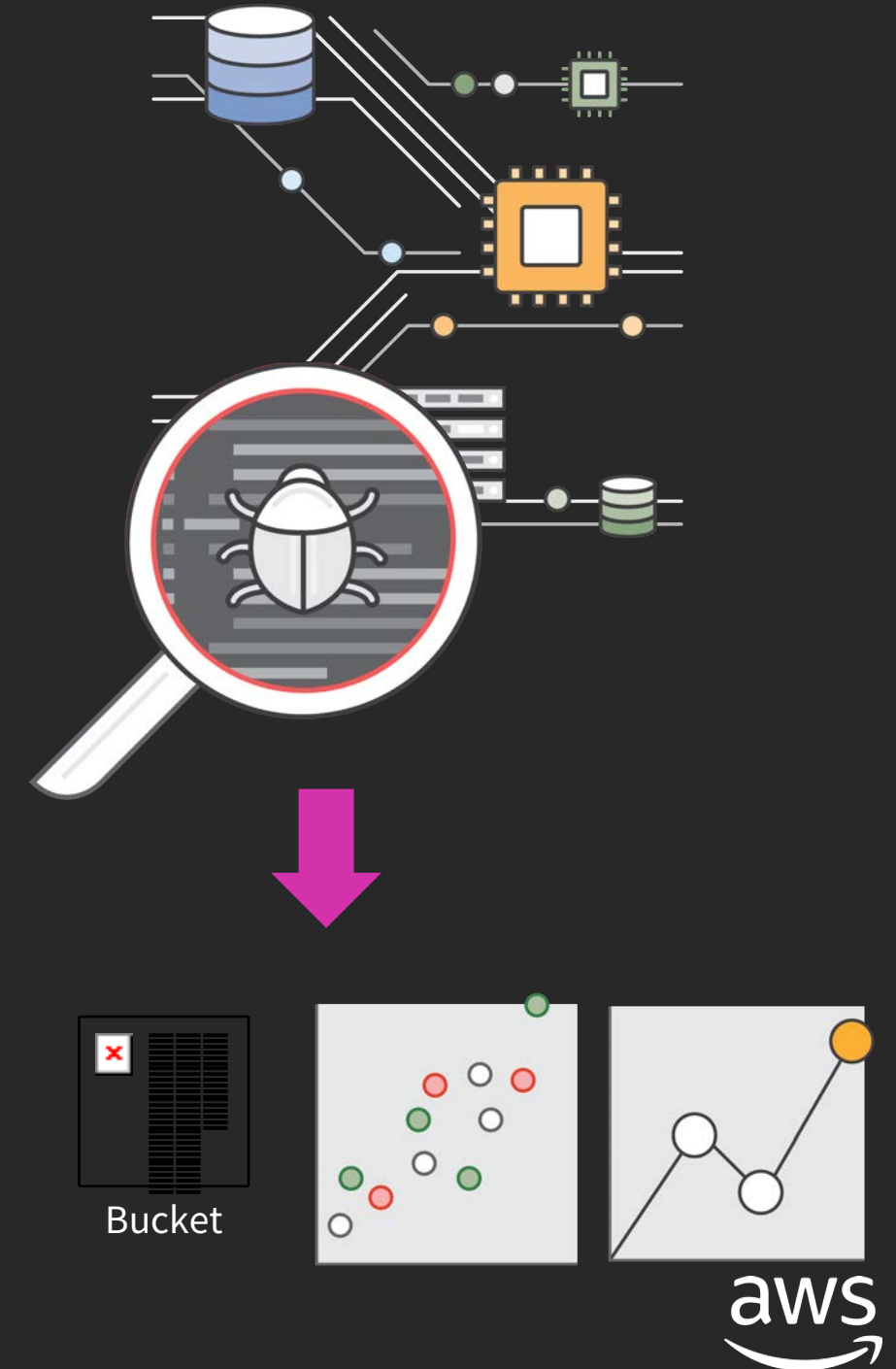
<https://aws.amazon.com/blogs/aws/amazon-sagemaker-processing-fully-managed-data-processing-and-model-evaluation/>

Amazon SageMaker Debugger

GA

- 学習時に不正な数値が出力される等の問題を検出する
- 問題検出のルールがあらかじめ与えられており（ルール追加も可能）、S3に出力される学習の途中計算結果を見てデバッグ可能
- Debug のコードを機械学習のコードに簡単に埋め込むための SageMaker Debugger SDK
- TensorFlow, Keras, Apache MXNet, PyTorch, XGBoost に対応

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



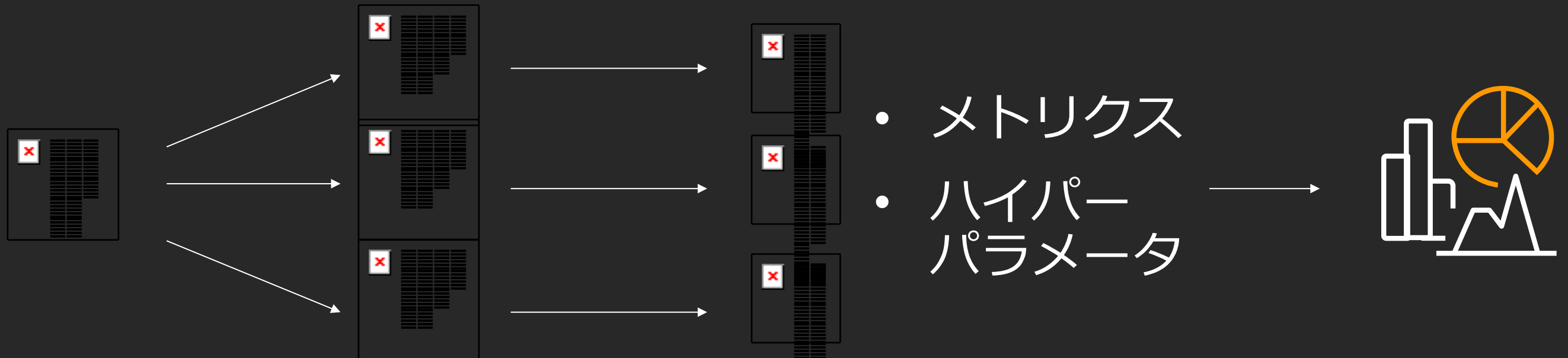
Amazon SageMaker Experiments

GA

- 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、グラフ化も可能な表形式のデータで確認できる

学習ジョブ

学習の結果

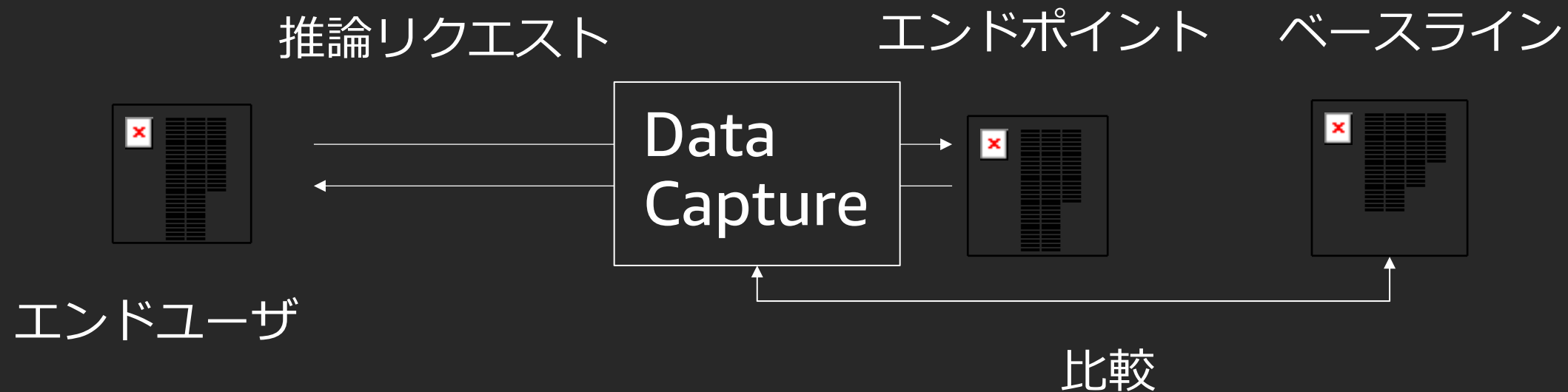


<https://aws.amazon.com/jp/blogs/news/amazon-sagemaker-experiments-organize-track-and-compare-your-machine-learning-trainings/>

Amazon SageMaker Model Monitor

GA

- SageMaker の推論エンドポイントの入出力を逐次監視し、ベースライン (主に学習に利用したデータ) と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

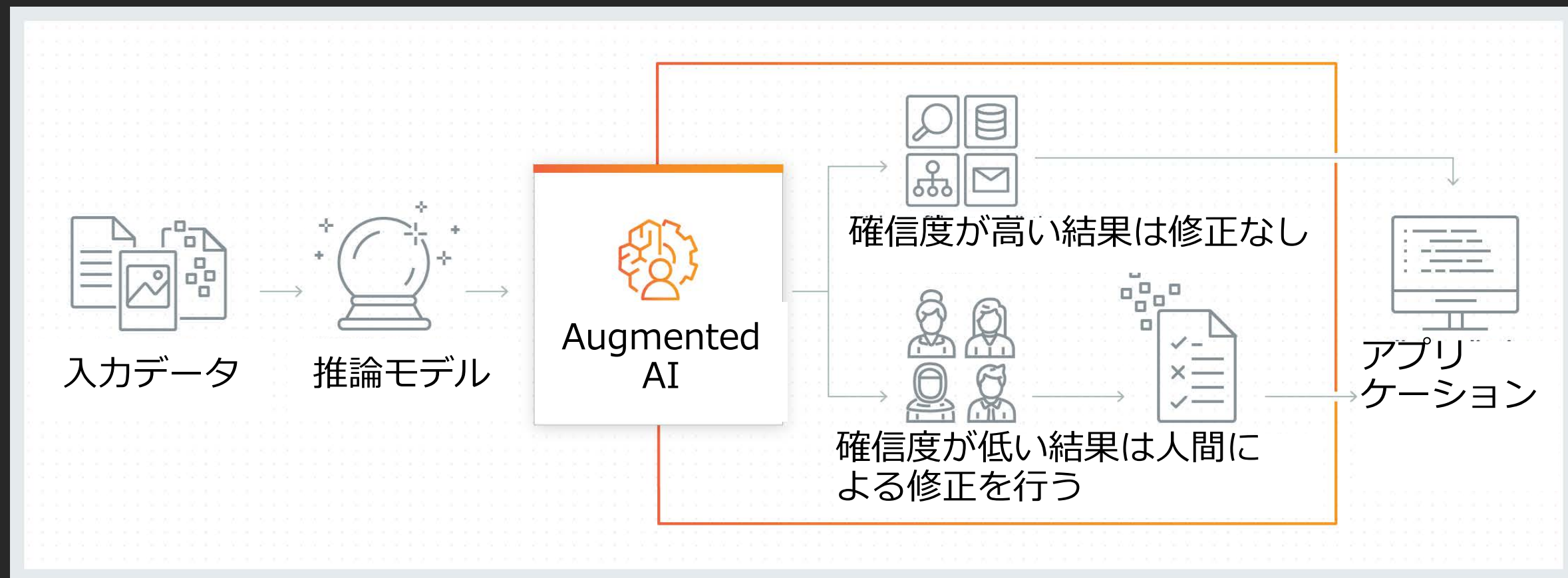


<https://aws.amazon.com/about-aws/whats-new/2019/12/introducing-amazon-sagemaker-model-monitor/>

Amazon SageMaker Augmented AI

Preview

- 推論結果を人間が修正するワークフローを構築する機能
- Ground Truth と同様、テンプレートを構築してワーカーにレビュー依頼



<https://aws.amazon.com/about-aws/whats-new/2019/12/announcing-amazon-augmented-ai/>

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Amazon SageMaker Operators for Kubernetes

GA

- Kubernetes を利用している開発者が、Amazon SageMaker を利用して機械学習モデルの学習・デプロイするための機能
- Kubernetes API や kubectl などのツールを利用して SageMaker を呼び出し、SageMaker でジョブの作成・操作が可能

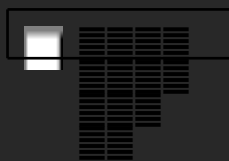


Amazon SageMaker



Amazon Elastic
Kubernetes Service

<https://aws.amazon.com/blogs/machine-learning/introducing-amazon-sagemaker-operators-for-kubernetes/>



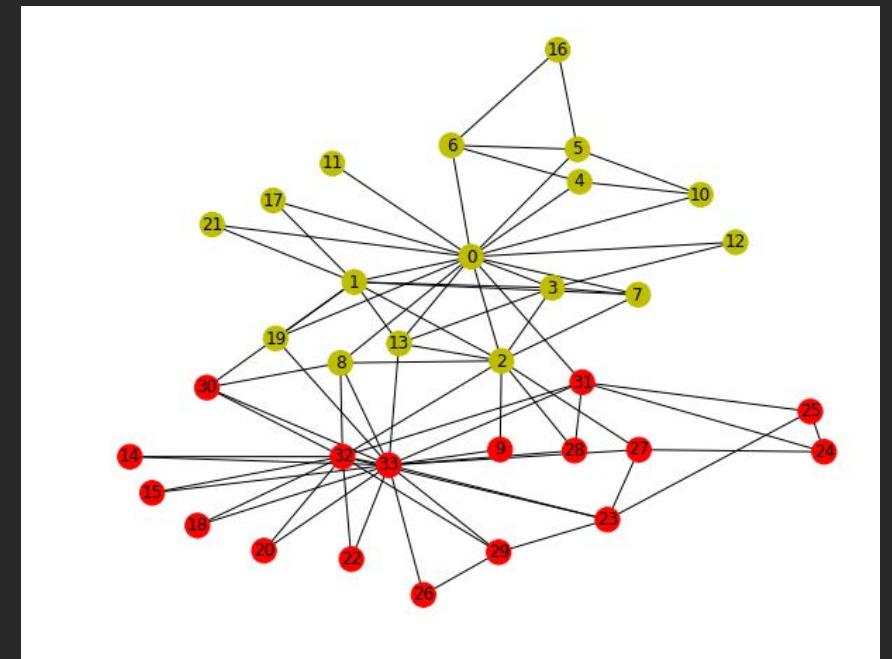
Deep Graph Library が SageMaker で利用可能

GA

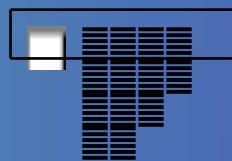
- ソーシャルネットワーク等のグラフを対象にしたニューラルネットワークを構築するライブラリ Deep Graph Library (DGL) をサポート
- 化学物質解析、リコメンデーション等への応用で注目されている
- DGL はオープンソースで PyTorch や Apache MXNetとの連携が可能

<https://www.dgl.ai/>

DeepGraphLibrary



ML フレームワークと インフラストラクチャ アップデート



Inf1 インスタンス

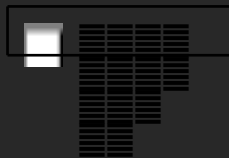
GA

- 機械学習の推論を低レイテンシかつ安価に実現する推論用チップAWS Inferentia を搭載した Inf1 インスタンスを発表
- G4インスタンスと比較して推論処理で最大3倍のスループット、最大40%のコスト削減効果を期待できる
- Amazon Deep Learning AMIs は、Inf1 向けに最適化したTensorFlow と Apache MXNet を含む (PyTorch も対応予定)
- 利用時にモデルのコンパイルを伴い、そのための Neuron SDK を提供
- 以下にも対応予定
 - Amazon SageMaker
 - Amazon ECS
 - Amazon EKS

Inf1ファミリ	Inferentia チップ数	vCPU	メモリ (GiB)	EBS帯域 (Gbps)	NW帯域 (Gbps)
inf1.xlarge	1	4	8	Max 3.5	Max 25
inf1.2xlarge	1	8	16	Max 3.5	Max 25
inf1.6xlarge	4	24	48	3.5	25
inf1.24xlarge	16	96	192	14	100

<https://aws.amazon.com/blogs/aws/amazon-ec2-update-inf1-instances-with-aws-inferentia-chips-for-high-performance-cost-effective-inferencing/>

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



機械学習のハンズオンのための サービスアップデート

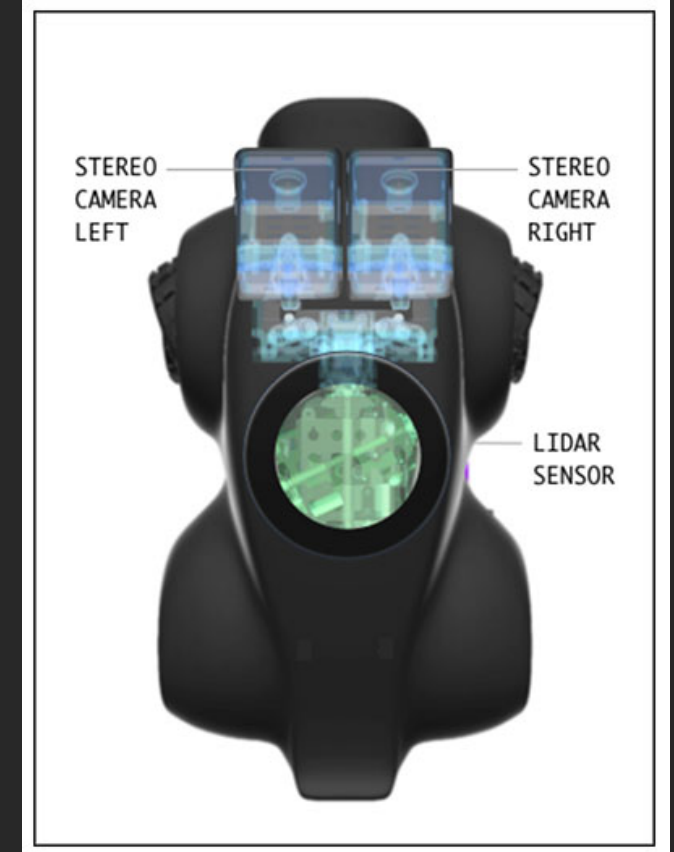


© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



AWS DeepRacer アップデート

- ステレオカメラとLIDAR (光計測と測距)を搭載した AWS DeepRacer Evo を発表
- 2台で競い合う head-to-head なレースが可能になり、障害物回避などにも取り組める
- 学習に利用可能なニューラルネットワークを3層と5層から選択可能に
- 2020年は DeepRacer Evo による新しい大会を予定



<https://aws.amazon.com/jp/blogs/aws/aws-deepracer-update-new-features-new-racing-opportunities/>



AWS DeepComposer

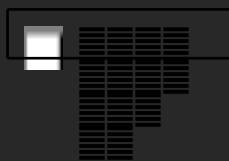
Preview

- メロディを入力すると、機械学習でジャンルに応じた伴奏を生成する
- 伴奏生成に利用されている GAN (Generative Adversarial Nets) を調整しながら仕組みを学べる
- キーボードは Amazon.com にて販売予定



<https://aws.amazon.com/deepcomposer/>

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

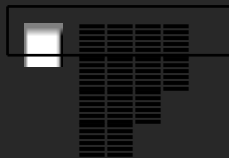


まとめ



まとめ

- AI サービスは、文書検索、不正検知など、独自に機械学習モデルを構築することが一般的だったユースケースに対応
- 機械学習向けの IDE である SageMaker Studio では効率よく機械学習を進めるための機能を簡単に利用可能
- inf1 インスタンスは、Production で課題となる推論コストの低減に有効
- DeepLens, DeepRacerにつづき DeepComposerを発表



ML re:Cap イベントのご案内

AWS re:Invent 2019 RE:CAP | AI/ML

日時: 2019年12月12日 (木)

開場: 13:45 開演: 14:15~17:15

(終了時刻変更の可能性あり)

<申込期限>12月10日(火) 17:00 まで

会場: アマゾン目黒オフィス

〒141-0021 東京都品川区上大崎3-1-1

目黒セントラルスクエア 21F



詳細は <https://aws-seminar.smktg.jp/public/seminar/view/892>

