

クイックスタート機械学習 Amazon SageMaker 編

尾原 颯

技術統括本部 ソリューションアーキテクト
Amazon ウェブ サービス ジャパン合同会社

自己紹介



尾原 颯

➤ 所属/役職：
ソリューションアーキテクト

➤ 好きなサービス：



Amazon SageMaker



AWS Lambda

本セッションの対象者



ビジネスアナリスト

業務のデータ解析で機械学習を
手軽に使ってみたいけど、
第一歩の踏み出しがわからない



エンジニア

社内に蓄積されたデータを活用した新機能実装のために
機械学習のモデルを作成してみたい
コードを使った柔軟な解析もしてみたい

本セッションのゴール

個人のレベル感に合わせて
AWS を利用した機械学習活用の第一歩を
踏み出すことができる

機械学習活用の実現方法

本セッションでは Amazon SageMaker での実現方法について扱う

AWS

AWS Partner

Do it yourself

AWS Professional Services

AWSクラウドを使用して期待する
ビジネス上の成果を実現できるよう
お客様をサポートできる、専門家からなる
グローバルチーム

Amazon ML Solutions Lab

チームと機械学習の専門家を組み合わせて組織
の投資収益率が最も高い機械学習に
対処する機械学習ソリューションを特定して
構築するよう支援

Find an AWS Partner

お客様のクラウドジャーニーを加速しAWSが
提供する全ての利点をお客様が
最大限活用するように支援

AWS Artificial Intelligence Services

MLの知識不要で簡単にアプリケーションに
インテリジェンスを追加できるサービス

Amazon SageMaker

フルマネージドインフラストラクチャ、
ツール、ワークフローを使用してあらゆるユースケース
向けの機械学習（ML）モデルを構築、
トレーニング、デプロイ

AWS-40「ビジネス課題を解決するための
AWS AI サービスの始め方」を参照

本セッションで扱う範囲



アジェンダ

1. データ活用に使われる技術
2. データ活用にまつわる課題
3. AWS を用いた機械学習の始め方 2 選
4. 機械学習習得のためのリソースの紹介
5. まとめ

データ活用に使われる技術

データサイエンス

データからビジネス的なインサイトを得るアプローチ

何が起きた？



どうして起きた？



次に何が起こる？

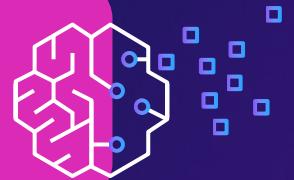


データサイエンス

統計



機械学習



機械学習実装の難点

データの曖昧性、コスト、スキルの不足、複雑さ

データ活用における課題

機械学習を活用してやりたいこと



ビジネスアナリスト

各店舗からの日毎の詳細なデータを集積し可視化できるようになった。
在庫管理の最適化のために、これらのデータを使って
未来の売り上げを予測したい。



エンジニア

サービス利用の感想を SNS 上から集めることはできたが、
そこからサービス改善のヒントとなるものを
自動的にピックアップしたい。

機械学習を活用するとできること

機械学習モデルから得られるインサイトからビジネス成果を達成

✓ 解約の削減

✓ メンテナンスの予測

✓ 財務指標の予測

✓ 不正検知

✓ 市場動向の検知

✓ 在庫最適化

✓ 売上予測

機械学習の専門知識が必要

外部からスペシャリストを雇うことは難しい

25万人

Quanthub は2020年の
データサイエンス人材の不足を
推定

12.2万件

2021年11月時点の
LinkedInにおける
データサイエンスジョブ
募集数

13%

VentureBeat による
DSMLプロジェクトが
製品化された割合

機械学習習得の難しさ

どこから学習を始めれば良いかわからない



初期段階で学ぶことが多い！

AWS ML Vision

機械学習をすべての人に

AWS での機械学習の始め方 2 選



機械学習の始め方 2 選

1. ノーコードで機械学習を実践



Amazon SageMaker Canvas

2. 統合的な開発環境を利用した更なる解析



Amazon SageMaker Studio

機械学習の始め方 2 選

1. ノーコードで機械学習を実践



Amazon SageMaker Canvas

2. 統合的な開発環境を利用した更なる解析



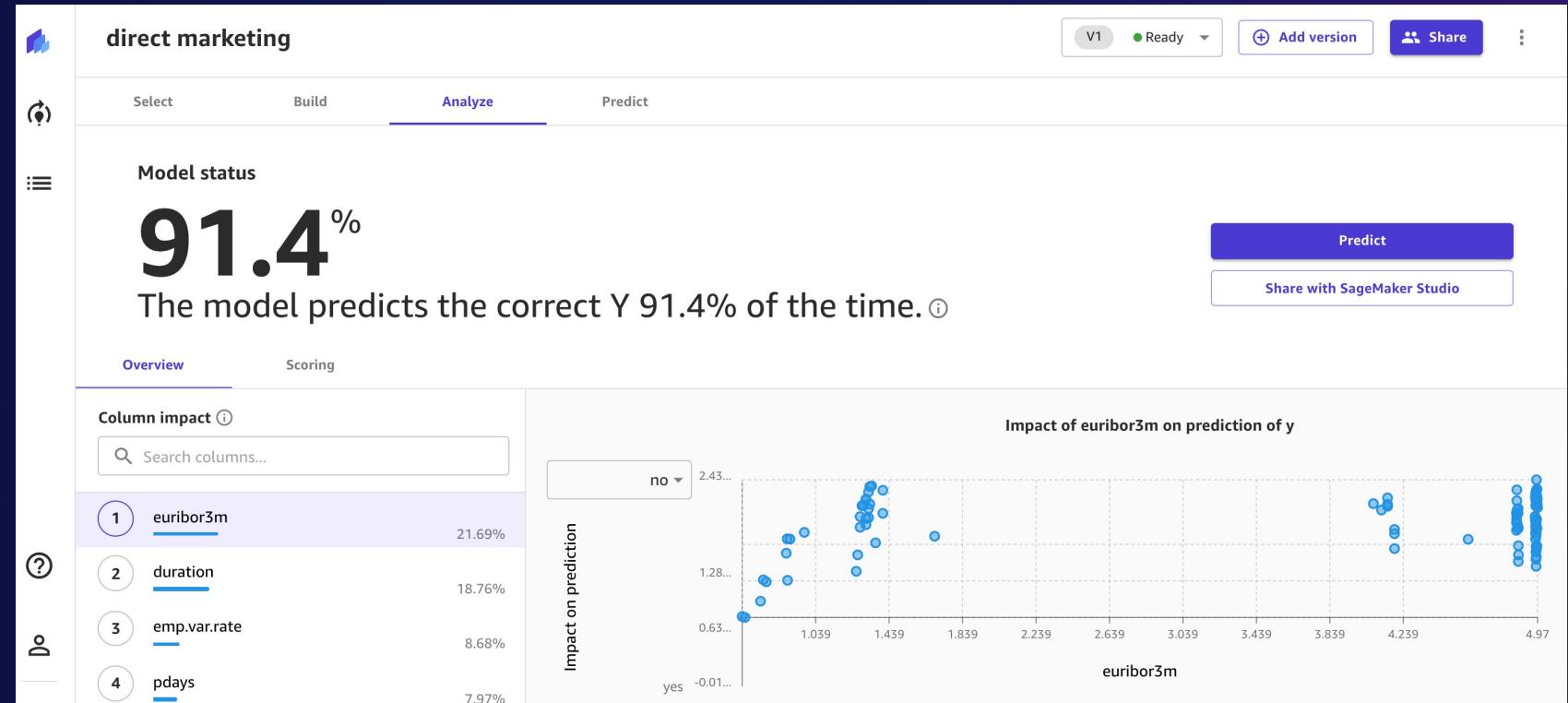
Amazon SageMaker Studio

Amazon SageMaker Canvas

ノーコードで、GUIでモデルを作成し予測を得られる AutoML サービス
機械学習の経験なしで始められる



Amazon SageMaker Canvas



SageMaker Canvas でできること

データの準備、モデリング、推測までを一気通貫で実施が可能



多様なデータソースへの
迅速な接続とアクセス



AutoML 技術をつかった
自動的な学習とモデルのビルト



作成したモデルでの
単体もしくはバッチ推論



SageMaker Studio
との連携による、
手軽なモデルの共有

多様なソースからデータを統合



多様なデータソースへの
迅速な
接続とアクセス

Import Data

Upload S3 Snowflake Redshift Crystal 1 Crystal 1

Connection name: ドラッグアンドドロップ

database1 database2 database3 database4

schema1 schema2

table1

S3からの読み込み

Snowflakeからの読み込み

Amazon Redshiftからの読み込み

Import preview

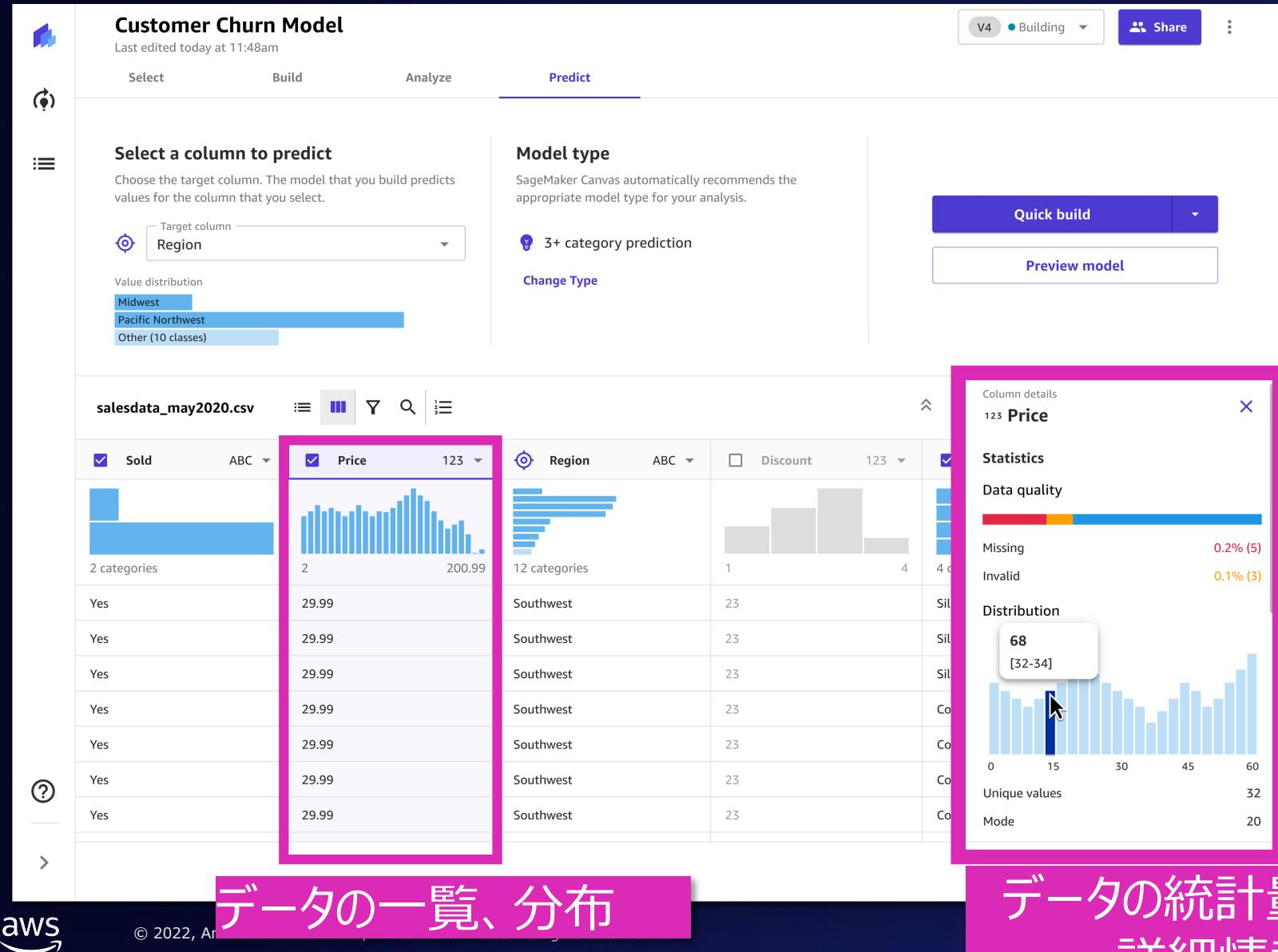
Sold	ABC	Price	123	Region	ABC	Discount	123	Fabric	ABC	Age	123
Yes		29.99		Southwest		23		Cotton		27	
Yes		29.99		Southwest		23		Silk		35	
Yes		29.99		Southwest		23		Silk		32	
Yes		29.99		Southwest		23		Silk		32	
Yes		29.99		Southwest		23		Cotton		30	

Previewing the first 100 rows

Close Import data

- ドラッグ・アンド・ドロップでCSVファイルをインポート可能
- Amazon Simple Storage Service (Amazon S3)、Amazon Redshift、Snowflakeといったソースからデータを統合可能

データを視覚的に理解



多様なデータソースへの 迅速な 接続とアクセス



- データの分布や
基本的な統計量、
欠損値の割合など
データ分析に
有用な情報を確認
 - 例：二値分類で、
目的変数との相関を確認

データの一覧、分布

データの統計量などの 詳細情報

ワンクリックでモデルを構築

Select a column to predict

Choose the target column. The model that you build predicts values for the column that you select.

Target column: Region

Region distribution: West (selected), Pacific Northwest, Other (10 classes)

data_may2020.csv

Order	ABC	Region	Price
1	✓	West	29.99
2		Pacific Northwest	29.99
3		Other (10 classes)	29.99
4		Southwest	23
5		Southwest	23
6		Southwest	23
7		Southwest	23
8		Southwest	23
9		Southwest	23
10		Southwest	23
11		Southwest	23
12		Southwest	23
13		Southwest	23
14		Southwest	23
15		Southwest	23
16		Southwest	23
17		Southwest	23
18		Southwest	23
19		Southwest	23
20		Southwest	23
21		Southwest	23
22		Southwest	23
23		Southwest	23
24		Southwest	23
25		Southwest	23
26		Southwest	23
27		Southwest	23
28		Southwest	23
29		Southwest	23
30		Southwest	23
31		Southwest	23
32		Southwest	23
33		Southwest	23
34		Southwest	23
35		Southwest	23
36		Southwest	23
37		Southwest	23
38		Southwest	23
39		Southwest	23
40		Southwest	23
41		Southwest	23
42		Southwest	23
43		Southwest	23
44		Southwest	23
45		Southwest	23
46		Southwest	23
47		Southwest	23
48		Southwest	23
49		Southwest	23
50		Southwest	23
51		Southwest	23
52		Southwest	23
53		Southwest	23
54		Southwest	23
55		Southwest	23
56		Southwest	23
57		Southwest	23
58		Southwest	23
59		Southwest	23
60		Southwest	23
61		Southwest	23
62		Southwest	23
63		Southwest	23
64		Southwest	23
65		Southwest	23
66		Southwest	23
67		Southwest	23
68		Southwest	23
69		Southwest	23
70		Southwest	23
71		Southwest	23
72		Southwest	23
73		Southwest	23
74		Southwest	23
75		Southwest	23
76		Southwest	23
77		Southwest	23
78		Southwest	23
79		Southwest	23
80		Southwest	23
81		Southwest	23
82		Southwest	23
83		Southwest	23
84		Southwest	23
85		Southwest	23
86		Southwest	23
87		Southwest	23
88		Southwest	23
89		Southwest	23
90		Southwest	23
91		Southwest	23
92		Southwest	23
93		Southwest	23
94		Southwest	23
95		Southwest	23
96		Southwest	23
97		Southwest	23
98		Southwest	23
99		Southwest	23
100		Southwest	23

Model type

SageMaker Canvas automatically recommends the model type based on your data.

Model type

Recommended

3+ category prediction
Classify rows into three or more categories.
Example business questions

- Will a delivery be early, on time, or late?
- Will a customer be a frequent customer, an average customer, or an infrequent customer?

Other

Numeric prediction
For the {target_column}, your model predicts numeric values.
Example business questions

- How many days will it take for a package to be delivered?

Cancel Change type

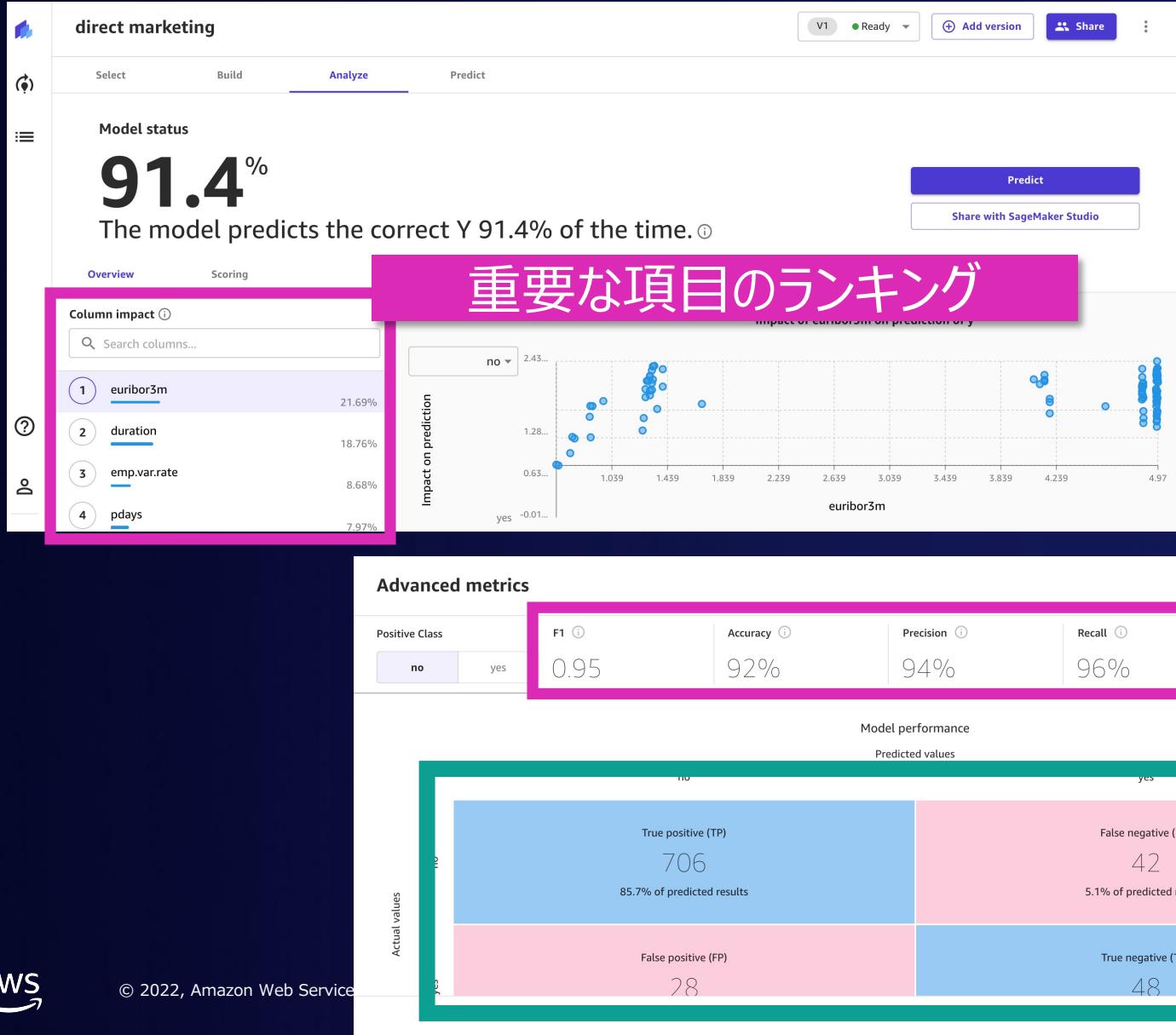
モデルの学習設定



AutoML 技術をつかった
自動的な学習と
モデルのビルト

- ・ クイックビルドと
スタンダードビルドの
2種類のビルト方法を
提供
- ・ まずは手軽に結果を
確認したいなら
クイックビルドを選択

重要な項目や詳しい性能指標を確認



AutoML 技術を使った
自動的な学習と
モデルのビルト

- 重要な項目のランキング
- 性能指標の確認

性能指標一覧

混同行列

バッチ予測や what-if シナリオの分析

Customer Churn Model
Last edited today at 11:48am

Select Build Analyze Predict

Predict target values

Batch prediction Single prediction

Modify value

Filter col

Column Column impact Value

Contract 61.3% ABC Two year

OnlineSecurity Month-to-month

TechSupport One year

InternetService Two year

PaymentMethod Fiber optic

OnlineBackup Electronic check

DeviceProtection No

MonthlyCharges Yes

PaperlessBilling 104.8

予測したいデータの値を手動で設定

Batch prediction で CSV ファイル全体に対する予測を実施

Churn prediction

No

Average probability New probability

No 71.5%

Yes 28.5%

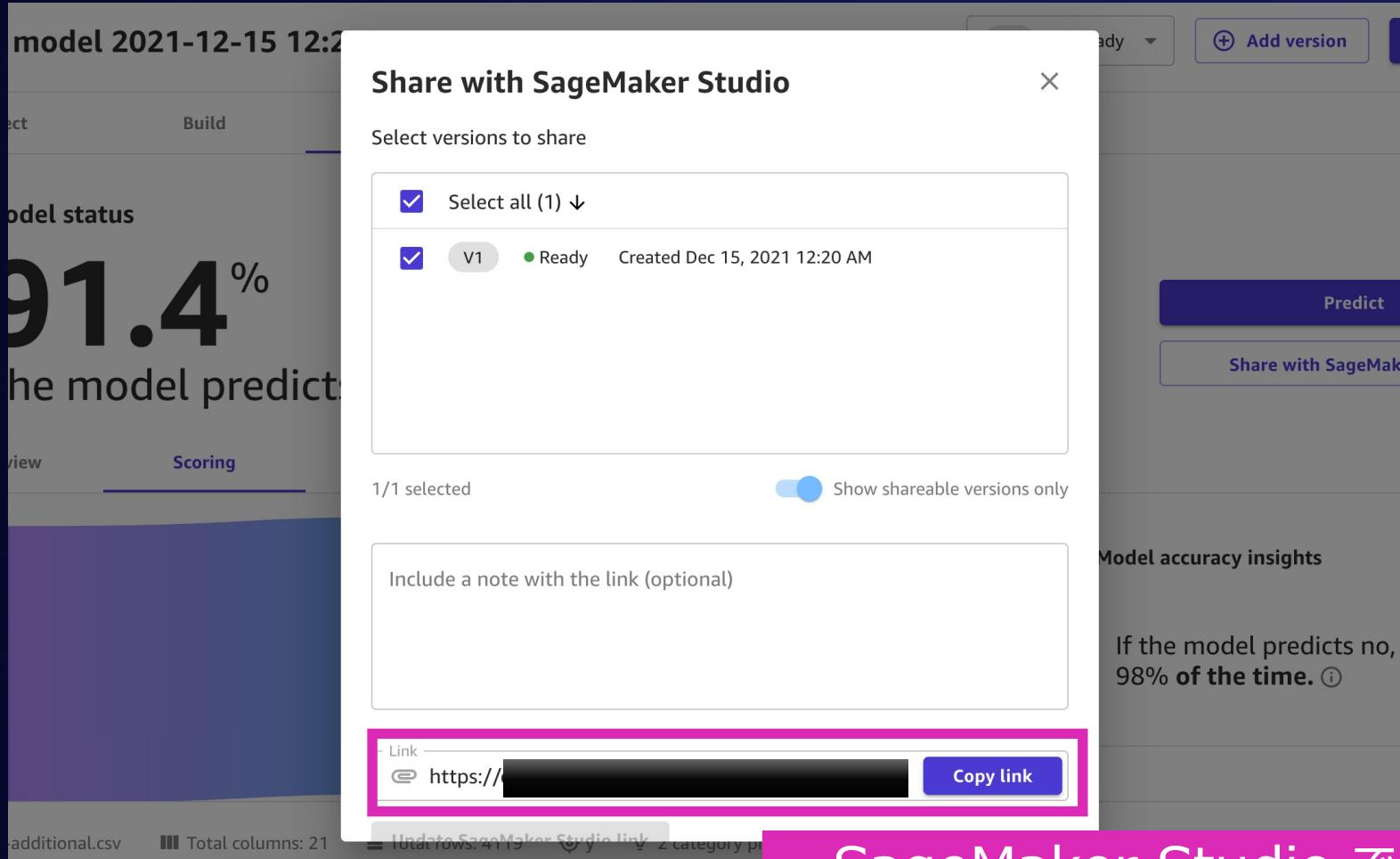
Close Download

© 2022, Amazon Web Services, Inc.



- ブラウザのUIで、特定の入力の特徴量を変更して、what-if シナリオの分析も可能
- CSV ファイルを指定して未知のデータセット全体に対する予測結果を取得

モデルやデータを他のメンバーに簡単に共有



SageMaker Studio との連携による、手軽なモデルの共有

- 共有ボタンをクリックするとすぐにリンクが生成
- 解析の次のステップへのスムーズな移行

SageMaker Studio で共有された
モデルやデータのリンクを取得

機械学習の始め方 2 選

1. ノーコードで機械学習を実践



Amazon SageMaker Canvas

2. 統合的な開発環境を利用した更なる解析



Amazon SageMaker Studio

SageMaker Canvas のその先へ



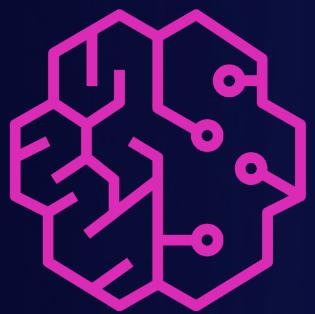
ビジネスアナリスト

SageMaker Canvas を使って、手軽にモデルを作成できた！
今度は、学習に使うデータとして商品画像を加えてみたい。



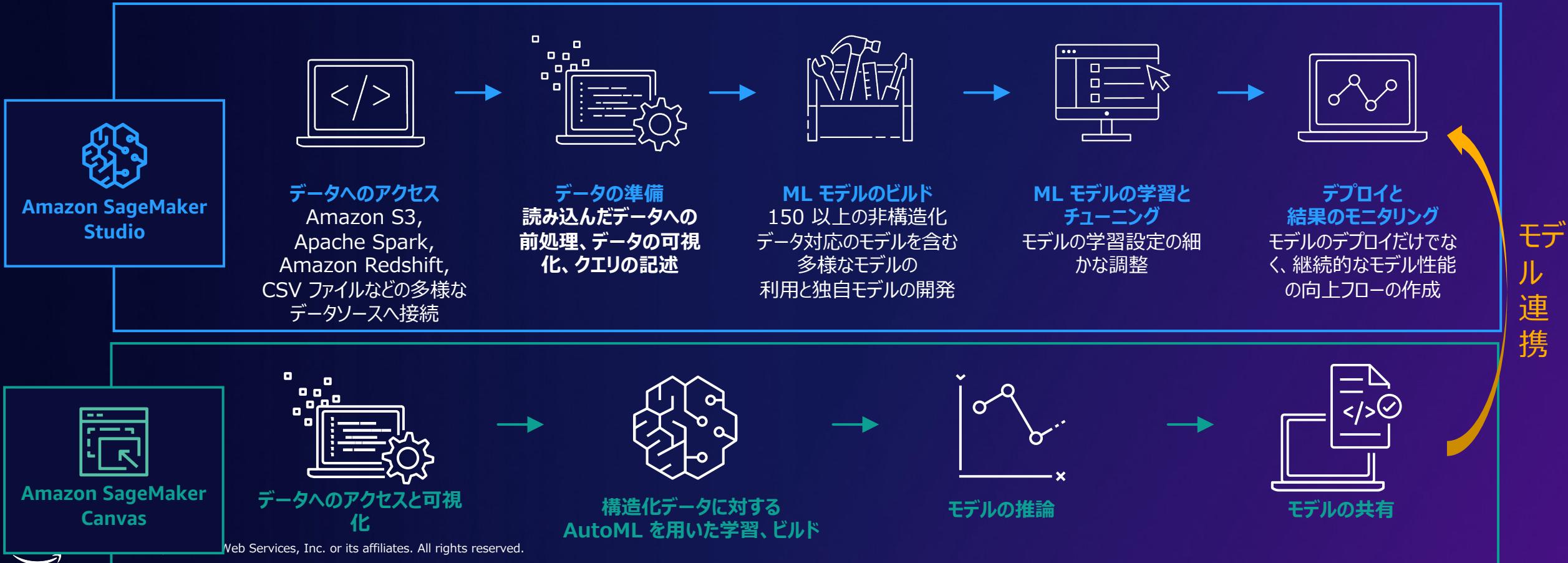
エンジニア

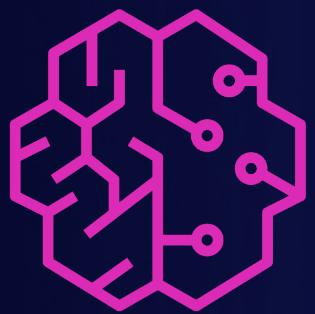
SageMaker Canvas ではまだ対応していない自然言語処理を使った
サービスを開発したい。



Amazon SageMaker

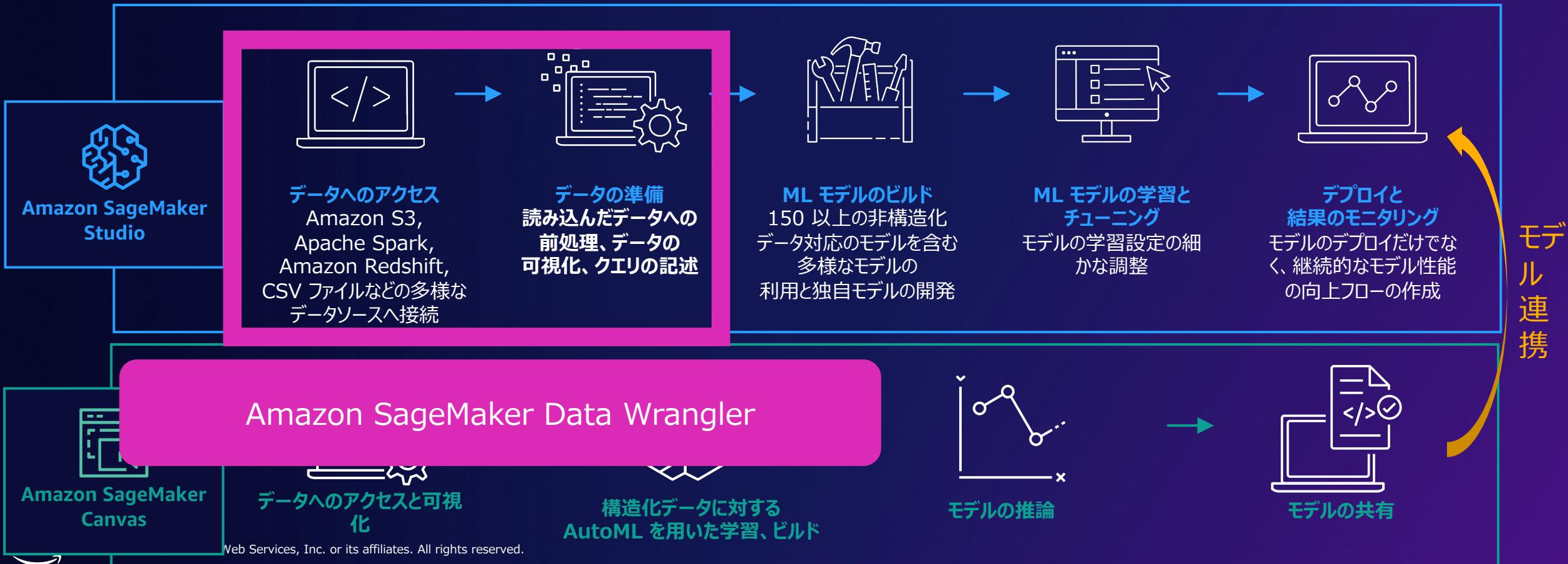
目的別に作成されたツール群。生産性を 10 倍に





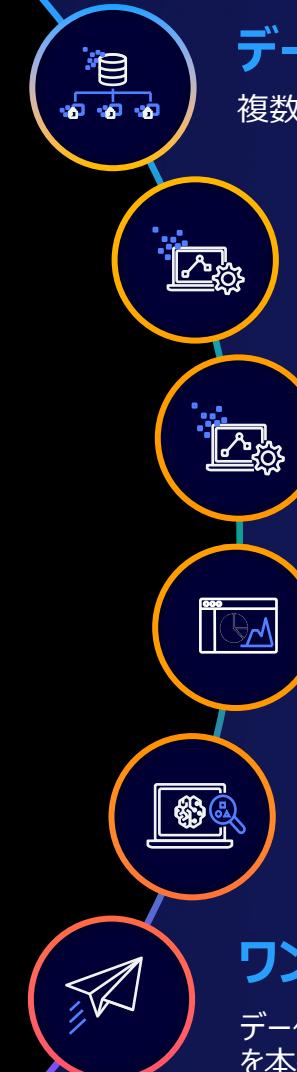
Amazon SageMaker

目的別に作成されたツール群。生産性を 10 倍に



SageMaker Data Wrangler

機械学習のためのデータを
準備する最速で
簡単な方法



データを素早く選択してクエリを実行

複数のソースからのデータをサポート

組み込みのデータ変換機能でデータを簡単に変換

組み込みのデータ変換を使用して、機械学習のために生データを特徴量に変換

データ変換をカスタマイズ

PySpark、SQL、Pandasで独自のカスタム変換を行う柔軟性

データを視覚的に理解

外れ値や極端な値を素早く検出 - コードを書かなくとも可能

MLモデルの精度を素早く推定

MLモデルの精度を妨げる可能性のあるデータ準備ワークフローの潜在的な問題を診断

ワンクリックでデータ準備のワークフローをデプロイ

データ準備のワークフローのすべてのステップを单一の視覚的なインターフェースで管理し、ワークフローを本番環境に迅速に運用可能

SageMaker Data Wrangler

ユースケース



データのクレンジングと探索

組み込みのデータ変換を使用して、
データのクレンジングと探索を高速化



データの可視化と理解

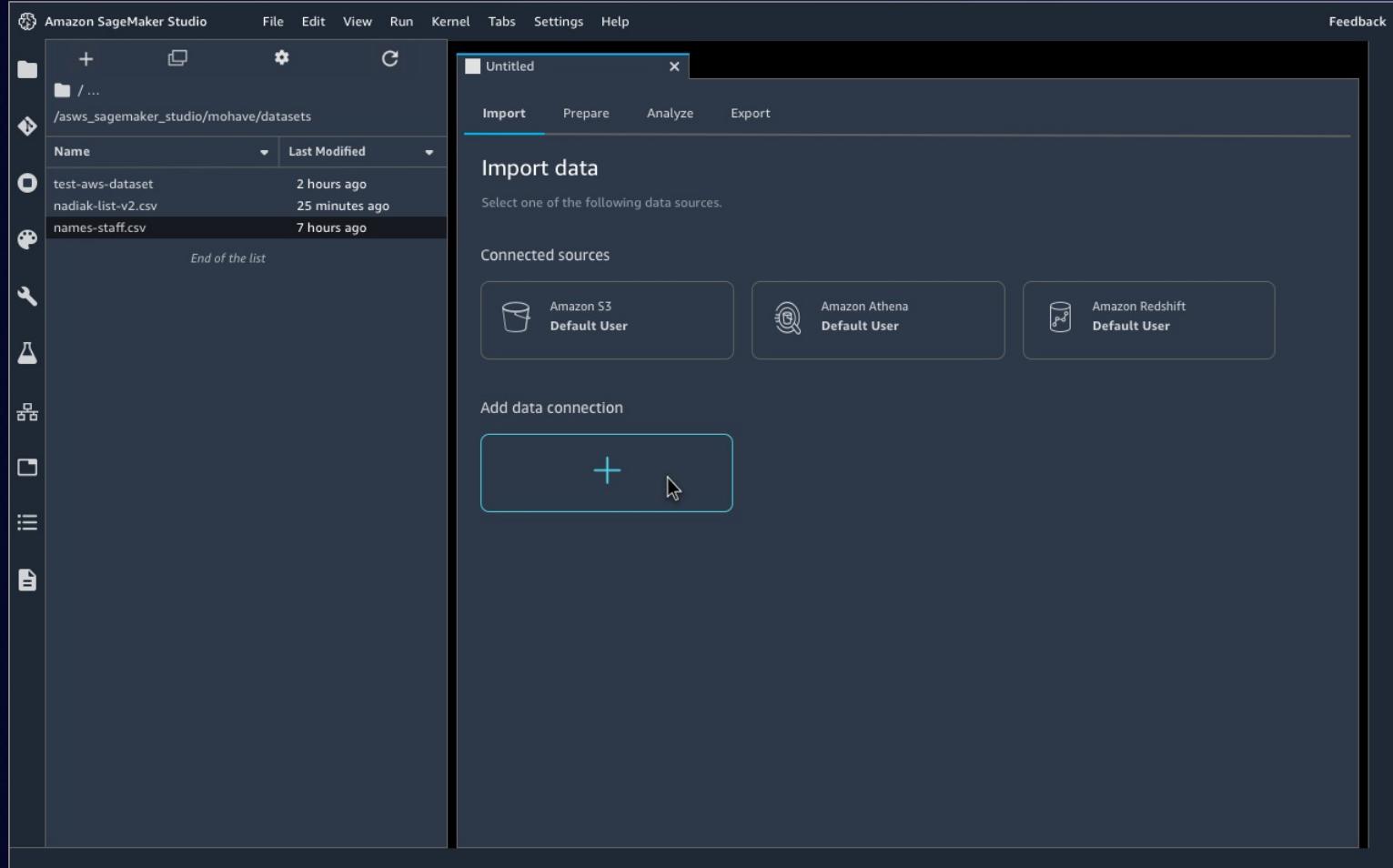
コードを書かなくても、
データセット内の外れ値や極端な値を
素早く検出



データのエンリッチ

組み込みのデータ変換ツールを使用して、
データを高精度なMLモデルを構築するために
使用できる形式に変換

多様なデータソースとの統合とクエリの実行



Amazon Athena、
Amazon Redshift、
Snowflake などの多様な
データソースとの統合

データをインポートする前に、
データソースに対してクエリを
記述することも可能

CSV ファイル、Parquet ファイル、
データベーステーブルなどの
ファイル形式のデータを直接
インポート

簡単にデータを変換

1行のコードも書かずに
データを変換

PySpark、SQL、
Pandasでカスタム変換も
作成可能

The screenshot shows the Amazon SageMaker Studio Data flow interface. On the left, there's a sidebar with various icons for file operations. The main area displays a CSV file named "S3-072320-campaignMovieSpecials.csv" with columns: String App, String Category, Decimal Rating, Integer Reviews, String Size, and String Installs. The data includes rows like "Photo Editor Coloring bookCamera &" and "Coloring book moana". To the right of the data table, there's a "TRANSFORMS" section with a "Custom transform" dropdown set to "PySpark". Below it is a code editor window containing the PySpark code: "1". At the bottom right of the code editor are "Clear", "Preview", and "Add" buttons.

Amazon SageMaker Studio

File Edit View Run Kernel Tabs Settings Help Feedback

Untitled

Import Prepare Analyze Export

Data flow / S3-072320-campaignMovieSpecials.csv

S3-072320-campaignMovieSpecials.csv

String App	String Category	Decimal Rating	Integer Reviews	String Size	String Installs
Photo Editor Coloring bookCamera &	ART_AND DESIGN	4.1	159	19M	10,000+
Coloring book moana	ART_AND DESIGN	3.9	967	14M	500,000+
U Launcher Lite	ART_AND DESIGN	4.7	87510	8.7M	5,000,000+
Themes, Hide Apps	ART_AND DESIGN	4.5	215644	25M	50,000,000+
Sketch - Draw & Paint	ART_AND DESIGN	4.3	967	2.8M	100,000+
Pixel Draw - Number Art Coloring Book	ART_AND DESIGN	4.4	167	5.6M	50,000+
Paper flowers instructions	ART_AND DESIGN	3.8	178	19M	50,000+
Smoke Effect Photo Maker - Smoke	ART_AND DESIGN	4.1	36815	29M	1,000,000+
Editor	ART_AND DESIGN	4.4	13791	33M	1,000,000+
Infinite Painter	ART_AND DESIGN	4.7	121	3.1M	10,000+
Garden Coloring Book	ART_AND DESIGN	4.4	13880	28M	1,000,000+
Kids Paint Free - Drawing Fun	ART_AND DESIGN	4.4	8788	12M	1,000,000+
Text on Photo - Fonteeve	ART_AND DESIGN	4.2	44829	20M	10,000,000+
Name Art Photo Editor - Focus n Filters	ART_AND DESIGN	4.6	4326	21M	100,000+
Tattoo Name On My Photo Editor	ART_AND DESIGN	4.4	1518	37M	100,000+
Mandala Coloring Book	ART_AND DESIGN	3.2	55	2.7M	5,000+
3D Color Pixel by Number - Sandbox	ART_AND DESIGN	4.7	3632	5.5M	500,000+
Art Coloring	ART_AND DESIGN	4.5	27	17M	10,000+
Learn To Draw Kawaii Characters	ART_AND DESIGN	4.3	194216	39M	5,000,000+
Photo Designer - Write your name with shapes	ART_AND DESIGN	4.6	224399	31M	10,000,000+
350 DIY Room Decor Ideas	ART_AND DESIGN	4.2	450	14M	100,000+
FlipaClip - Cartoon animation	ART_AND DESIGN	4.6	4.6	10M	100,000+
ibis Paint X			4		

Back to data flow

TRANSFORMS

Learn more about transforms.

Add Previous steps (5)

Suggested transformers

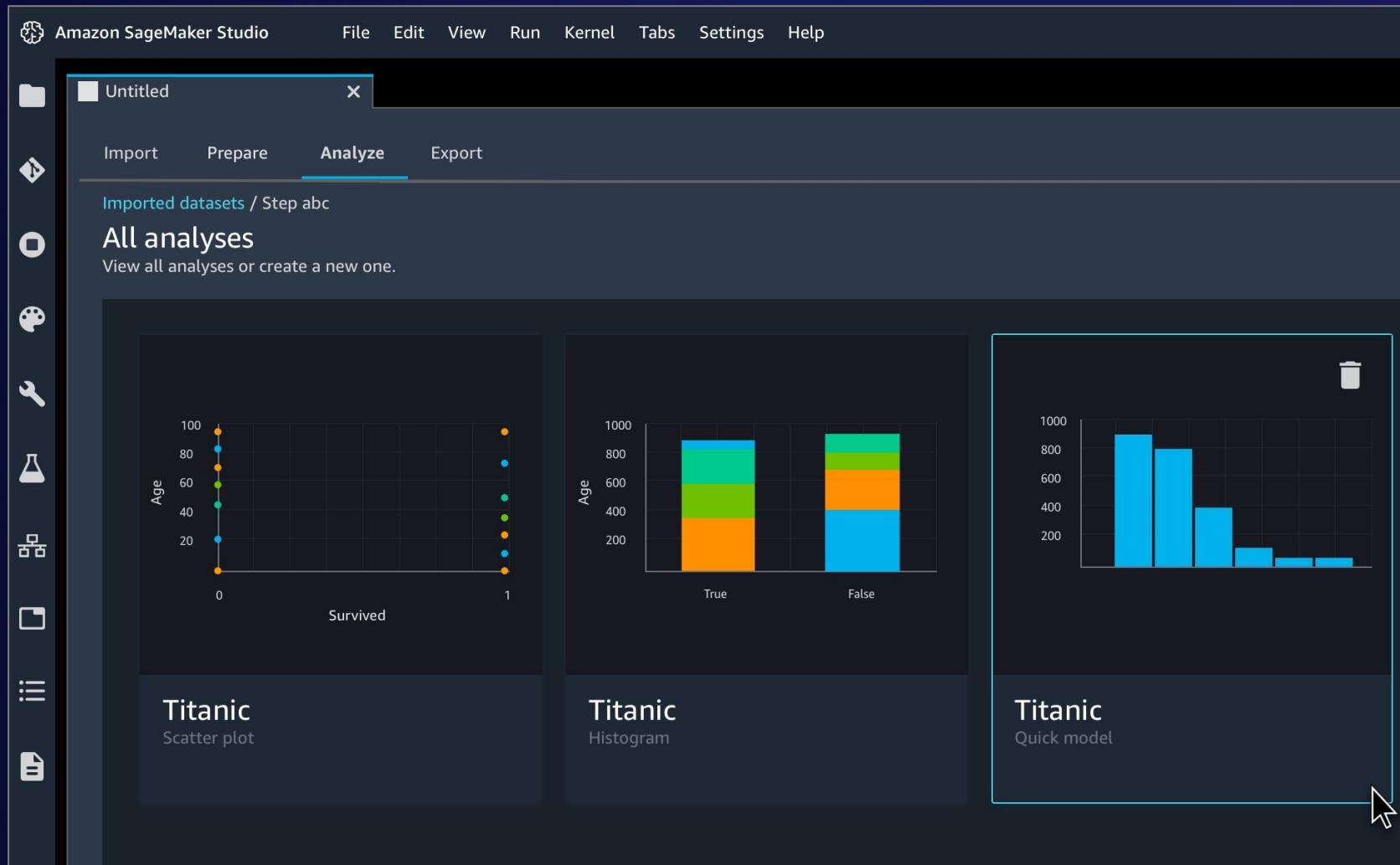
- > Find-Replace
- > Rename column
- > Replace rare
- > Impute missing categorical
- > Impute missing numeric
- > Tokenizer
- > TF-IDF text embedding
- > Ordinal encode
- > Onehot encode
- > Drop column
- > Duplicate column
- > Flatten vector
- > Split by delimiter
- > Type conversion
- > Custom transform

PySpark

1

Clear Preview Add

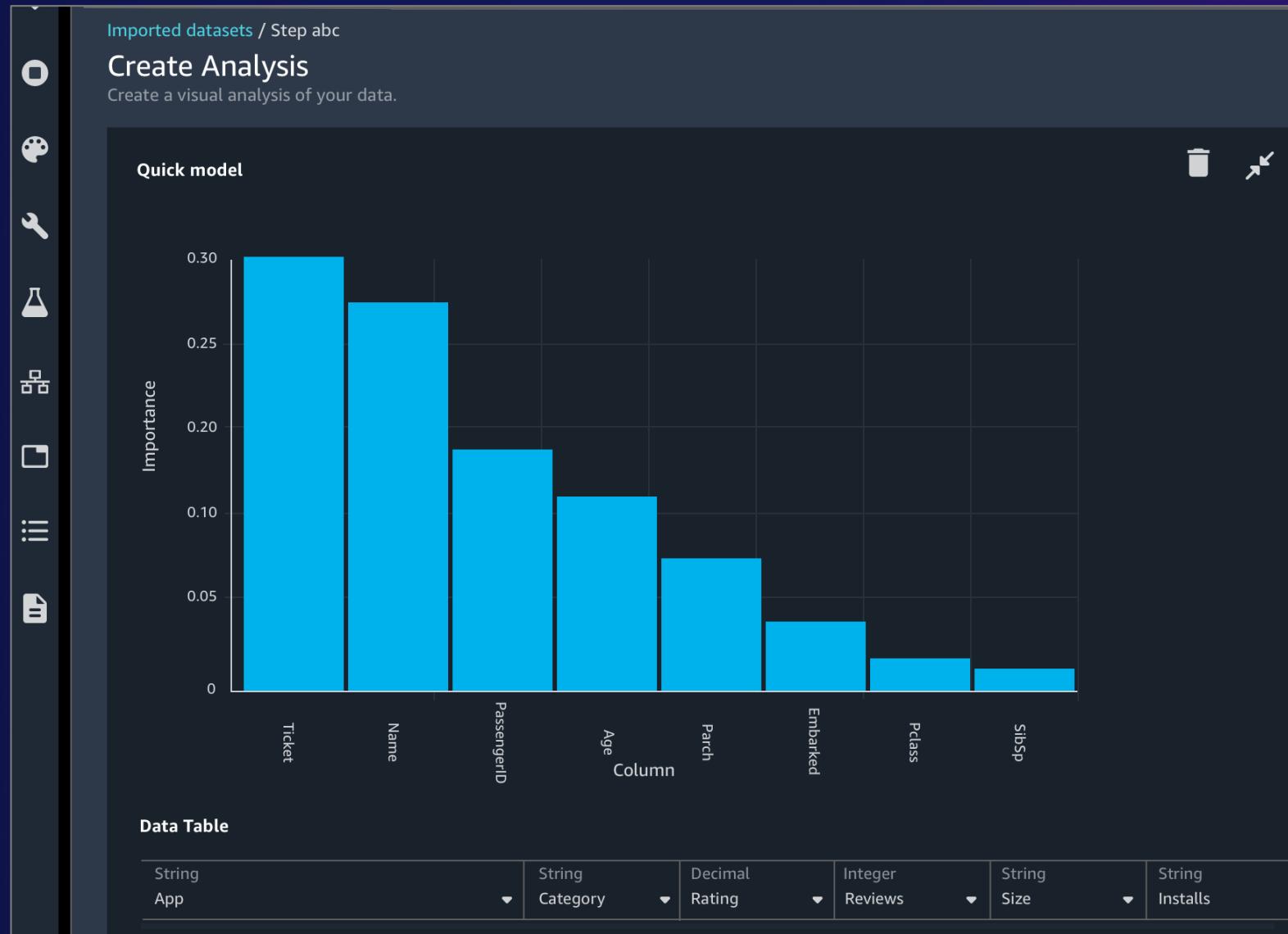
データを視覚的に理解する



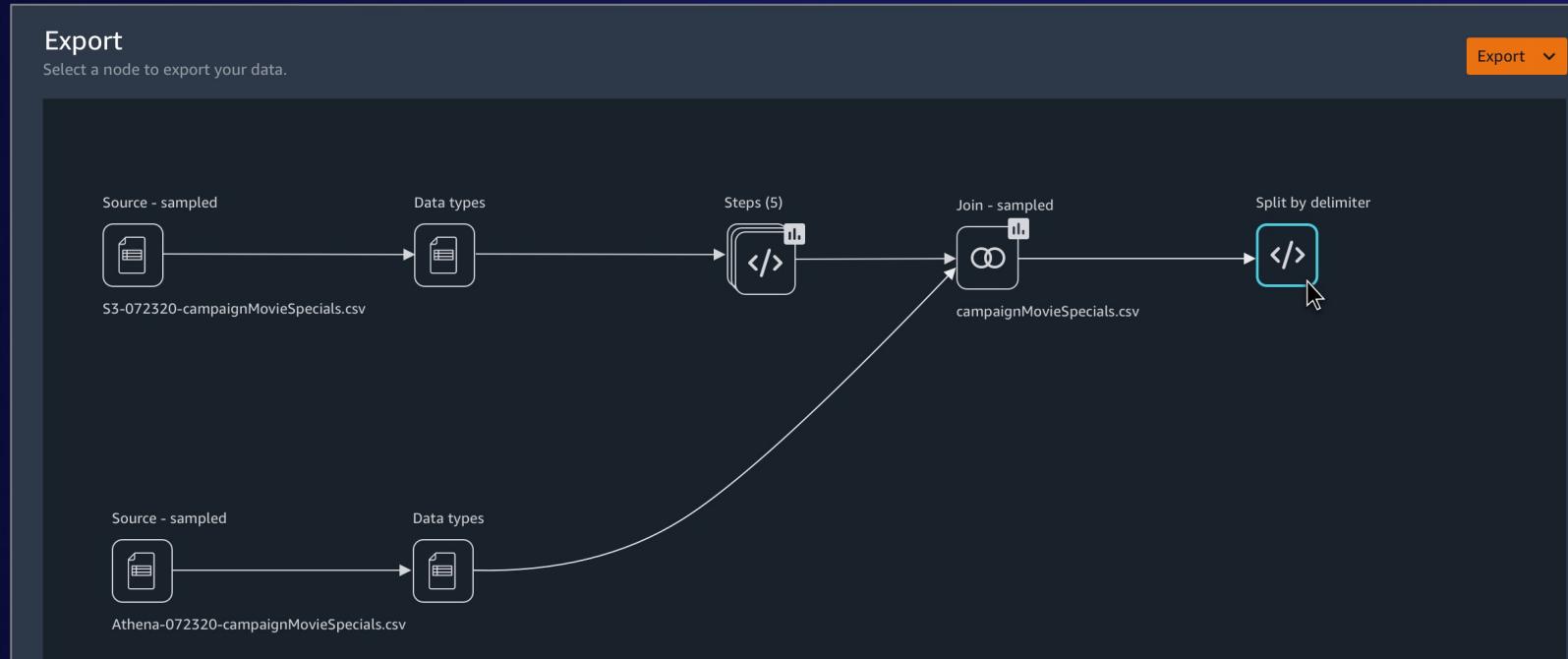
モデルの精度を素早く推定する

クリックでモデル構築
重要な特徴量を確認

リーケージやデータの
バイアスの分析が可能



データ準備のワークフローを本番環境にデプロイ



データ準備のワークフローをノートブックや
Python コードにエクスポート

SageMaker Pipelines とご自身のワークフローを統合し、
モデルのデプロイと管理を自動化

機械学習習得のための リソースの紹介

Amazon SageMaker を利用してみたいけど。。



Amazon SageMaker を利用すると、SageMaker Canvas 以上に柔軟な解析や学習、フローの構築ができることがわかった。
だけど、こんなに豊富な機能を使いこなせるか不安。



そもそも機械学習の知識もあまり詳しくないのに、適切な前処理やモデル選択ができるとは思えない。

学ぶアプローチ

1. 機械学習について学ぶ



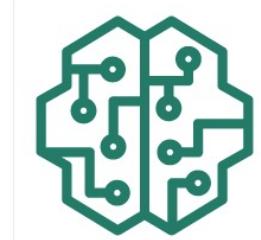
Amazon SageMaker Studio Lab



Amazon DeepRacer

2. ツールを使ってみる

SageMaker Immersion Day



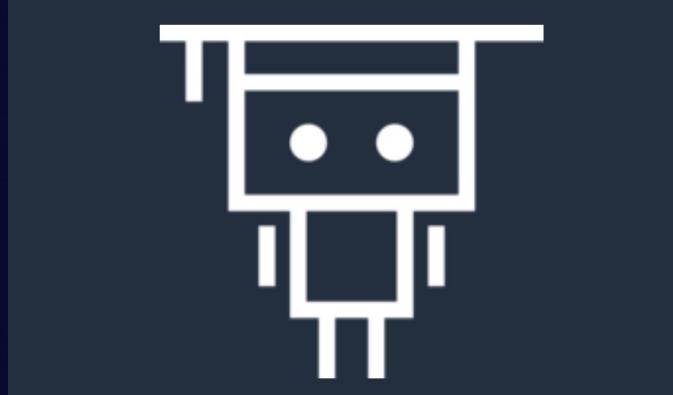
SageMaker Immersion Dayへようこそ

Amazon SageMaker Immersion Dayは、特徴量エンジニアリングからさまざまなビルトインアルゴリズムの理解、本番環境のようなシナリオでのMLモデルのトレーニング、チューニング、デプロイまで、MLユースケース構築のエンドツーエンドの理解をお客様やパートナーに提供します。独自のエコシステムの上に構築された高度なMLアーキテクチャをAmazon SageMakerプラットフォームに利用する機会を得られます。

機械学習について学ぶためのリソース

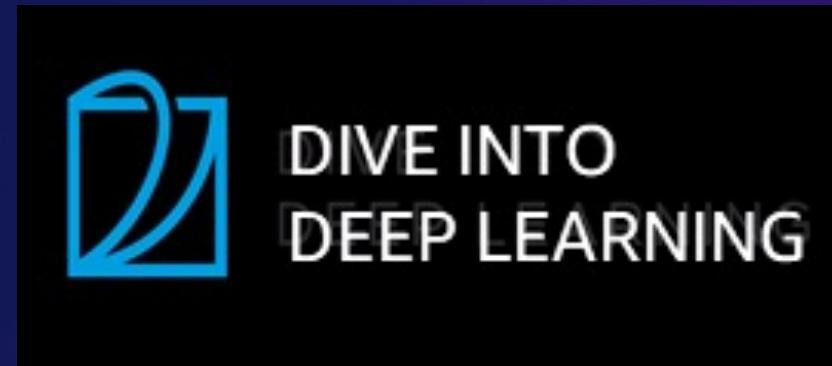
手を動かしながら機械学習を知る

Amazon の開発者を
トレーニングしていたものと
同じ機械学習コースを受講
(英語)



AWS Machine Learning University
(MLU)

コードや数式、議論を含む
インタラクティブな
ディープラーニングの本
(英語/一部日本語)



Dive into Deep Learning(D2L)

手を動かすための準備が必要

手を動かすために準備するもの



計算リソース



コードを動かすための
環境



これらを準備するのは何かと面倒
お金もなるべくかけたくない

Amazon SageMaker Studio Lab

手軽に機械学習を学び、
実験することのできる無料で設定不要なサービス

→ Jupyter Notebook 環境

JupyterLab ベース

→ コスト効率/簡単設定

無料で設定不要

→ CPU / GPU 計算

CPU & GPU 環境の提供

→ コード化までの時間

プロジェクトを保存可能な15 GB

→ バージョンコントロール管理

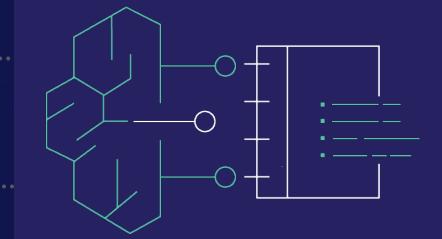
Git 統合

→ コミュニティサポート

GitHub 統合

→ シェルコマンドのフルサポート

ターミナルアクセス



Amazon SageMaker
Studio Lab

SageMaker Studio Lab でできることとは？



学ぶ

ハンズオンチュートリアル
や演習、ラボのための
ML環境に素早く無料で
アクセスする



教える

ML環境のトラブルシュート
の時間を減らし、MLコース
を準備し教えることにより
多くの時間を費やす



トレーニングする

参加者の負担やセットアップ
時間を減らして
ワークショップや
トレーニングを実施する

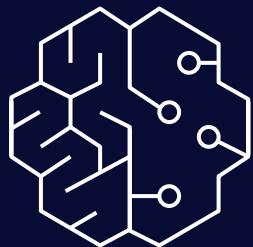


共同作業する

事前設定された環境や
ノートブックを使って
大規模なMLハッカソンや
チャレンジを実施する

AWS DeepRacer で学ぶ

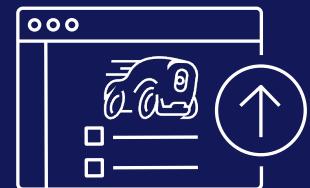
ハンズオン体験を通じて機械学習を学ぶ



強化学習の基礎を
知る



AWS DeepRacer の
構成要素について学ぶ



最初のモデルを
構築



構築したモデルを提出して
世界ランキングに挑戦

学ぶアプローチ

1. 機械学習について学ぶ



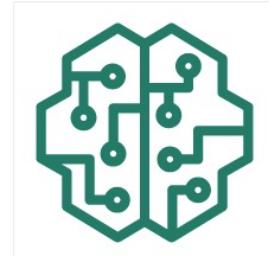
Amazon DeepRacer



Amazon SageMaker Studio Lab

2. ツールを使ってみる

SageMaker Immersion Day



SageMaker Immersion Dayへようこそ

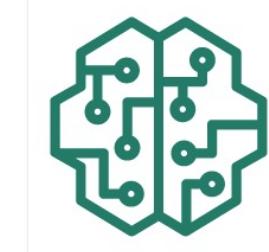
Amazon SageMaker Immersion Dayは、特徴量エンジニアリングからさまざまなビルトインアルゴリズムの理解、本番環境のようなシナリオでのMLモデルのトレーニング、チューニング、デプロイまで、MLユースケース構築のエンドツーエンドの理解をお客様やパートナーに提供します。独自のエコシステムの上に構築された高度なMLアーキテクチャをAmazon SageMakerで実現する方法について

SageMaker Immersion Day

SageMaker の使い方を網羅的に学べるワークショップ資料

- AWS の知識なしで始められる
- Amazon SageMaker Studio 上で動かせる機械学習における各ステップの Lab を用意
- 1 つの Lab は 1 時間程度で完了（現在 7 つの Lab を提供）
- 資料として公開されているので自分の好きなタイミングで取り組める

SageMaker Immersion Day



SageMaker Immersion Dayへようこそ

Amazon SageMaker Immersion Dayは、特微量エンジニアリングからさまざまなビルトインアルゴリズムの理解、本番環境のようなシナリオでのMLモデルのトレーニング、チューニング、デプロイまで、MLユースケース構築のエンドツーエンドの理解をお客様やパートナーに提供します。独自のモデルを持ち込んで、オンラインでのML開発をA to Zで学ぶことができます。

<https://sagemaker-immersionday.workshop.aws/ja-JP>

AWS SageMaker Examples on GitHub

オープンソースなサンプルとラボ、チュートリアル

- ・日本語版の GitHub レポジトリも存在
- ・SageMaker の使い方について学べるリソースを多数公開



Amazon SageMaker

Amazon SageMaker Examples

Example Jupyter notebooks that demonstrate how to build, train, and deploy machine learning models using Amazon SageMaker.

<https://github.com/aws/amazon-sagemaker-examples>



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

AWS Example の例

商品レビューに対する感情分析を行うサンプル

BlazingText を利用した感情分析

概要

このノートブックでは、Amazon の商品レビューに対する感情分析、つまり、そのレビューが Positive (Rating が 5 or 4) か、Negative (Rating が 1 or 2) なのかを判定します。これは、文書を Positive か Negative に分類する2クラスの分類問題なので、**BlazingText**による教師あり学習を適用することができます。

データの準備

Amazon の商品レビューデータセットは [Registry of Open Data on AWS](#) で公開されており、以下からダウンロード可能です。このノートブックでは、日本語のデータセットをダウンロードします。

- データセットの概要
<https://registry.opendata.aws/amazon-reviews/>
- 日本語のデータセット(readme.htmlからたどることができます)
https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_multilingual_JP_v1_00.tsv.gz

https://github.com/aws-samples/amazon-sagemaker-examples-jp/blob/master/nlp_amazon_review/BlazingText/blazingtext.ipynb



各種学習リソースへのリンク

AWS Machine Learning University

<https://aws.amazon.com/machine-learning/mlu/>

Dive into Deep Learning

<https://ja.d2l.ai/index.html>

AWS SageMaker Examples

<https://github.com/aws/amazon-sagemaker-examples>

AWS ML-JP (日本語)

<https://github.com/aws-samples/aws-ml-jp>

SageMaker Immersion Day

<https://sagemaker-immersionday.workshop.aws/ja-JP>



まとめ



まとめ

各々のレベル感に合わせて機械学習活用を始めてみよう



みなさんも、機械学習の世界に飛び
込んでみませんか？

まずは SageMaker Canvas から

Thank you!



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