

T2-1

ノーコードではじめる機械学習 Amazon SageMaker Canvas の 使いどころ

小杉 知己

アマゾン ウェブ サービス ジャパン合同会社
プロフェッショナルサービス本部 クラウドアプリケーションアーキテクト

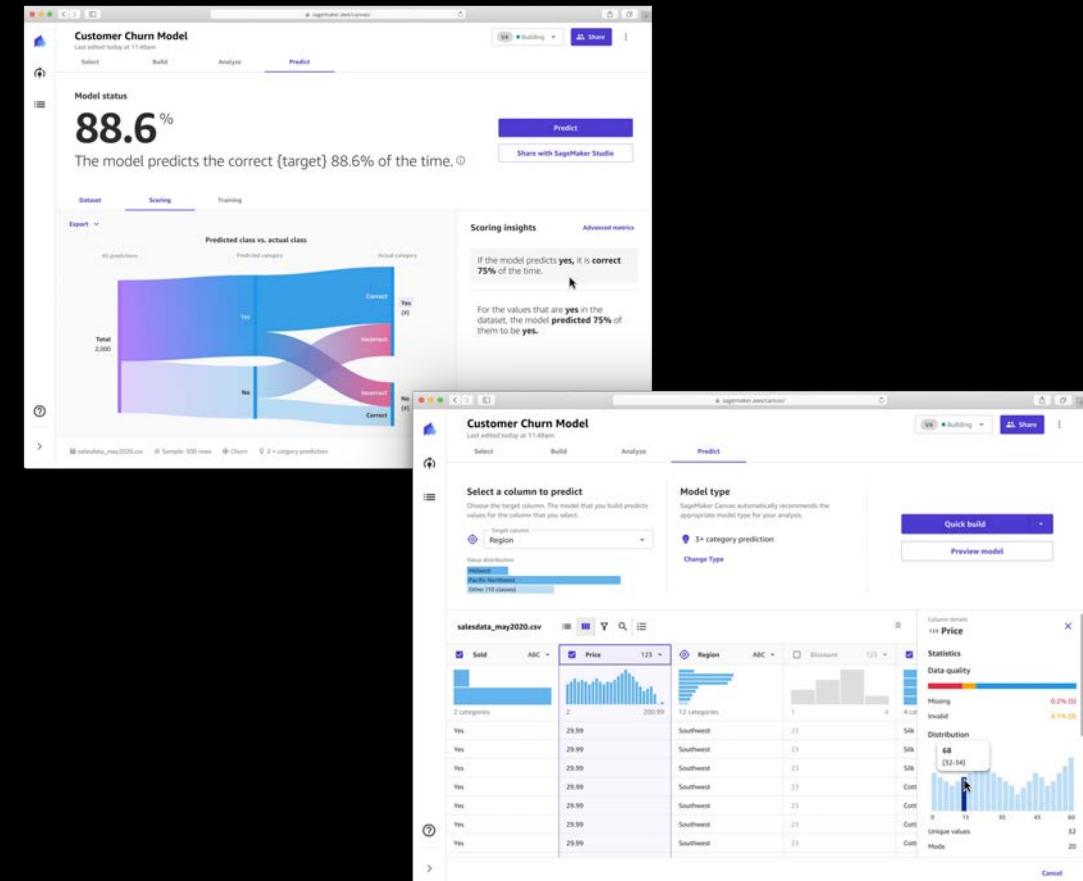


本セッションの概要

- Amazon SageMaker Canvasの使い方・使いどころ
- ビジネス課題へ機械学習を適用するはじめの一歩



GUI上でコーディングなしに実施可能



本セッションの想定ユーザー像

- ・ まだ機械学習の知識・スキルはないけど、業務に機械学習を活用したい！
 - ビジネスリーダー
 - ドメインエキスパート
 - ビジネスアナリスト
- ・ データ分析・機械学習のワークフローを効率化したい！
 - データサイエンティスト
 - データエンジニア
 - 機械学習エンジニア



自己紹介

小杉 知己 (こすぎともき)

略歴

- ・ エンジニア・PMとしてR&Dプロダクトの開発
- ・ データ活用・機械学習プロジェクトにおけるアプリケーション開発
- ・ データサイエンティストではございません



得意なこと/好きなこと

- ・ 機械学習やデータ活用を絡めたアプリケーションやシステム開発
- ・ Container, IaC, CI/CD

好きなAWSサービス

ECS, SageMaker



Agenda

1. データ活用・機械学習の需要と課題
2. Amazon SageMaker Canvasの概要・使いどころ
3. Amazon SageMaker Canvasの使い方（デモ）
4. 事例紹介
5. まとめ



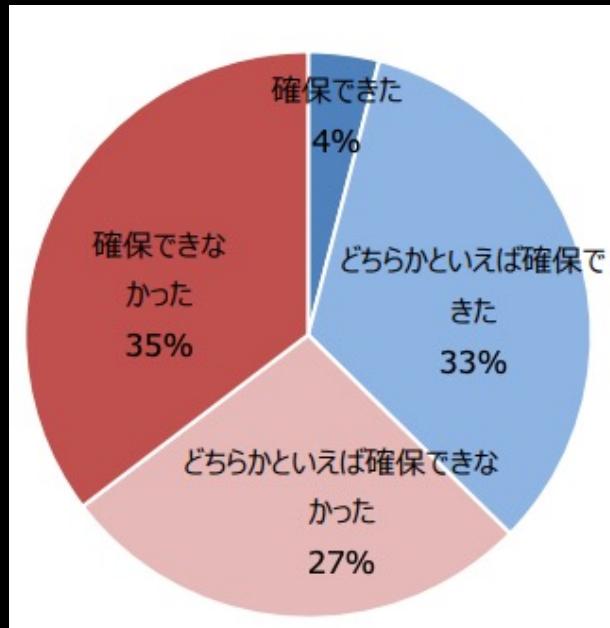
データ活用・機械学習の 需要と課題



データサイエンティスト人材の需要

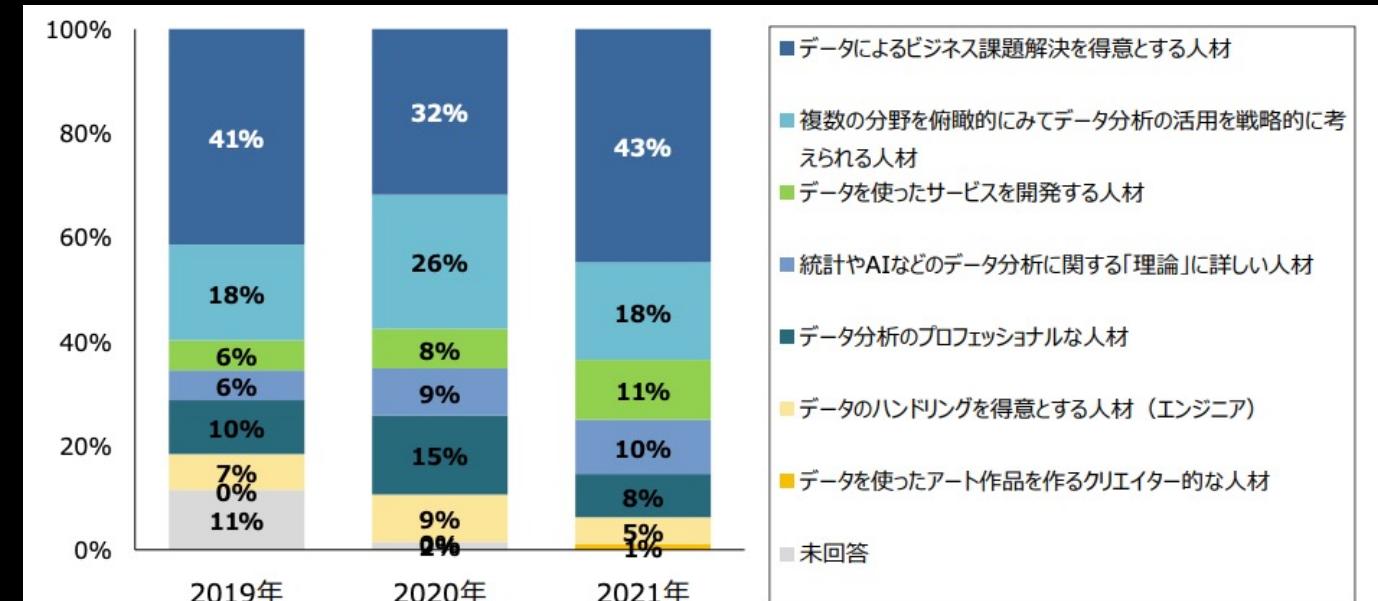
引用：データサイエンティストの採用に関するアンケート（2021年回答）／データサイエンティスト協会

Q. この1年間で、貴社が目標としていた人数のデータサイエンティストを確保できましたか。



「データサイエンティストを目標通り確保できなかつた」企業が62%

Q. 今後3年間で、貴社が採用・育成したいデータサイエンティストの人材像をお答えください。



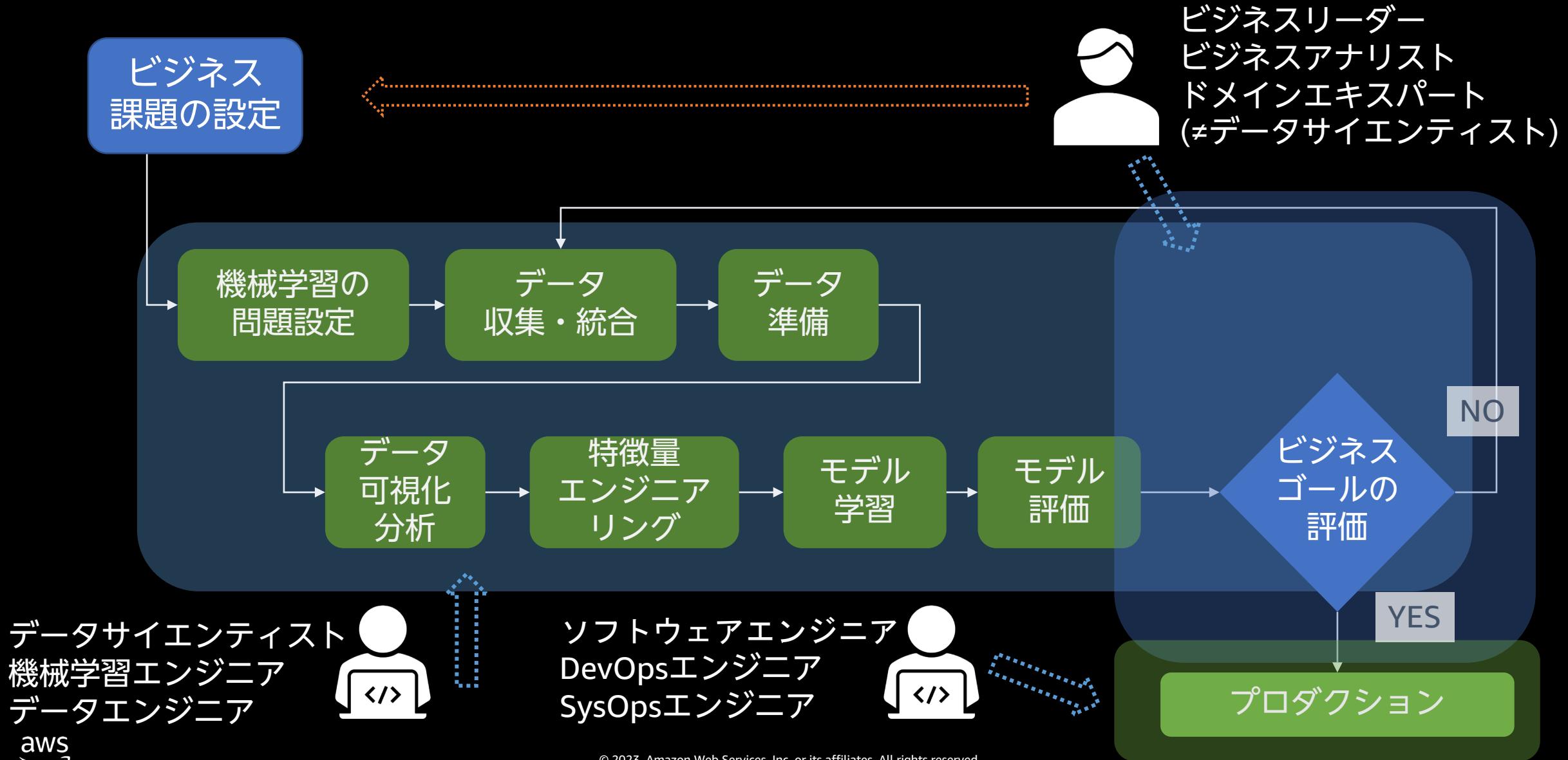
「データによるビジネス課題解決得意とする人材」の需要が43%でトップ

https://www.datascientist.or.jp/common/docs/corporate_research2021.pdf

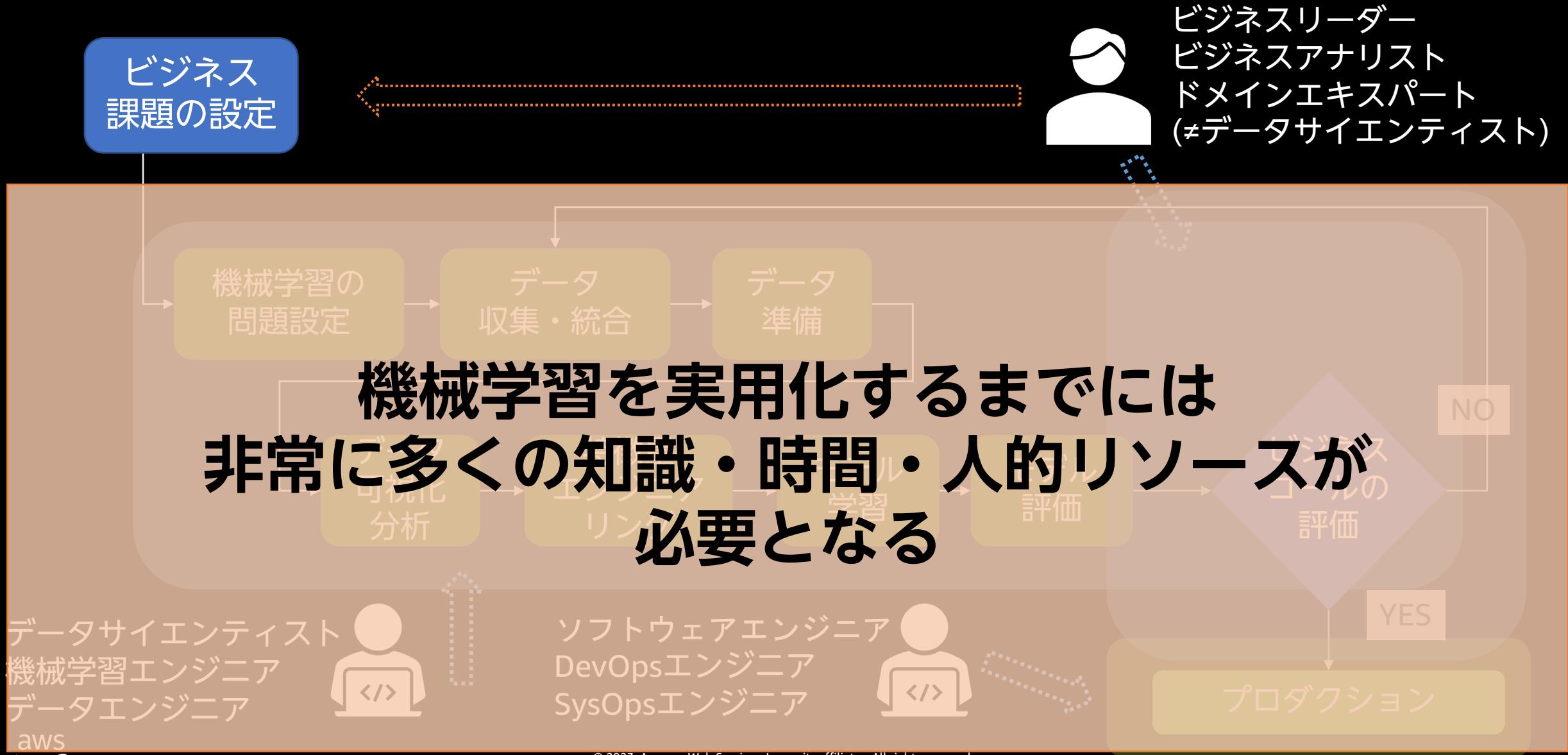
機械学習実用化までの典型的なワークフロー



機械学習実用化までの典型的なワークフロー



機械学習実用化までの典型的なワークフロー



機械学習へのチャレンジにおける課題

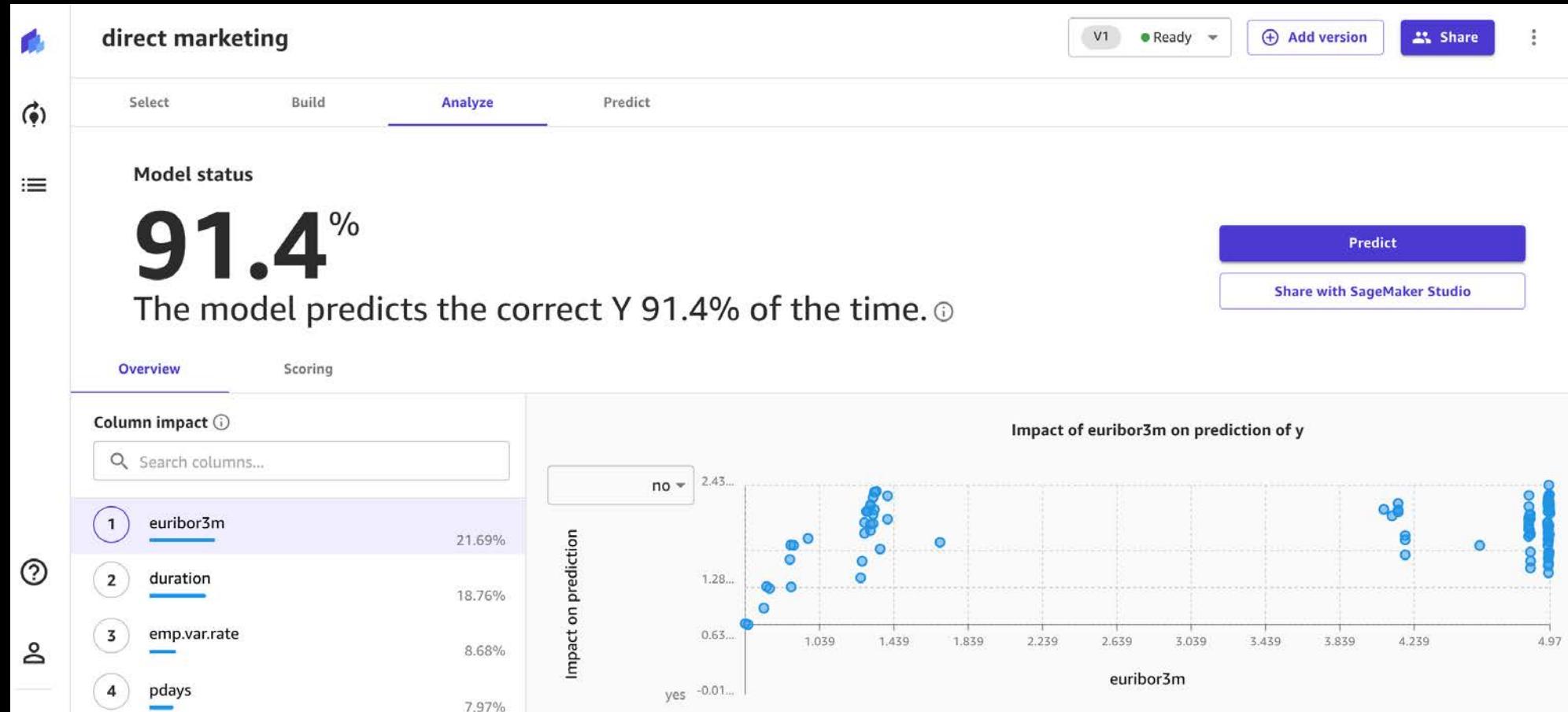
- 1 機械学習のニーズは高まる一方で、データサイエンスチームの供給が足りていない
→ 採用・育成による即時的な解決はなかなか難しい...
- 2 機械学習の知識・スキルを短時間で習得することは困難
→ ノーコード機械学習サービスによって
ビジネスアナリストに作業を移譲
- 3 機械学習の実用化には時間がかかり多くのコーディング・作業が必要
→ 既存のデータサイエンスチームがノーコード/ローコードの
サービスを活用することで効率化

Amazon SageMaker Canvasの概要・使いどころ



Amazon SageMaker Canvas 一言でいうなら

機械学習モデルの作成や予測をノーコード・GUIで実行できるサービス



これでAI/MLできるやんサイコー





機械学習をはじめるその前に

機械学習はあくまで

ビジネス課題を解くための1つの手段です。

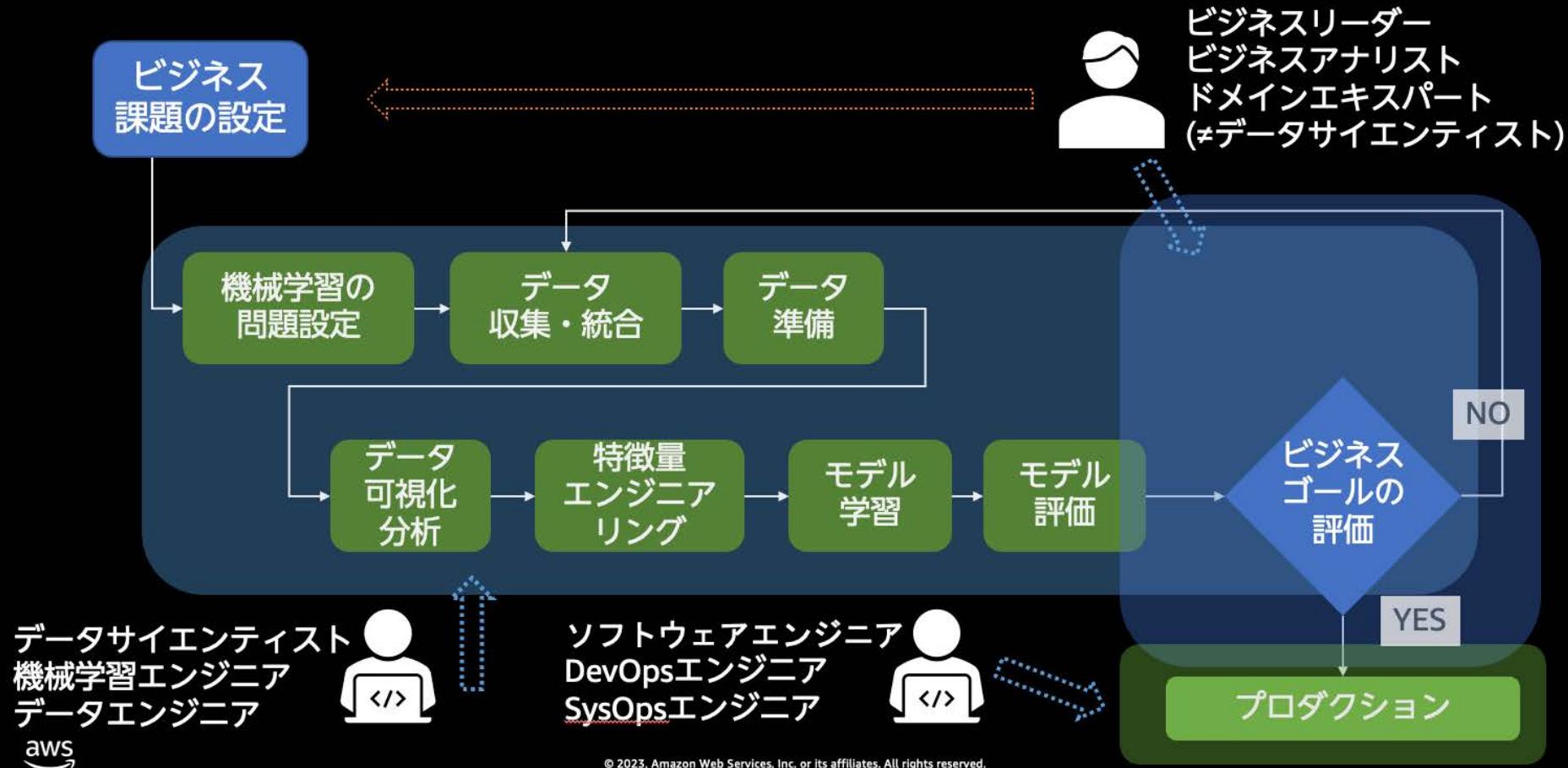
「手段の目的化」に陥っていないかは

常に注意しましょう。

Amazon SageMaker Canvasの使いどころ

(再掲)

機械学習実用化までの典型的なワークフロー



Amazon SageMaker Canvasの使いどころ

(再掲)

機械学習実用化までの典型的なワークフロー



Amazon SageMaker Canvasの機能概要



ローカルPCや
Amazon S3,
Amazon Redshift,
Snowflakeといった
さまざまなソースか
らデータを統合可能

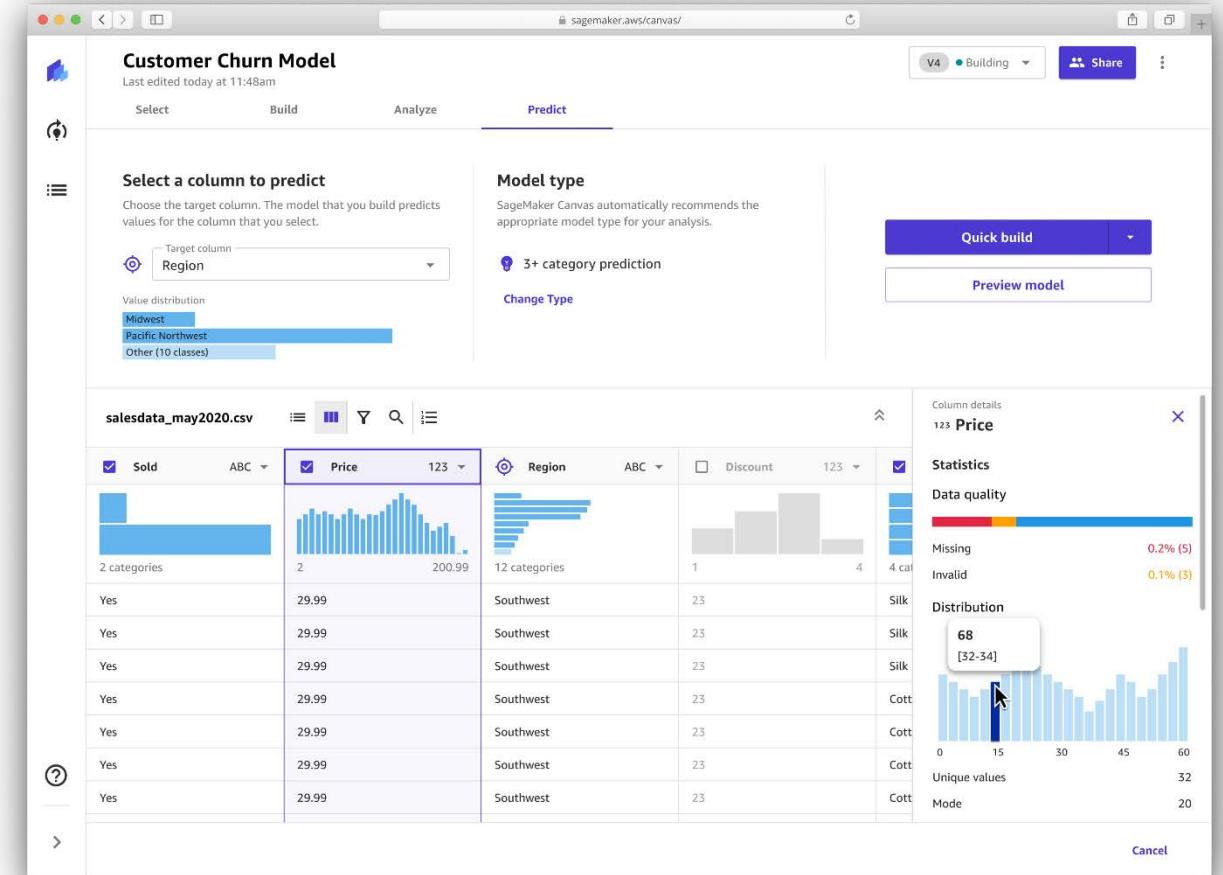
…その他データソースも対応予定

The screenshot shows the AWS SageMaker Canvas interface titled "Import Data". At the top, there are connection options: "Upload", "S3", "Snowflake Crystal 1", and "Redshift Crystal 1". Below that is a "Connection name" dropdown with "database1", "database2", "database3", "database4", "schema1", and "schema2" listed. A tree view shows "table1" under "schema2". On the right, a diagram illustrates a data flow from "table1.csv" to a central node, which then connects to "table2.csv". The "Import preview" section displays a table with 100 rows of data, including columns like "Sold", "ABC", "Price", "Region", "Discount", "Fabric", "Age", and "123". A "Show dropped columns" toggle is on. At the bottom, buttons for "Close" and "Import data" are visible.

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	



データを視覚的に 素早く理解



ワンクリックで自動的に機械学習モデルを構築



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

Customer Churn Model
Last edited today at 11:48am

Select Build Analyze Predict

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

Value distribution:

- Midwest
- Pacific Northwest
- Other (10 classes)

Model type
SageMaker Canvas automatically recommends the appropriate model type for your analysis.

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

salesdata_may2020.csv

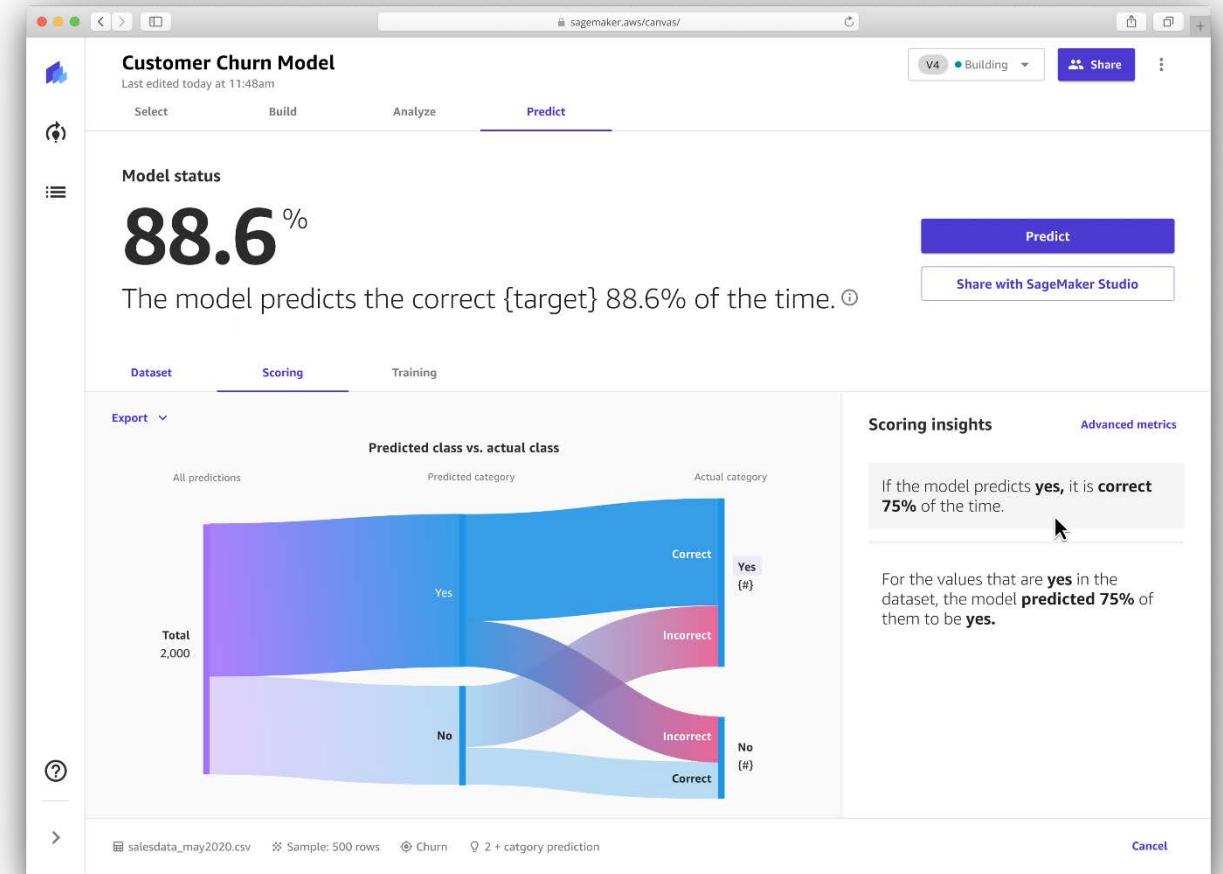
Sold	ABC	P
2 categories	2	2
Yes	29.99	
Yes	29.99	
Yes	29.99	
Yes	29.99	Southwest

Column details: Price

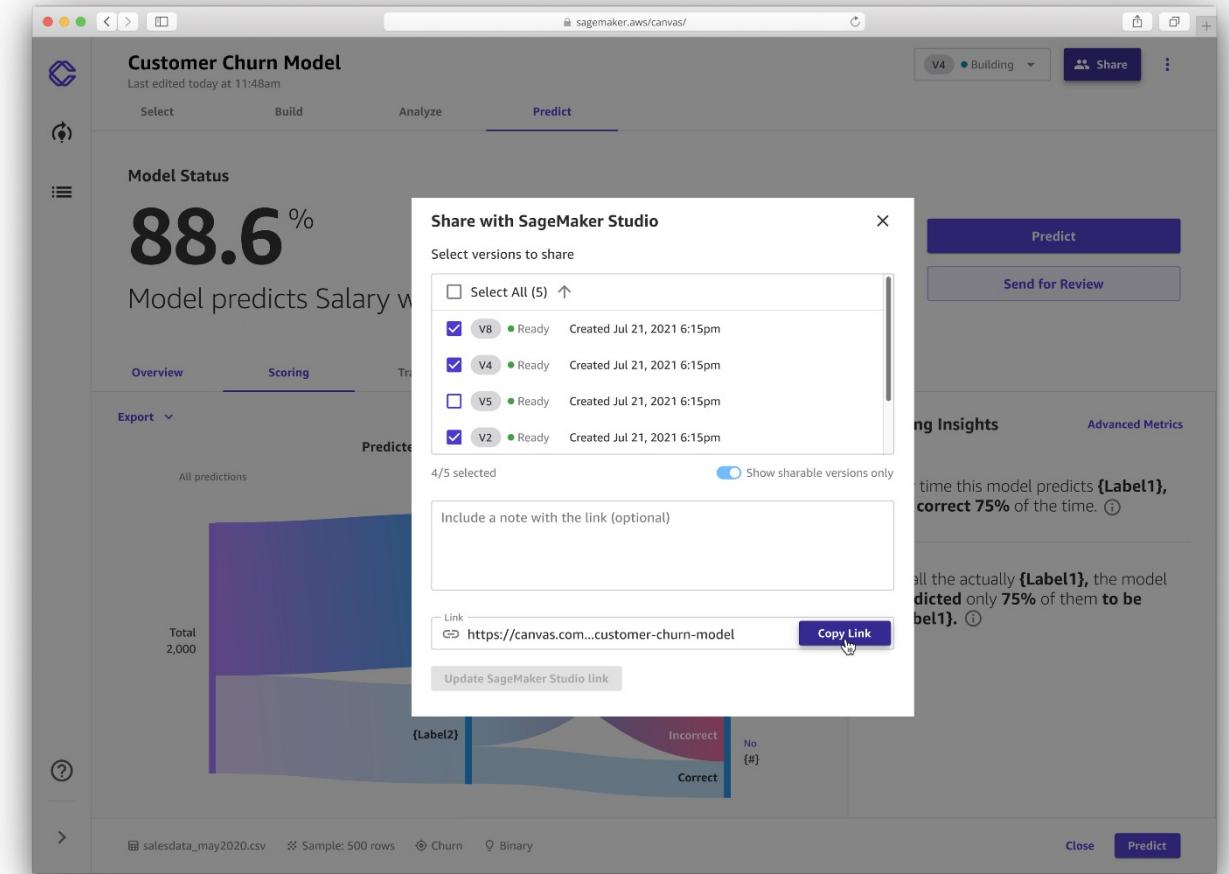
Statistics: Data quality

Distribution: Unique values, Mode

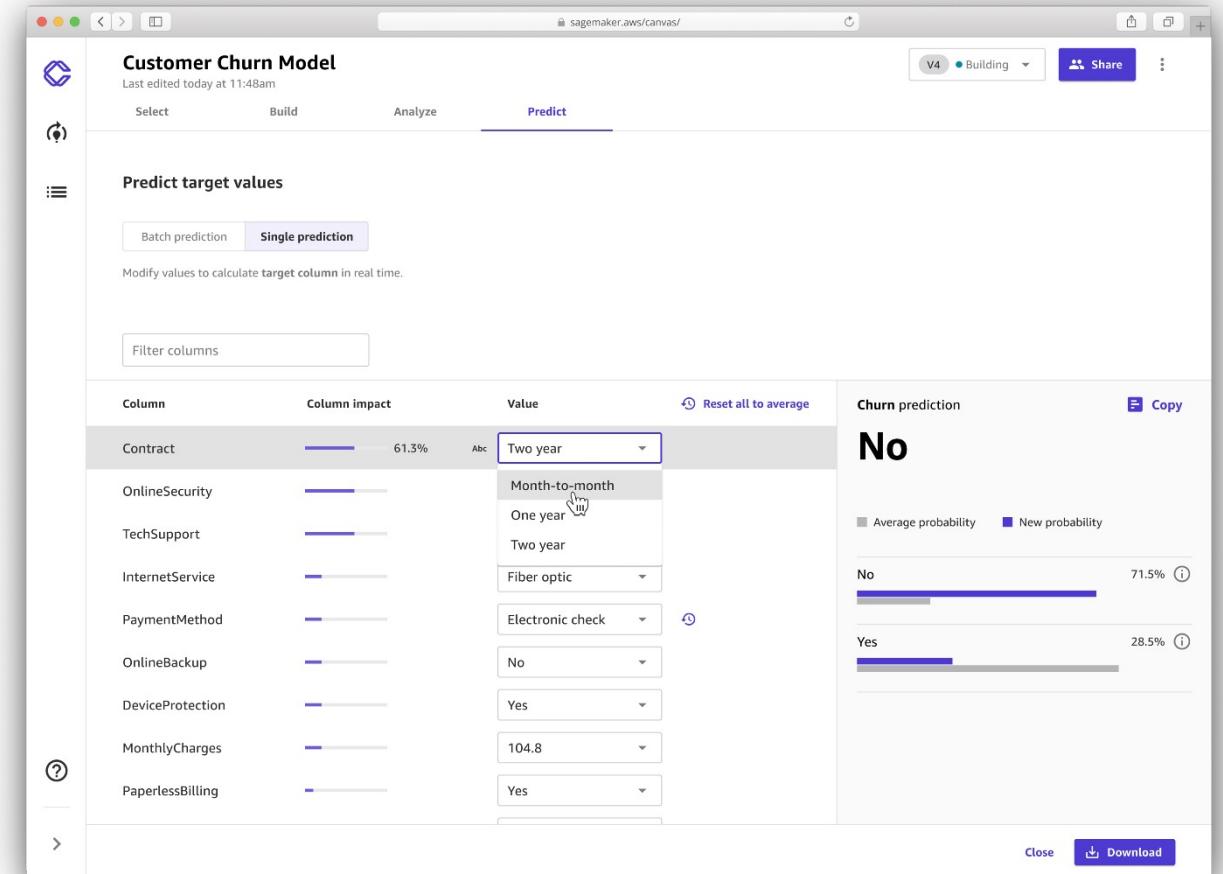
詳しい評価指標で予測の説明性を理解



モデルをデータサイエンティストと共有できる



バッチ予測や what-ifシナリオの 分析



Amazon SageMaker Canvasの料金体系 概要



セッション料金

SageMaker Canvasにログインしている間に課金

1.9USD / 1時間

※明示的にログアウトするまで課金



トレーニング料金

学習データのセル数に応じて課金

最初の1000万セル：30USD / 100万セル

<https://aws.amazon.com/jp/sagemaker/canvas/pricing/>

Amazon SageMaker Canvas ユースケース

異なる事業または、業種にまたがる膨大な数のユースケース



セールス とマーケティング

1. 販売コンバージョン
2. 売上予測
3. 顧客の離反傾向
4. 顧客の Lifetime Value (LTV) 予測
5. Marketing Mix Modeling (MMM) 分析



ファイナンス と会計

1. 信用リスクスコアリング
2. 支払い遅延予測
3. 不正検知
4. ポートフォリオ最適化
5. 買掛金処理の自動化



小売

1. 需要予測
2. 在庫管理とスケジューリング
3. 配達時間予測

その他多数…

Amazon SageMaker Canvasの使い方（デモ）



デモのシチュエーション

- サービスを解約しそうなユーザーに対して
キャンペーンなどの施策を実施する

解約しそうなユーザーを
予測できるか試してみる

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1 State	Account Leng	Area Code	Phone	Int'l Plan	VMail Plan	VMail Messa	Day Mins	Day Calls	Day Charge	Eve Mins	Eve Calls	Eve Charge	Night Mins	Night Calls	Night Charge	Intl Mins	Intl Calls	Intl Charge	CustServ C	Churn?	
2 PA	163	806	403-2562	no	yes		300	8.16220402	3	7.5791737	3.93303499	4	6.50863888	4.06575946	100	5.11162391	4.92816021	6	5.67320304	3	True.
3 SC	15	836	158-8416	yes	no		0	10.0189927	4	4.22628882	2.32500455	0	9.97259242	7.14103987	200	6.43618762	3.22174762	6	2.55974916	8	False.
4 MO	131	777	896-6253	no	yes		300	4.70848992	3	4.76815988	4.53746628	3	4.5667145	5.36323468	100	5.14245087	7.13902305	2	6.25415659	4	False.
5 WY	75	878	817-5729	yes	yes		700	1.26873377	3	2.56764159	2.52874788	5	2.33362363	3.77358558	450	3.81441342	2.24577933	6	1.08069204	6	False.
6 WY	146	878	450-4942	yes	no		0	2.69617724	3	5.90891587	6.01533707	3	3.67040775	3.75167337	250	2.796812	6.90554529	4	7.13434291	6	True.
7 VA	83	866	454-9110	no	no		0	3.63477597	7	4.80489189	6.05194434	5	5.27843696	2.93787999	300	4.81795844	4.94881609	4	5.13532274	5	False.
8 IN	140	737	331-5751	yes	no		0	3.22942026	4	3.16508206	2.44015281	8	0.26454309	2.3522739	300	3.86917635	5.39343925	4	1.78476482	4	False.
9 LA	54	766	871-3612	no	no		0	0.56792018	6	1.95009839	4.50702683	0	4.47308576	0.68878515	400	6.13213669	5.01274724	5	0.41742094	8	False.
10 MO	195	777	249-5723	yes	no		0	5.81111576	6	4.33106502	8.10412615	2	4.47503429	4.20835201	250	5.97457535	4.75015291	7	3.32031079	7	True.
11 MI	104	657	757-7502				0	0.71111000	7	5.12065602	0.50000100	6	0.22112101	0.21017507	200	1.20005573	2.12005500	5	5.67320304	1	False.

過去の顧客ユーザーデータから
パターンを学習し、新規顧客ユーザー
に対する予測をできるようにする

「解約したかどうか」の情報

U	Churn?
	True.
	False.
	False.
	False.
	True.
	False.
	False.
	True.
	False.

データ概要

列	備考
State	顧客が居住する米国の州。2文字の略語で示されます。
Area Code	顧客の電話番号の3桁の市外局番
Phone	残りの7桁の電話番号
Day Mins, Day Calls, Day Charge	日中の通話時間、通話件数、請求費用
Eve Mins, Eve Calls, Eve Charge	夕方の通話時間、通話件数、請求費用
Night Mins, Night Calls, Night Charge	夜間の通話時間、通話件数、請求費用
Churn?	顧客が解約したか (true/false)

...その他、全部で21列のCSVデータ

データのリンク：<https://sagemaker-sample-files.s3.amazonaws.com/datasets/tabular/synthetic/churn.csv>



Amazon SageMaker Canvas

初回セットアップ

The screenshot shows the AWS Management Console with the search bar set to "Search" and the keyboard shortcut "[Option+S]" highlighted. The navigation bar includes the AWS logo, Services icon, and a search bar. The main content area shows the "Amazon SageMaker > Setup SageMaker Domain" path. The title "Setup SageMaker Domain" is displayed, followed by a sub-instruction: "Use SageMaker Domain as the central store to manage the configuration of SageMaker for your organization." A prominent blue box contains the "Quick setup (1 min)" section, which describes letting Amazon SageMaker configure the account and set up permissions for the SageMaker Domain. It lists several checked options: "Public internet access, and standard encryption", "SageMaker Studio Integration", "Sharable SageMaker Studio Notebooks", "SageMaker Canvas", and "IAM Authentication". Below this box, a note states: "Perfect for single user domains and first time users looking to get started with SageMaker." At the bottom of the screen, a dark bar displays the text "Domain name".

ドキュメント：https://docs.aws.amazon.com/ja_jp/sagemaker/latest/dg/canvas-getting-started.html#canvas-prerequisites

参考ワークショップ：<https://catalog.workshops.aws/canvas-immersion-day/ja-JP/0-prerequisites/self-paced>

[デモ] データセットのインポート



[デモ] データセットのインポート

The screenshot shows the Amazon SageMaker Canvas interface. On the left is a sidebar with icons for Amazon SageMaker (blue square), Canvas (purple square), Models (green square), and Datasets (blue square). The Datasets icon is highlighted with a light purple background. The main area is titled "Datasets" and contains a table with two rows of data. The table columns are: Name, Source, Columns, Rows, Cells, Created, and Status. The data rows are:

Name	Source	Columns	Rows	Cells	Created	Status
ChurnTest.csv	Local	21	500	10,500	11/21/2022 12:04 AM	Ready
ChurnTrain.csv	Local	21	4,500	94,500	11/21/2022 12:04 AM	Ready

A search bar at the top right contains the placeholder text "Dataset".

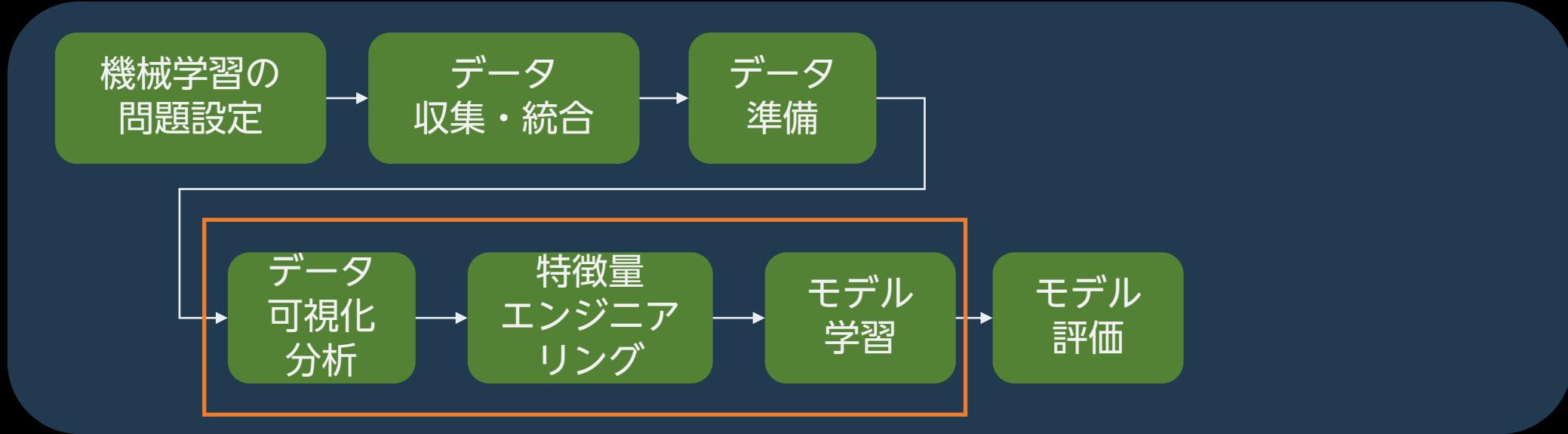
Getting started with using Amazon SageMaker Canvas

<https://docs.aws.amazon.com/sagemaker/latest/dg/canvas-getting-started.html>

[デモ] モデルの作成



[デモ] モデルの作成



この作業は通常多くのコーディング作業が必要になるが、SageMaker Canvasではノーコードで簡単に実行することが可能

使ってはいけないデータの例



ユニークな識別子

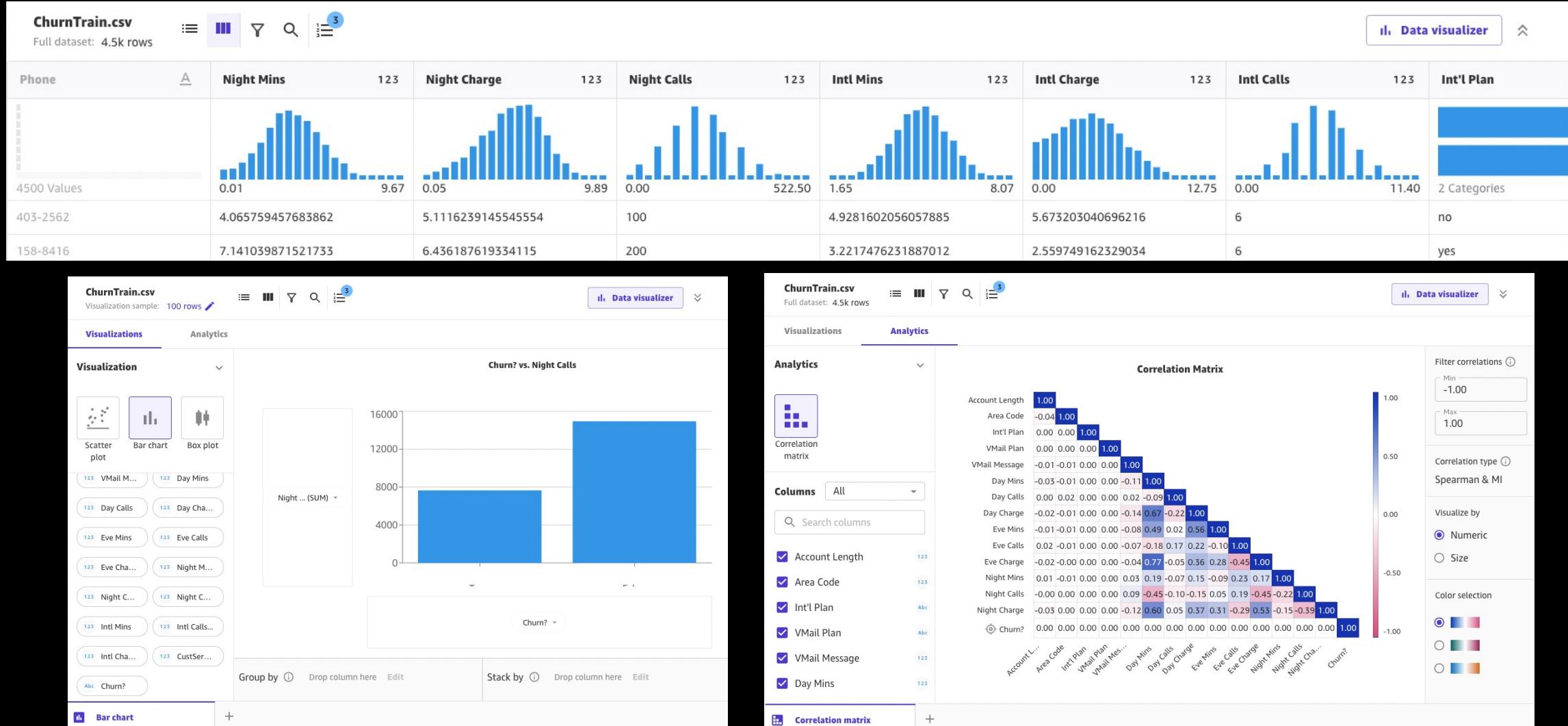
Phoneのような各ユーザを一意に特定できるようなデータを学習しても未知のデータは予測できないモデルになってしまう



実際の業務で手に入らないデータ

予測実行時よりも未来のデータなど

データの可視化



<https://docs.aws.amazon.com/sagemaker/latest/dg/canvas-explore-data.html>

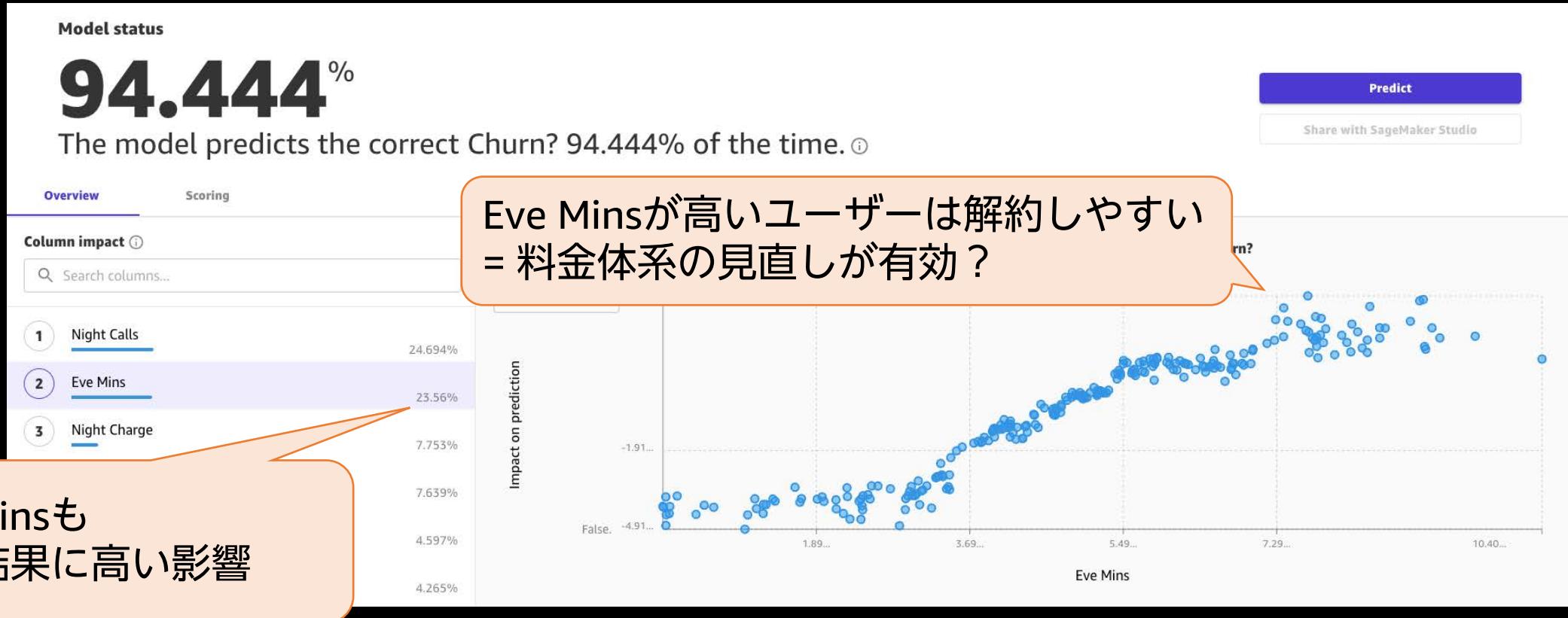


[デモ] モデルの確認・評価

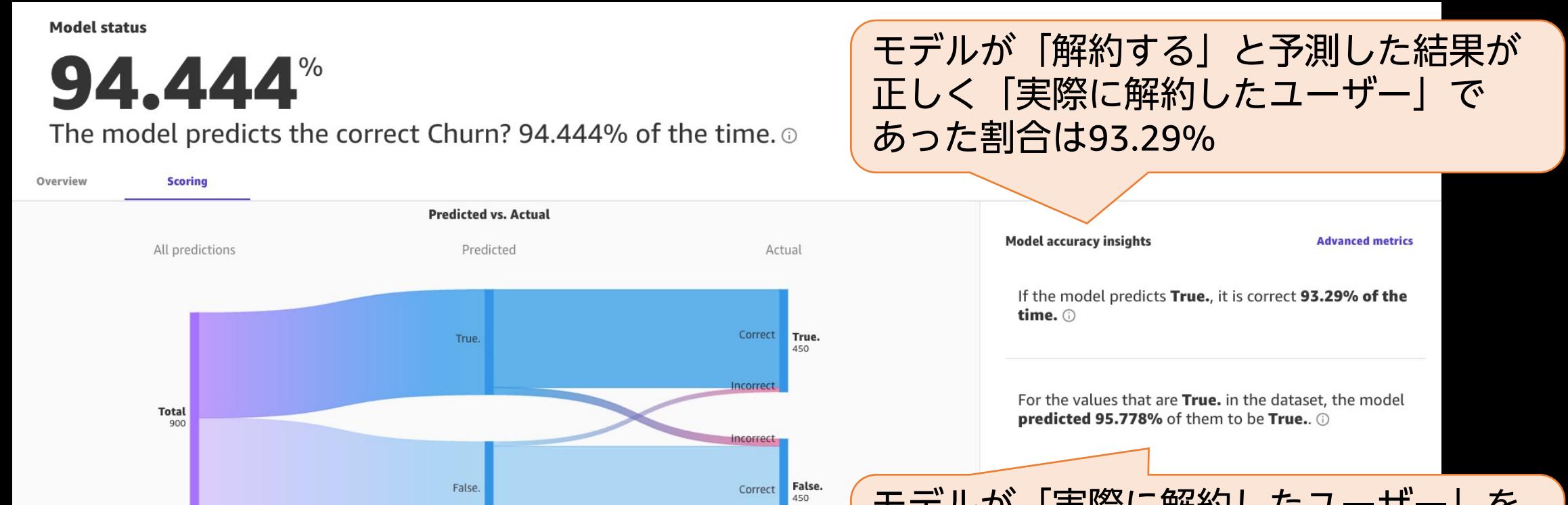


Feature Importance

モデルが解約する/しないを判断するのに使った列の重要度



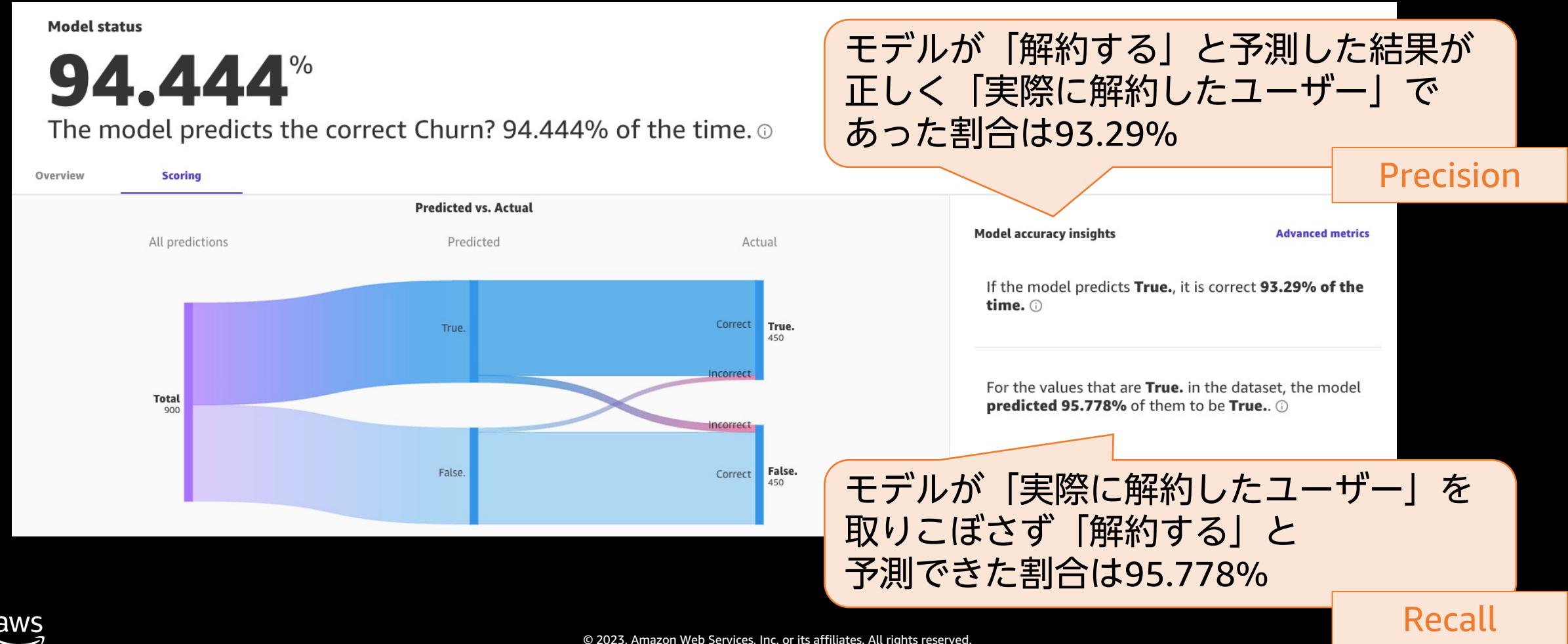
モデルの評価



モデルが「解約する」と予測した結果が正しく「実際に解約したユーザー」であった割合は93.29%

モデルが「実際に解約したユーザー」を取りこぼさず「解約する」と予測できた割合は95.778%

モデルの評価



モデルの評価



ビジネスに利用できるかの評価例

解約すると予測したのでキャンペーン

キャンペーンによって
N%解約を防止
できると見込む

解約しないのに
無駄にキャンペーン
してしまう



解約しないと予測したので放置

解約したが
放置

放置したが
解約しなかった

$(431 \times N\% \times \text{解約を防止出来たときの利益}) > ((431 + 31) \times \text{キャンペーン単価} + \text{システムコスト})$
を満たす場合はビジネスに導入できそう

[デモ] 予測の実行



[デモ] 予測の実行

CustomerChurn

V1 • Ready Add version Share :

Select Build Analyze Predict

Predict target values

Batch prediction Single prediction

Modify values to predict Churn? in real time.

Filter columns

Column	Feature importance	Value
Eve Mins	30.089%	0.00465930017438509
Day Mins	17.591%	0.00021480176744415
Day Charge	13.354%	0.00477677863207759
Night Charge	12.642%	0.05486252011786785
Night Mins	7.424%	0.00846773672108724
Night Calls	2.953%	200
Day Calls	2.952%	4

Churn? Prediction Copy

False.

Average prediction

True. 0.288% ⓘ

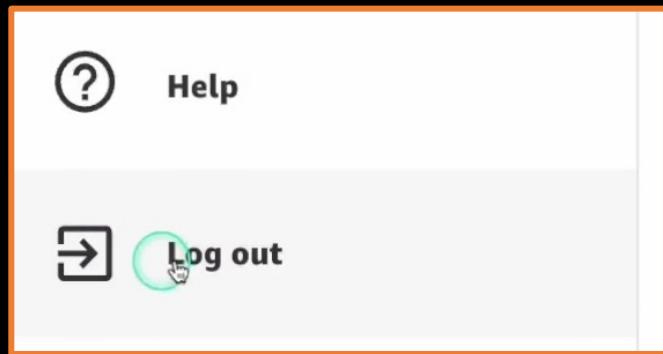
False. 99.712% ⓘ

The screenshot shows the 'CustomerChurn' notebook in AWS SageMaker Studio. The 'Predict' tab is selected. On the left, there's a table of feature importance and values for a single prediction. On the right, a bar chart shows the predicted 'Churn?' status as 'False.' with a confidence of 99.712%.

[デモ] ログアウト



[デモ] ログアウト



The screenshot shows the 'Datasets' page in the Amazon SageMaker Canvas interface. The page has a header with a search bar labeled 'Dataset'. Below the header is a table with columns: Name, Source, Columns, Rows, Cells, Created, and Status. The table lists ten datasets:

Name	Source	Columns	Rows	Cells	Created	Status
ChurnTest.csv	Local	21	500	10,500	11/21/2022 12:04 AM	Ready
ChurnTrain.csv	Local	21	4,500	94,500	11/21/2022 12:04 AM	Ready
canvas-sample-loans-part-2.csv	S3	5	1,000	5,000	11/20/2022 10:53 PM	Ready
canvas-sample-loans-part-1.csv	S3	19	1,000	19,000	11/20/2022 10:53 PM	Ready
canvas-sample-shipping-logs.csv	S3	12	1,000	12,000	11/20/2022 10:53 PM	Ready
canvas-sample-sales-forecasting.csv	S3	5	1,000	5,000	11/20/2022 10:53 PM	Ready
canvas-sample-maintenance.csv	S3	9	1,000	9,000	11/20/2022 10:53 PM	Ready
canvas-sample-housing.csv	S3	10	1,000	10,000	11/20/2022 10:53 PM	Ready
canvas-sample-diabetic-readmission.csv	S3	16	1,000	16,000	11/20/2022 10:53 PM	Ready
canvas-sample-product-descriptions.csv	S3	5	120	600	11/20/2022 10:53 PM	Ready

作業終了後は
ログアウトを忘れずに！

事例紹介



株式会社マルハン：これまで手作業だった店舗の売上予測業務を、AIを使って素早く高精度に実施



お客様プロフィール



INDUSTRY
アミューズメント事業

COUNTRY
日本

パチンコ事業を中心に、ボウリング、アミューズメント、シネマなどレジャー事業に関する業務の経営



ビジネスの課題

- 店舗毎の売上予測をしている業務があるが、役職者の業務時間を取っていることと、知識の属人化と経験の差により、予測精度の誤差が大きくなっている。



ソリューション

- Amazon SageMaker Canvasを活用し、AIによる売上予測を実施
- GUIによって、直感的な操作で活用可能



導入効果

- 役職者の売上予測業務に関する時間削減
- 従来の予測と比較して、より優れた精度での予測を実現
- 将来的に、予測モデルの活用範囲を広げて、バックエンド業務の効率化を視野に入れる。

「 Amazon SageMaker Canvasを使用することで本来機械学習を行うのに必要な知識が無くとも機械学習を実行することが出来ます。またその気軽さは属人化を防ぎAIをより身近にし、企業の成長と人の成長に大きく寄与すると考えます。」

南 力心 氏

株式会社マルハン

北日本カンパニー 営業推進部 営業戦略課 チーフ



まとめ



本セッションのまとめ

- 機械学習の需要は増えている
 - 人材の供給が追いついていないという課題
- 機械学習を導入する際に時間と人的リソースを削減するためまずはノーコードのAmazon SageMaker Canvasでクイックに検証をはじめてみる
- 評価と検証を繰り返していく
 - 必要に応じて他のサービスを利用していく



More Information

- Amazon SageMaker Canvas ドキュメント
 - <http://aws.amazon.com/sagemaker/canvas>
 - <https://docs.aws.amazon.com/sagemaker/latest/dg/canvas.html>
- Amazon SageMaker Canvas Immersion Day ワークショップ
 - <https://catalog.workshops.aws/canvas-immersion-day/ja-JP>

aws workshop studio

Amazon SageMaker Canvas Immersion Day

▶ 前提条件

- ラボ 1 - 解約予測 (マーケティング)
- ラボ 2 - 住宅価格 (不動産)
- ラボ 3 - 売り上げ予測 (小売)
- ラボ 4 - 債務不履行予測 (金融)
- ラボ 5 - 糖尿病患者の再入院予測 (ヘルスケア & ライフサイエンス)
- ラボ 6 - 機器の故障タイプ予測 (製造)
- ラボ 7 - 貨物到着日数予測 (運送 & 物流)

見る [YouTube](#)

Amazon SageMaker Canvas - Generate accurate M... リンクをコ...

re:Invent 2021 における SageMaker Canvas のプレゼンテーションはこちらの動画です。 (1:36:01 ~)

aws workshop studio

Amazon SageMaker Canvas Immersion Day

▶ 前提条件

ラボ 1 - 解約予測 (マーケティング)

ラボ 2 - 住宅価格 (不動産)
ラボ 3 - 売り上げ予測 (小売)
ラボ 4 - 債務不履行予測 (金融)
ラボ 5 - 糖尿病患者の再入院予測 (ヘルスケア & ライフサイエンス)
ラボ 6 - 機器の故障タイプ予測 (製造)
ラボ 7 - 貨物到着日数予測 (運送 & 物流)
ラボ 8 - Predicting website shopper sales conversion (Retail)

Overview タブで各列のインパクトを確認しましょう。この情報で提供できるインサイトは、マーケティングチームが顧客解約を減らすのに役立ちます。例えば、CustServ Calls が少ない場合も多い場合も、解約の可能性が高くなることがあります。このインサイトから、マーケティングチームは顧客解約を防ぐためのアクションを取ることができます。たとえば、顧客サービスへの問い合わせを減らすために Web サイトで詳細 FAQ を作成したり、その FAQ で顧客との教育キャンペーンを実行してエンゲージメントを向上させることができます。

Model status
95.9%
The model predicts the correct Churn? 95.9% of the time.

Model resource insights
Model accuracy
Model performance

More Information



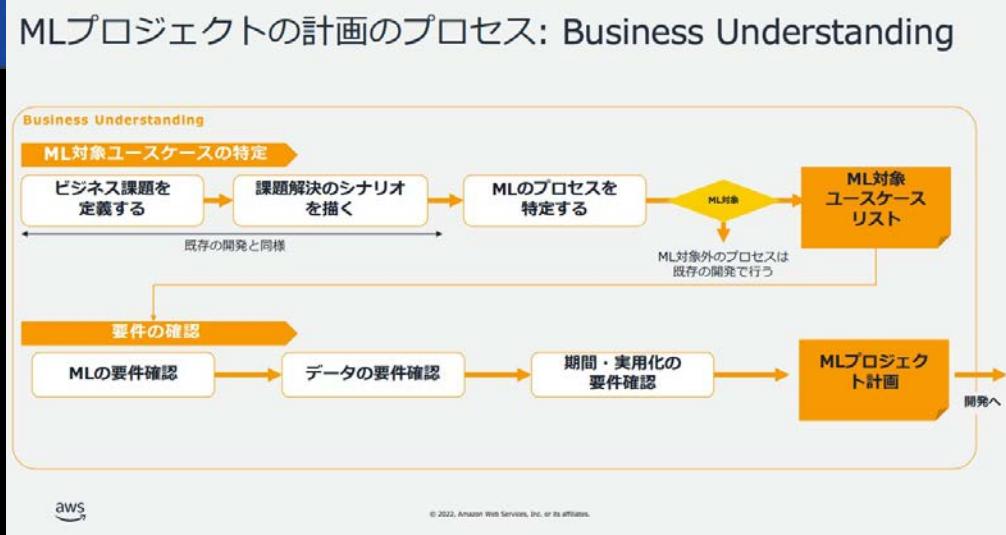
aws

ML Enablement Workshop: ML Input module #2

MLプロジェクトの 計画

久保 隆宏
Developer Relation
Machine Learning

© 2022, Amazon Web Services, Inc. or its affiliates.

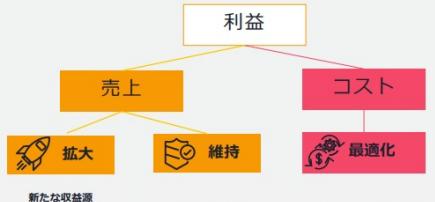


ビジネス課題を定義する

ビジネス課題はTo-beとAs-isのギャップにより定義される。

ビジネスKPIによる課題定義

To-beのKPIとAs-isのKPIのギャップ=課題
Ex: To-beの利益率はx%だがAs-isはy%である。



利益

売上

コスト

拡大

維持

最適化

新たな収益源
既存売上の増加

顧客の囲い込み

オペレーションの効率化

ユーザー体験による課題定義

To-beの体験とAs-isの体験のギャップ=課題
Ex: To-beは銀行口座があれば自動的に確定申告が完了するが
As-isはユーザー自身が事業経費を仕分けて申告する必要がある。



To-be

個人事業主 確定申告をする 入出金を合算した 入出金 事業の経費を分類した 事業経費 日別額を計算した 事業報告を提出した 中古

As-is

個人事業主 確定申告をする 能役者を探した 個人情報を提出している 個人情報を提出している 事業報告が会社が確認した 事業経費 信頼性を計算した 事業報告を提出した 中古

© 2022, Amazon Web Services, Inc. or its affiliates.

ML Enablement Workshop

<https://github.com/aws-samples/aws-ml-enablement-workshop/blob/main/docs/presentations/ml-enablement-workshop-module2.pdf>



More Information

- AI/ML Black Belt Online Seminar @YouTube 始めました

Light Part

製品やサービスに機械学習を導入するプロジェクトの進め方

<https://bit.ly/3M1F9as>



Step Up!!

Dark Part

機械学習モデルの開発や運用をマネジドサービスで効率的に行う方法

<https://bit.ly/3927PCN>



More Information

セッション視聴後に、ハンズオンにチャレンジ！
プラットフォームからいつでもアクセスいただけますので
自由な時間にお試しいただけます。
「Z1: (関連資料) ハンズオン」より今すぐアクセスいただけます！



AWS TRAINING & CERTIFICATION

AWS Skill Builder の 500+ の 無料デジタルコースで学ぼう

30以上のAWSソリューションの中から、自分に最も関係のあるクラウドスキルとサービスにフォーカスし、自習用のデジタル学習プランとRamp-Upガイドで学ぶことができます。

- 自分のペースでAWSクラウド上を活用した未来を切り開く
- 学習プランでスキルや知識を向上
- AWS認定資格でクラウドの専門知識を証明する



自分に合ったスキルアップ方法をで学びましょう
[EXPLORE.SKILLBUILDER.AWS »](https://explore.skillbuilder.aws)



AWS Builders Online Series に ご参加いただきありがとうございます

楽しんでいただけましたか? ぜひアンケートにご協力ください。
本日のイベントに関するご意見/ご感想や今後のイベントについての
ご希望や改善のご提案などございましたら、ぜひお聞かせください。



aws-apj-marketing@amazon.com



twitter.com/awscloud_jp



facebook.com/600986860012140



<https://www.youtube.com/user/AmazonWebServicesJP>



<https://www.linkedin.com/showcase/aws-careers/>



twitch.tv/aws



Thank you!

