

AWS Builders Online Series

もし CEO が『我が社も AI を』と言ったら ～ Amazon SageMaker Canvas による処世術～

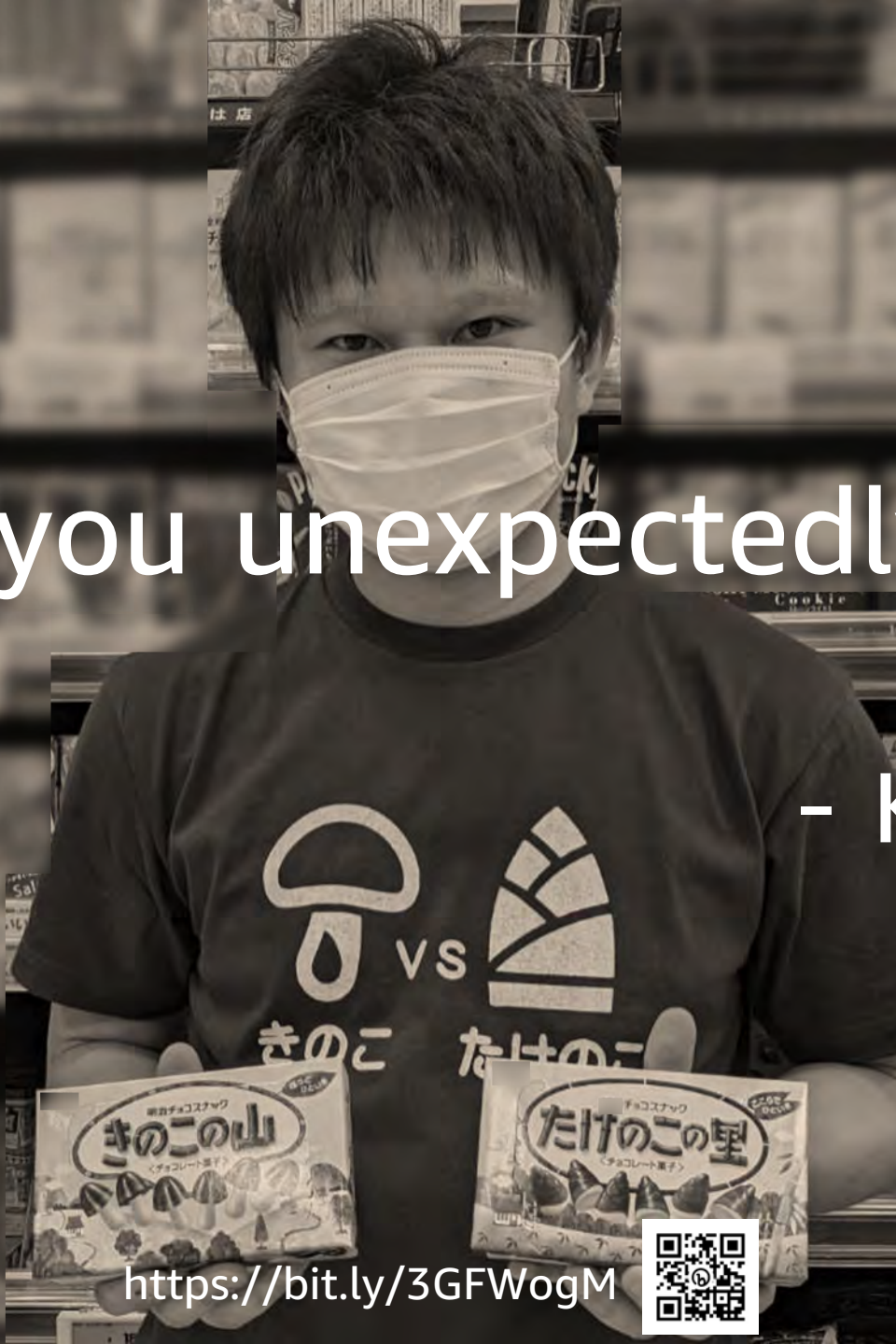
呉 和仁

アマゾン ウェブ サービス ジャパン合同会社
AWS 技術統括本部
機械学習ソリューションアーキテクト



ML comes to you unexpectedly.
Be prepared.

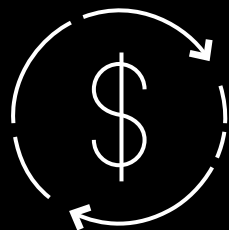
- K. Go(1986~)



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



いつでも誰でも機械学習の当事者になりえる



加速する投資

2024年までに、人工知能に対する投資は1100億ドルに到達する

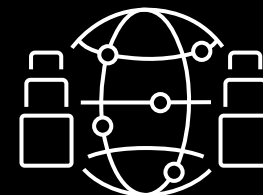
—IDC



検討から運用へ

2024年末までに、75%の企業が検討フェイズから運用フェイズへ移行する。

—Gartner



AIによる変革

2025年までに新しいエンタープライズアプリケーションの90%以上にAIが組み込まれる

—Deloitte

セッションの概要

- 機械学習にいつエンカウトしても大丈夫なように今日から使える処世術の1つをロールプレイを通して紹介
- 想定視聴者は以下のいずれかに該当する方
 - CEO/上司が突然 AI という方
 - Amazon SageMaker Canvas に興味がある方
- 対象としない方
 - ~~上司に No と言える方~~
 - 皆様聞いてください

自己紹介

- 機械学習ソリューションアーキテクト
- コンピュータに仕事を全部任せたい
- 好きな AWS サービス
 - AI サービス全般
 - Amazon SageMaker
- たけのこ > きのこ



目次

1. 機械学習は突然に
2. 機械学習、その前に
3. 私も機械学習ひとりでできるもん
4. 完全に理解した私とその後

目次

1. 機械学習は突然に

2. 機械学習、その前に

3. 私も機械学習ひとりでできるもん

4. 完全に理解した私とその後

CEO の黙示録

我が社も AI でなにか

うちの本部も
機械学習でなにか

離反防止を機械学習
でなにか

ナンダッテー!!!

Wakkan Ai CEO

携帯電話の
通信キャリアを経営

マーケティング
本部長

新規加入や離反防止を担当
離反数の多さに不満を抱える

カスタマー
レーション部長

離反するのは回線プラン
の問題だと思っている

平社員

G くん

離反抑止のための
施策立案実行を担う

機械学習をやったことはない

(前提として)そのアプローチの筋は良くはない

ビジネス課題を解く方法の1つとして機械学習がある

改善したいビジネスの内容を決める(why, what)

ex1) イベントでお客様の入退場の時間がかかりすぎているから半分にしたい

ex2) 生産ラインで不良品が多発しているから1/10にしたい



How を決めてできるか検証

ex1) 機械学習を用いた顔認証システムを使えるかもしれないから試してみる

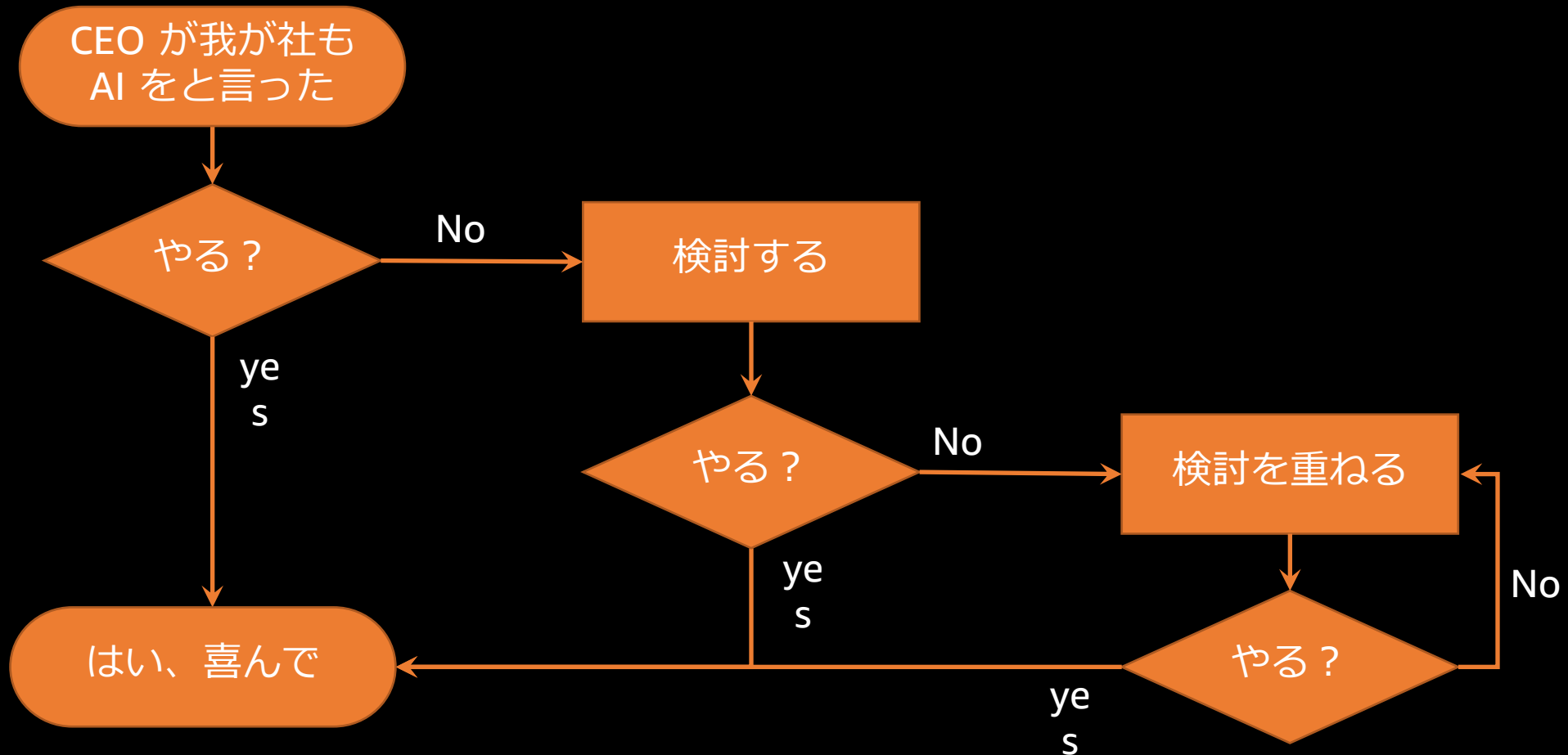
ex2) まずは原因をデータ分析してみる (≠機械学習)

→工程や道具の改善で改善するのがよいかもかもしれない

<https://docs.aws.amazon.com/wellarchitected/latest/machine-learning-lens/machine-learning-lens.html>

取れる選択肢は少ない

前向きな姿勢を見せる必要があることは多い



目次

1. 機械学習は突然に

2. 機械学習、その前に

3. 私も機械学習ひとりでできるもん

4. 完全に理解した私とその後

機械学習を始める前に知っておきたいこと

「教師あり学習」で「回帰」と「分類」をまず抑えてほしい

- データから連続値を予測する「回帰」
 - 過去の売上データから明日の売上を予測
 - インフラの敷設状況から残存寿命予測
 - 広告のクリック率を予測
- データからクラスを予測する「分類」
 - 画像に映っているお菓子を予測
 - トランザクションデータが正常か不正かを予測
 - ユーザーの利用履歴から継続か離反を予測

最初に必要なのは、緑(学習データ)と
オレンジ(教師ラベル)の過去データのセット



[機械学習 離反抑止] < カチッ

検索しユースケースなどを見ておく

- 先行事例を学ぶ
- 業界が違ってくるならば許容
- どんなデータを使って予測しているのかを確認

(参考) AWS における資料

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

<https://aws.amazon.com/jp/events/aws-innovate/machine-learning/resources/>

<https://aws.amazon.com/jp/ai/aiml-resources/>

<https://aws.amazon.com/jp/blogs/news/lekker-building-customer-churn-model-with-sagemaker-debugger/>



Amazon Web Services ブログ

lekker 社が顧客離反モデルの開発で Amazon SageMaker Debugger を利用して、より多くの洞察を得た方法

by Katsuhisa Niwa | on 16 JUL 2021 | in Amazon SageMaker, Artificial Intelligence, Industries, Power & Utilities | Permalink |

Share

この記事は、「How lekker got more insights into their customer churn model with Amazon SageMaker Debugger」を翻訳したものです。

離反予測にはテーブルデータ

機械学習は一般的に画像やテキスト、テーブル(表)データをよく使う



機械学習で画像から
お菓子を検出する例



<https://go.aws/3GA5uLX>



小説からキーワードなどを検出して
読書感想文を楽に書く例



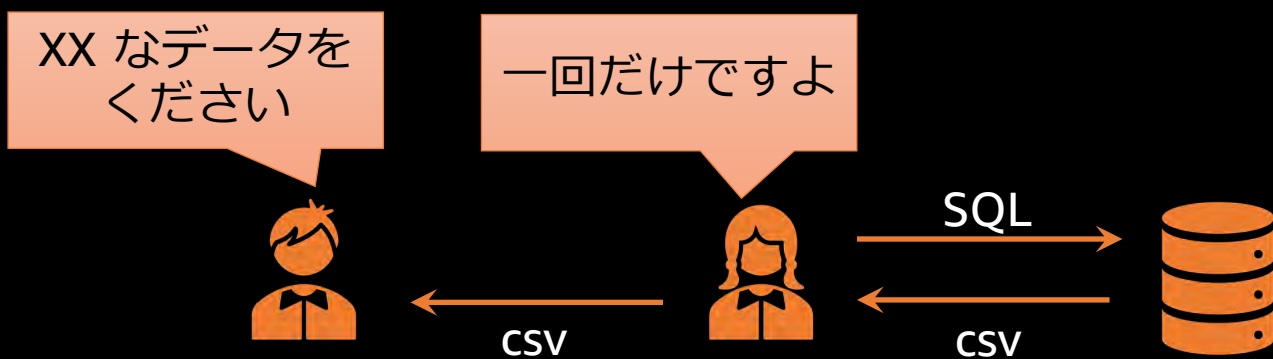
<https://go.aws/3N5mN9P>

扱いやすいテーブルデータ (csv) で用意することをまず検討

csv を入手する

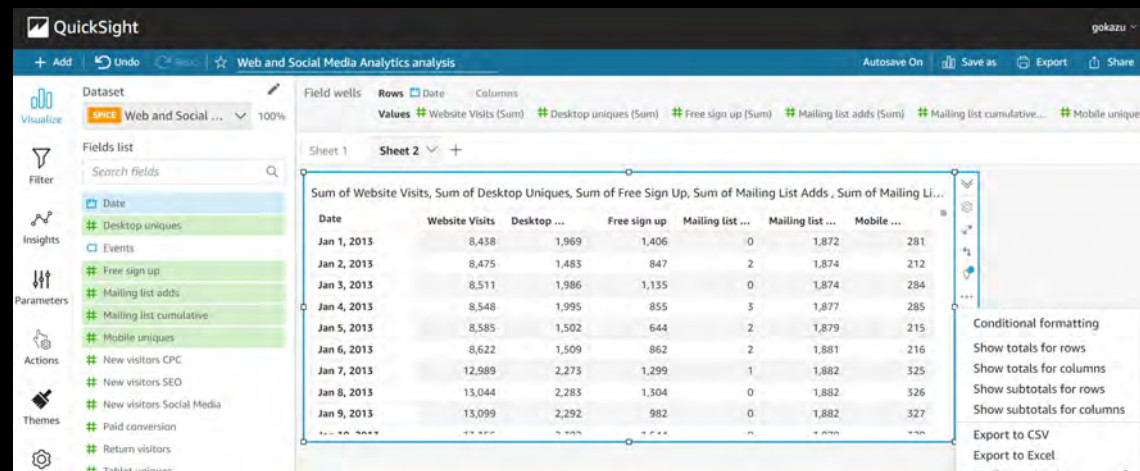
csv 入手経路 3 例

データベースにアクセスできる人(社内SE,DBA,ビジネスアナリスト,etc...)に出力してもらおう



自分が Excel ファイルを持っている
場合は変換する

BI ツールから出力



入手した(とする)データセット

s3://sagemaker-sample-files/datasets/tabular/synthetic/churn.txt

<https://sagemaker-sample-files.s3.us-east-1.amazonaws.com/datasets/tabular/synthetic/churn.txt>

学習データ

教師ラベル

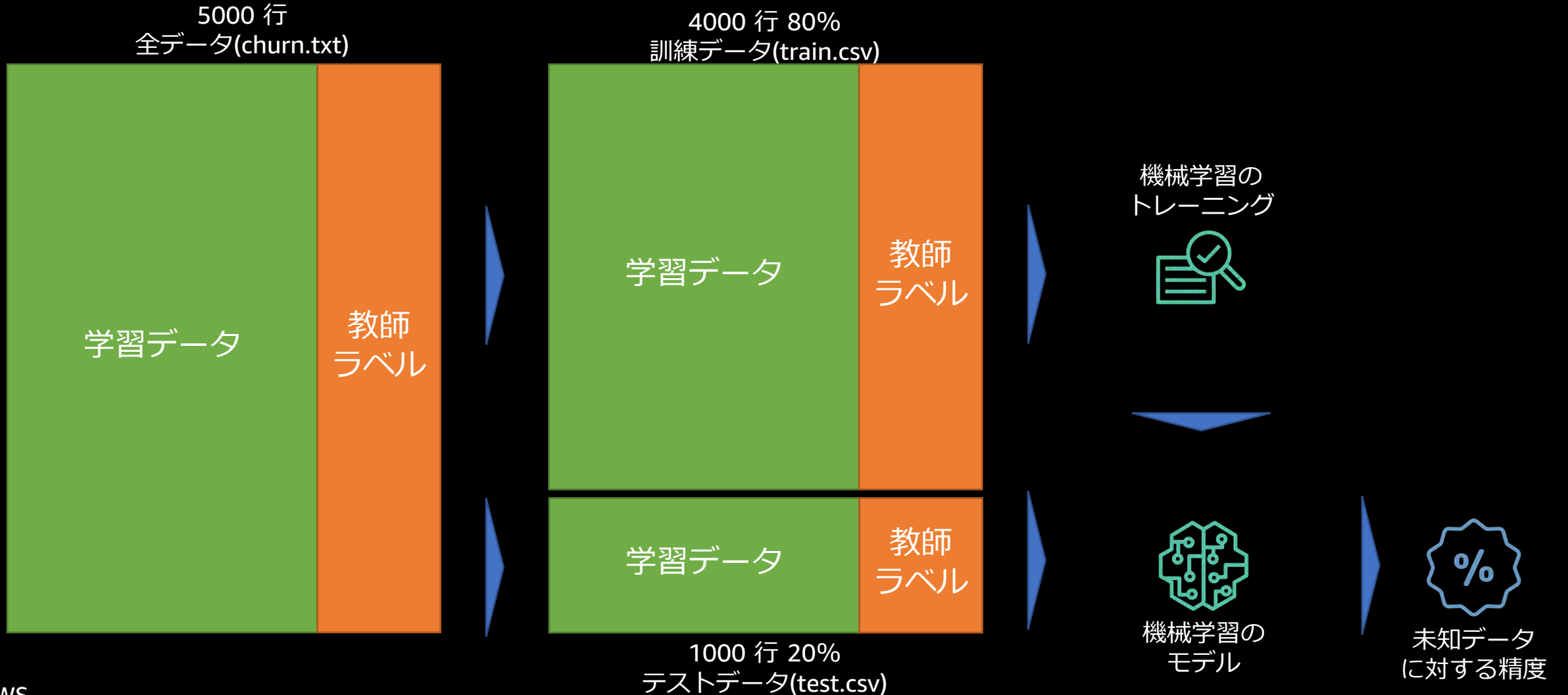
State	Account Length	Area Code	Phone	Int'l Plan	VMail Plan	VMail Message	Day Mins	Day Calls	Day Charge	Eve Mins	Eve Calls	Eve Charge	Night Mins	Night Calls	Night Charge	Intl Mins	Intl Calls	Intl Charge	CustServ Calls	Churn?
PA	163	806	403-2562	no	yes	300	8.162204022	3	7.579173703	3.933034994	4	6.508638877	4.065759458	100	5.111623915	4.928160206	6	5.673203041	3	True.
SC	15	836	158-8416	yes	no	0	10.01899266	4	4.226288822	2.325004553	0	9.972592415	7.141039872	200	6.436187619	3.221747623	6	2.559749162	8	False.
MO	131	777	896-6253	no	yes	300	4.708489917	3	4.768159882	4.537466276	3	4.566714502	5.363234681	100	5.142450865	7.139023053	2	6.254156588	4	False.
WY	75	878	817-5729	yes	yes	700	1.268733772	3	2.567641586	2.528747882	5	2.33362363	3.773585578	450	3.814413422	2.245779328	6	1.080692041	6	False.
WY	146	878	450-4942	yes	no	0	2.696177238	3	5.908915869	6.015337072	3	3.670407748	3.751673372	250	2.796811998	6.905545286	4	7.134342911	6	True.
VA	83	866	454-9110	no	no	0	3.634775967	7	4.804891887	6.051944336	5	5.278436965	2.937879987	300	4.81795844	4.948816088	4	5.135322741	5	False.
IN	140	737	331-5751	yes	no	0	3.229420256	4	3.165082061	2.44015281	8	0.26454309	2.352273898	300	3.869176348	5.393439249	4	1.78476482	4	False.
LA	54	766	871-3612	no	no	0	0.567920183	6	1.950098388	4.507026828	0	4.473085756	0.688785148	400	6.132136689	5.012747241	5	0.417420939	8	False.
MO	195	777	249-5723	yes	no	0	5.811115762	6	4.331065017	8.104126146	2	4.475034293	4.208352015	250	5.97457535	4.750152913	7	3.32031079	7	True.
AL	104	657	767-7682	yes	no	0	2.714429979	7	5.138669204	8.529943777	6	3.321121008	2.342176971	300	4.328965743	3.433554182	5	5.677057853	4	False.
CA	137	676	185-4876	no	yes	800	3.85904161	3	3.80590292	1.958122519	2	5.865065968	5.601397036	350	3.436449463	4.955366293	9	1.035231207	7	False.
RI	78	827	276-8625	yes	no	0	6.872939364	2	6.534157344	3.696520174	2	5.826586242	3.497215844	150	6.710051923	6.194516763	4	3.072914598	7	True.
GA	184	716	815-7162	no	no	0	1.477758591	6	5.532156653	6.718964555	8	1.061149272	2.059022754	350	0.959632984	6.617656599	7	6.191955344	3	False.
PA	150	806	610-2756	yes	no	0	8.735239799	3	5.607474259	6.959893299	1	8.167985577	2.839336261	150	5.149192276	5.342917778	6	8.143195635	4	True.
ND	114	786	968-6867	yes	yes	300	6.215980907	4	6.705125867	5.839379503	4	4.240211919	2.700893798	250	7.533492316	5.491468775	5	0.470645092	8	True.
MS	114	778	756-1787	no	no	0	3.333268667	3	4.888256309	4.779260734	1	4.86477525	2.84992155	200	5.206421652	4.598937925	3	4.77138846	8	True.
WY	48	878	889-7926	yes	yes	100	9.813813712	3	7.759409291	8.670985701	3	6.472732176	4.728587845	250	6.254585699	4.284906031	5	5.661948756	5	True.
NY	142	788	248-9207	yes	no	0	0.631607108	4	0.81736356	4.399597543	0	2.024482319	3.156504871	200	6.124743097	5.644293003	3	1.532017352	10	True.
NJ	111	787	843-5355	yes	yes	700	11.61561856	2	7.898472493	9.791103721	0	7.37576447	5.144235389	250	6.855467798	4.413320414	4	8.499379731	6	False.
NM	63	787	360-2799	yes	yes	300	7.331254865	3	8.456739886	9.617906177	6	3.818931666	6.528778715	250	4.539917114	4.952046947	3	8.911840816	5	False.
ID	124	736	199-4929	yes	no	0	6.350164019	6	5.278668211	6.174900894	4	5.825235277	5.097930528	150	5.93179335	4.882246831	3	5.576146181	6	True.
DE	111	686	620-9126	no	yes	500	2.018493489	6	5.820356743	8.19113534	6	2.428186871	1.333267799	400	4.967204287	5.256095958	5	4.673462287	5	False.
WY	182	878	671-4825	yes	yes	200	1.67522982	6	3.999441248	0.642391947	8	4.680588723	3.544147898	100	5.530137404	5.882881558	1	3.308140601	4	False.
IN	162	737	822-7473	yes	yes	500	1.637895653	4	1.253564333	2.857913575	3	4.654391689	1.858334478	150	4.655465974	3.909770306	3	4.20492245	8	False.
IL	128	737	247-5360	no	no	0	3.255026006	3	6.43792988	7.138773609	4	3.031720999	2.851625128	300	5.172175165	4.161727937	4	3.922241844	7	True.
CA	14	676	641-1882	no	no	0	5.435168039	1	5.036441163	7.246933156	0	6.213823231	2.978703792	300	5.506355457	4.740097542	3	6.964495186	7	True.
NM	56	787	822-3322	yes	yes	200	4.579843133	2	2.001178048	2.525309572	5	0.024721852	3.731893514	150	1.66765635	3.882070668	2	6.690489283	3	False.
WA	17	876	870-2858	yes	no	0	3.648750847	6	5.789514629	6.374691811	8	4.005674596	4.315006821	200	3.851060791	5.904362433	3	7.297146003	2	False.
ND	48	786	998-3997	yes	yes	100	1.087469195	4	1.779403327	5.075619562	1	2.214029199	1.323951293	150	4.425674917	5.844585094	4	3.718192501	6	False.

a

served.

データ分割

モデルのテスト用にデータを予め分割しておく



目次

1. 機械学習は突然に

2. 機械学習、その前に

3. 私も機械学習ひとりでできるもん

4. 完全に理解した私とその後

Amazon SageMaker Canvas

ML モデルを構築し、正確な予測を生成
— コーディングは不要



機械学習のためのデータへの迅速なアクセスと準備



組み込み AutoML でモデルを構築し、正確な予測を生成



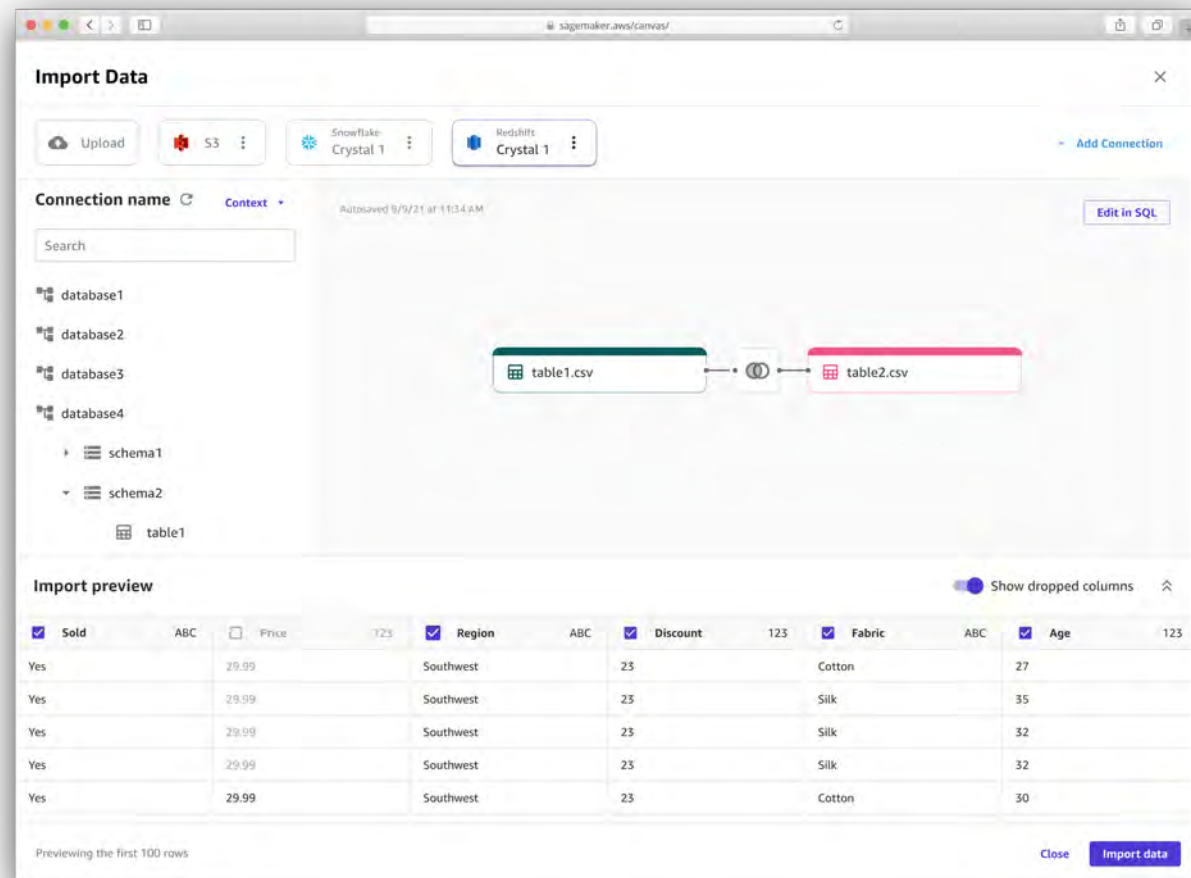
ML モデルの共有とデータサイエンスチームとの連携



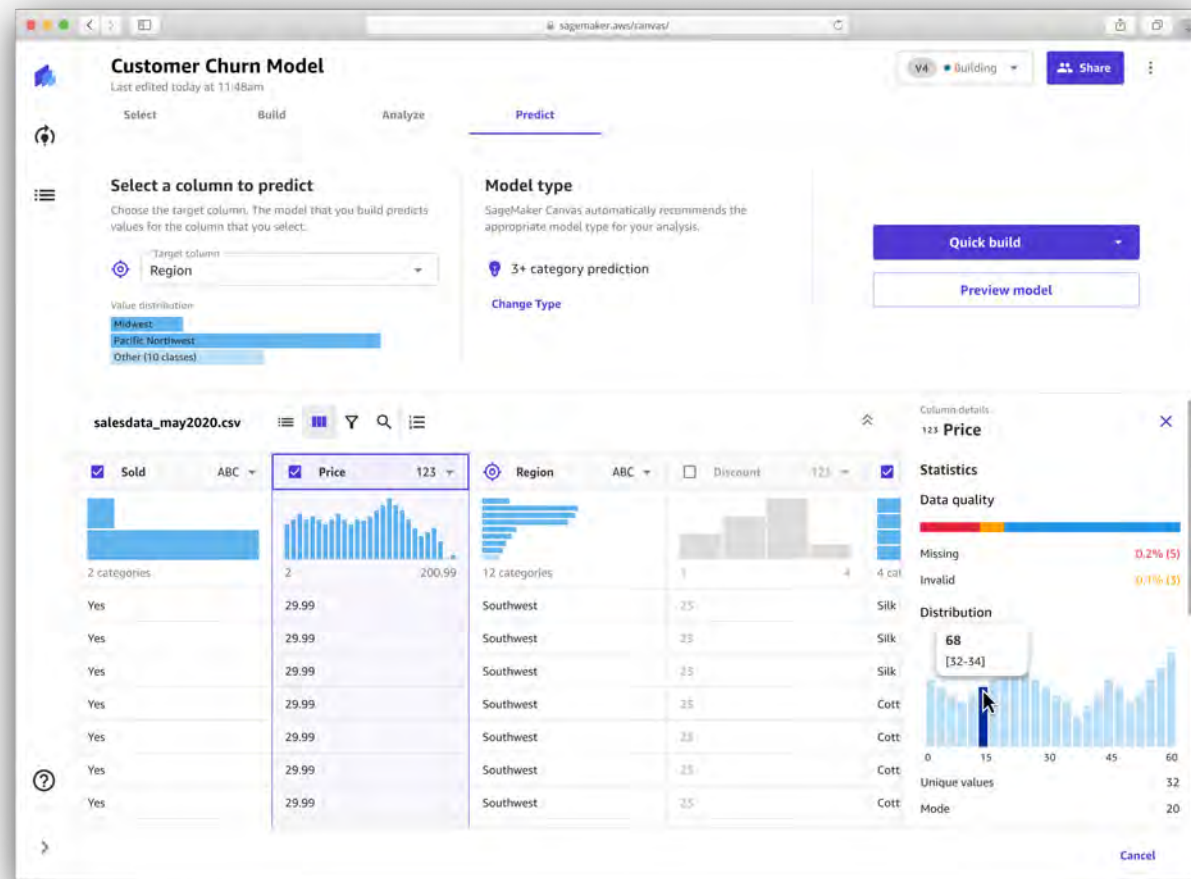
ライセンス料を回避し、TCO を削減するための使用量ベースの価格設定

ローカルディスク、
Amazon S3、Amazon
RedShift、Snowflake
など、
さまざまなソースから
のデータセットを結合
することが可能

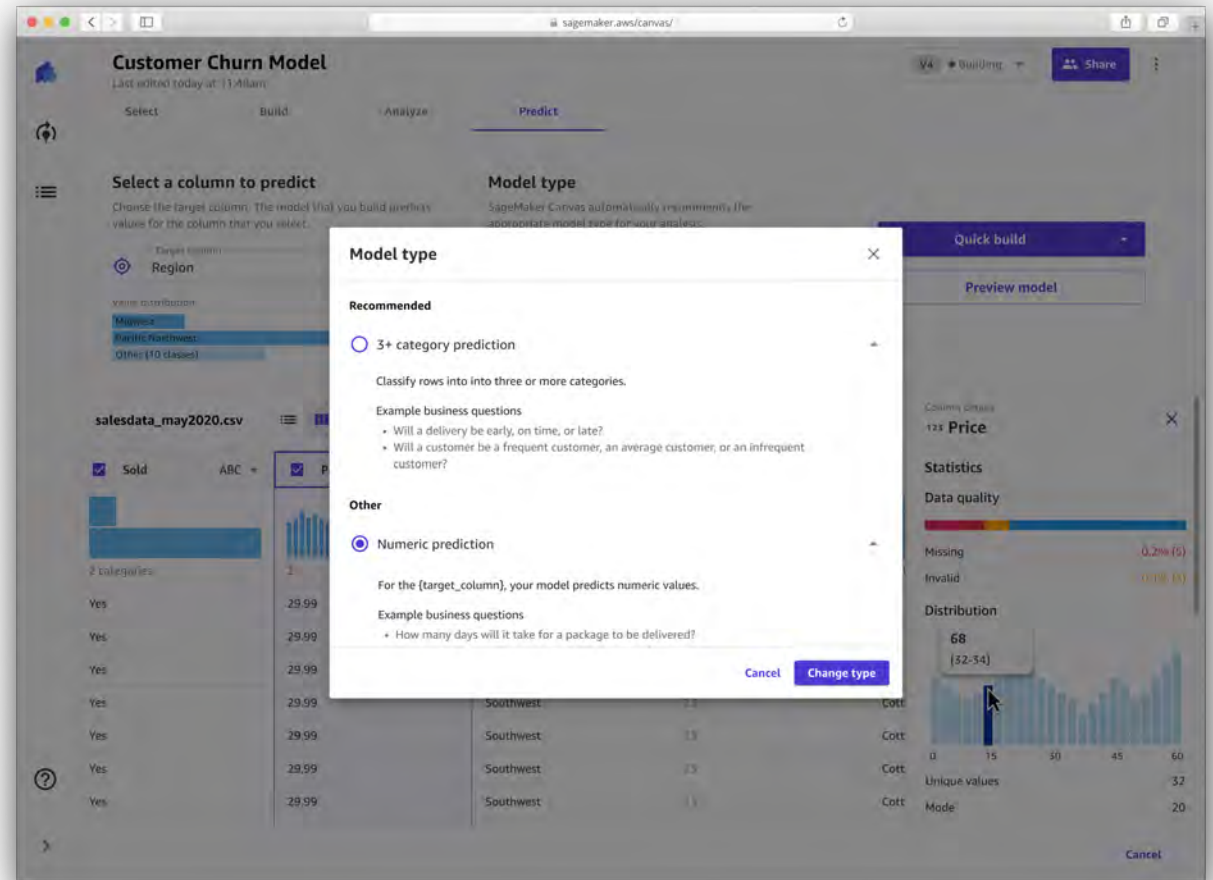
…他のデータソースも近日追加予定



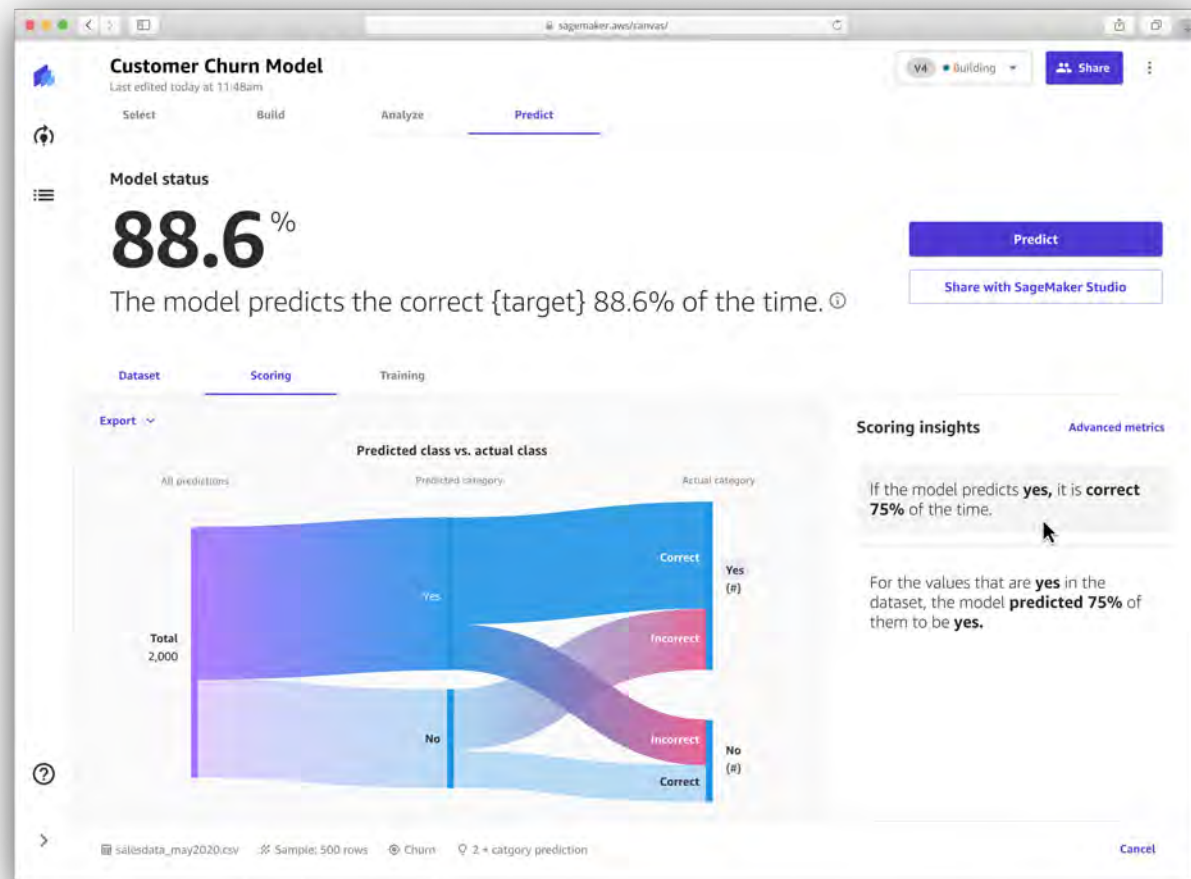
視覚的なインター
フェースにより、
データを素早く理解
し、準備する



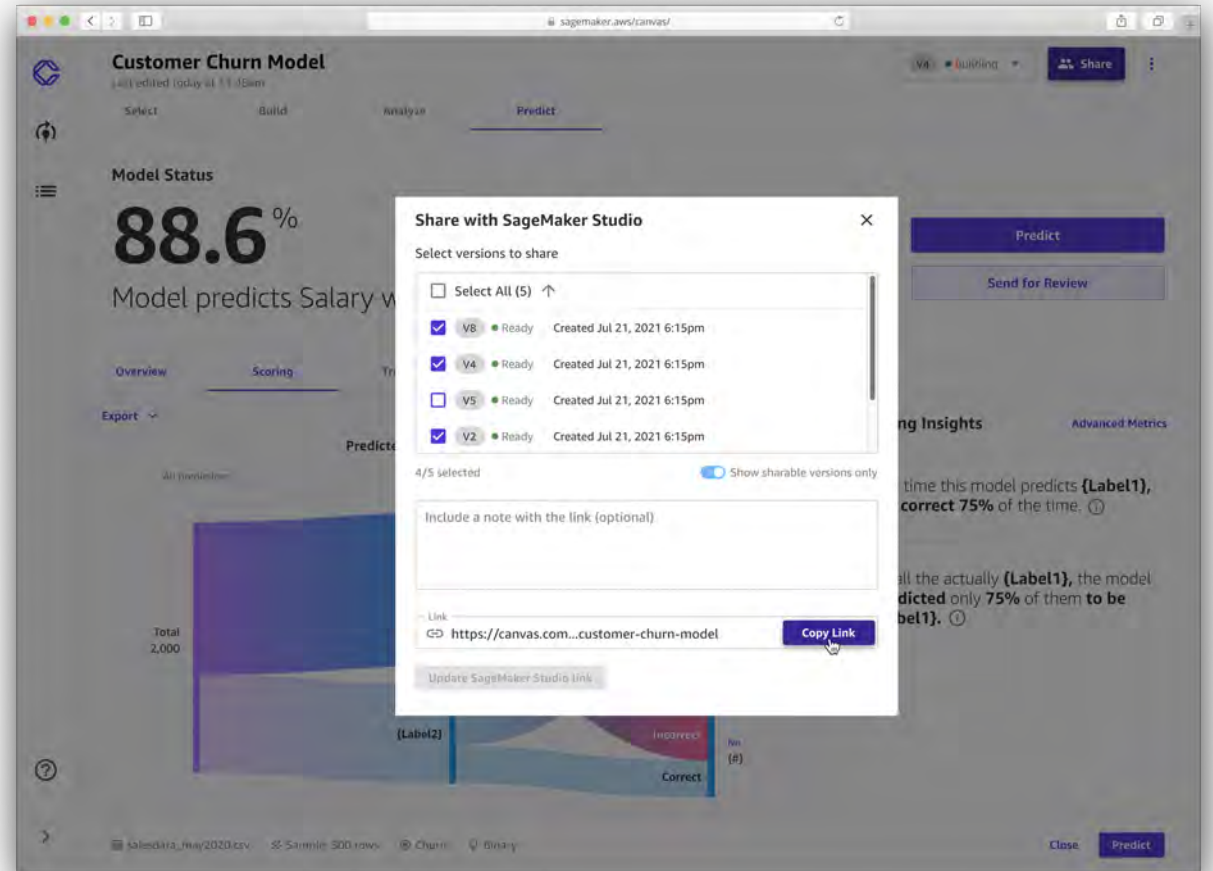
データセットに対する 正確な ML モデルを 自動で構築する



数分で最初の ML モデルを生成し、予測を理解し説明するために、高度なメトリクスと特徴量の重要性を確認



作成したモデルを
データサイエンティ
ストと共有し、簡単
にフィードバックを
得ることができる



what-if シナリオの
実行や、データセット
全体の予測を取得

Customer Churn Model
Last edited today at 11:48am

Select Build Analyze **Predict**

Predict target values

Batch prediction **Single prediction**

Modify values to calculate target column (in real time)

Filter columns

Column	Column impact	Value	Reset all to average
Contract	61.3%	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	
		Yes	

Churn prediction Copy

No

Average probability New probability

No 71.5%

Yes 28.5%

Close Download

SageMaker Canvas のユースケース

異なるビジネスファンクションやパーティカル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



オペレーション とロジスティクス

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
4. 予知保全

その他多数...

コンソールからSageMaker Canvas を立ち上げる

aws Services Search for services, features, blogs, docs, and more [Alt+S] Tokyo gokazu @ 5520-8113-2209

S3 Amazon SageMaker Elastic Container Registry

Console Home [Info](#) Actions ▾

Welcome to AWS

- [Getting started with AWS](#)
Learn the fundamentals and find valuable information to get the most out of AWS.
- [Training and certification](#)
Learn from AWS experts and advance your skills and knowledge.
- [What's new with AWS?](#)
Discover new AWS services, features, and Regions.

AWS Health [Info](#)

Open issues
0 Past 7 days

Scheduled changes
0 Upcoming and past 7 days

Other notifications
0 Past 7 days

[Go to AWS Health](#)




Amazon S3 に転送

aws Services Search for services, features, blogs, docs, and more [Alt+S] Tokyo gokazu @ 5520-8113-2209

S3 Amazon SageMaker Elastic Container Registry

Console Home [Info](#) Actions ▾

Welcome to AWS

-  [Getting started with AWS](#)
Learn the fundamentals and find valuable information to get the most out of AWS.
-  [Training and certification](#)
Learn from AWS experts and advance your skills and knowledge.
-  [What's new with AWS?](#)
Discover new AWS services, features, and Regions.

AWS Health [Info](#)

Open issues
0 Past 7 days

Scheduled changes
0 Upcoming and past 7 days

Other notifications
0 Past 7 days

[Go to AWS Health](#)

Feedback Looking for language selection? Find it in the new [Unified Settings](#)

© 2022, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

Sagemaker Canvas 🚀 Quick Build

Amazon SageMaker
Canvas

Models

Search models

+ New model

Models

Datasets

Help

Account

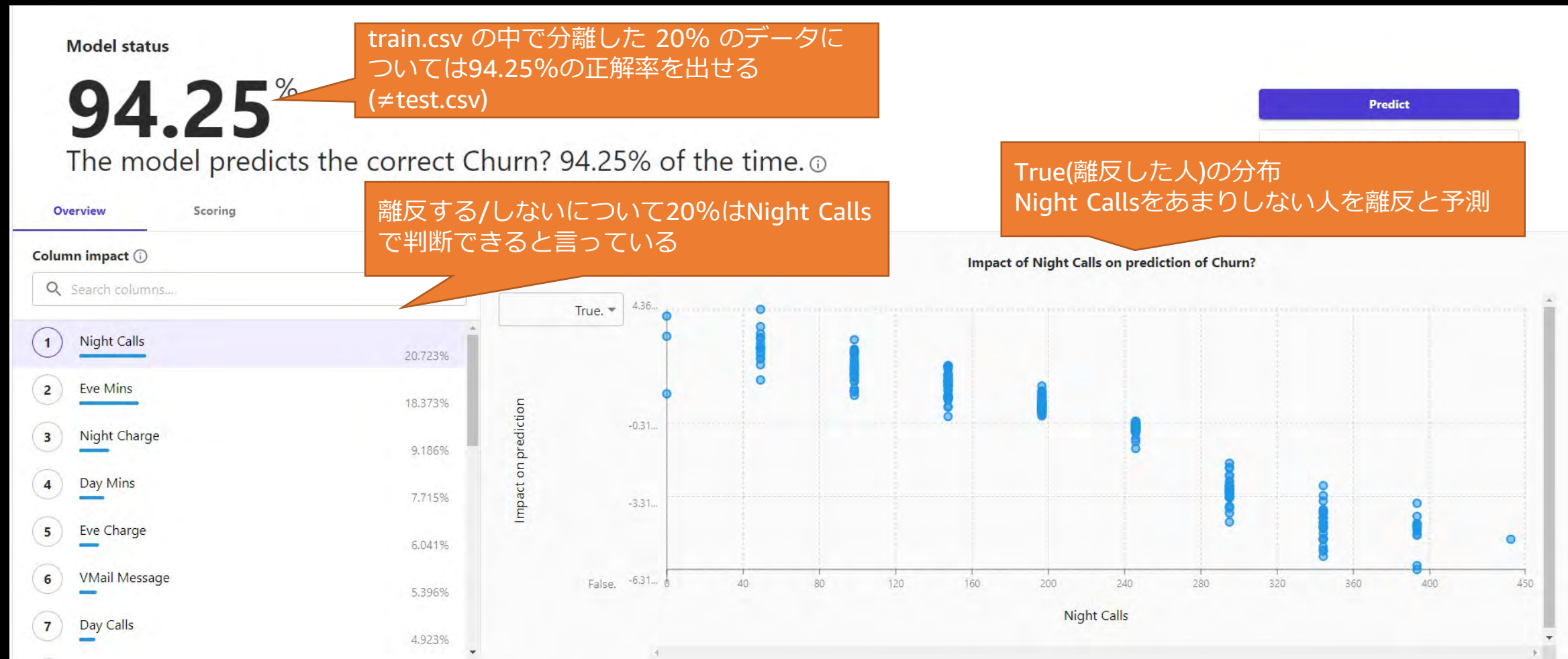
You haven't created any models yet.

Create a new model.

+ New model

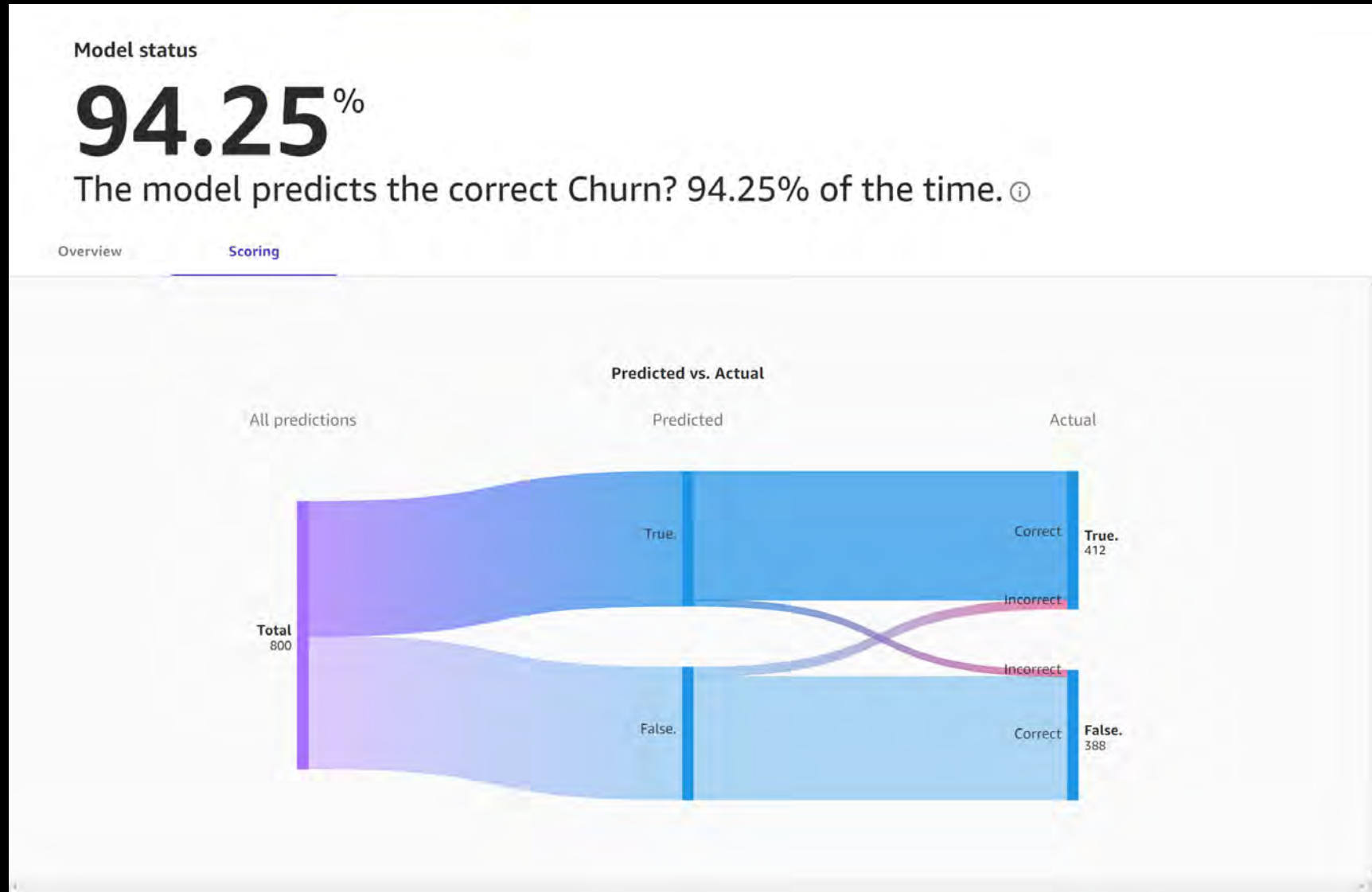
Feature Importance

- モデルが離反する/しないを判断するのに使った重要度



モデルの精度

- 評価用データ(train.csvの20%を自動で分離)で精度を知れる



使ってはいけないデータ

Phone 列を使ってはいけない



ビルド前に気づくことも出来た

Column name ↓	Data type	Missing ⓘ	Mismatched ⓘ	Unique ⓘ	Mean / Mode	Correlation to target ⓘ	Feature importance ⓘ
Phone	Text	0.00% (0)	0.00% (0)	4,000	100-2118	N/A	0.113279

使ってはいけないデータ例



ユニークな識別子

Phone の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



教師ラベルに直結するデータ

審査で最終審査結果を予測する場合に、
学習データに一次審査結果が入っている場合など



実際の業務で手に入らないデータ

前日のデータなど予測段階で間に合わないケースなど

Standard Build でモデルの再ビルド

my_first_churn

V1 Ready Add version Share

Select Build Analyze Predict

Model status

94.25%

The model predicts the correct Churn? 94.25% of the time. ⓘ

Predict

Share with SageMaker Studio

Overview Scoring

Column impact ⓘ

Search columns...

1	Night Calls	20.723%
2	Eve Mins	18.373%
3	Night Charge	9.186%
4	Day Mins	7.715%
5	Eve Charge	6.041%
6	VMail Message	5.396%
7	Day Calls	4.923%

Impact of Night Calls on prediction of Churn?

Impact on prediction

Night Calls

train.csv Total columns: 21 Total rows: 4,000 Total cells: 84,000 Churn? 2 category prediction

Predict

未知データで予測して精度検証

my_second_churn

V1 Ready Add version Share

Select Build Analyze Predict

Model status

97.794%

The model predicts the correct Churn? 97.794% of the time.

Predict

Share with SageMaker Studio

Overview Scoring

Column impact

Search columns...

Rank	Column	Impact
1	Night Calls	12.737%
2	Eve Mins	11.252%
3	Eve Charge	10.366%
4	CustServ Calls	9.545%

Impact of Night Calls on prediction of Churn?

Impact on prediction

False

True

train.csv Total columns: 20 Total rows: 4,000 Total cells: 80,000 Churn? 2 category prediction

Predict

業務に適用できるか検討

離反抑止でクーポンを配布する例

離反すると予測したのでキャンペーン

離反しないと予測したので放置

キャンペーンによって X%離反を防止		離反予測	継続予測
	離反	472	14
離反しないのに 無駄にキャンペーン	継続	21	493

離反してしまうが
放置

継続するだけ

A. $472 \times \text{離反阻止率} \times \text{解約阻止出来たときの利益} > (472+21) \times \text{キャンペーン単価} + \text{機械学習システムコスト}$
を満たす場合はビジネスに導入すべし

What-if 分析でモデルの挙動確認

- Single Prediction から入力値を自由に設定して予測結果を確かめることができる

Predict target values

Batch prediction Single prediction

Modify values to predict Churn? in real time.

Filter columns

Column	Feature importance	Value
Eve Mins	26.055%	0.005548398
Night Charge	23.393%	0.05486252
Day Charge	15.996%	0.004776779
Day Mins	11.453%	0.000214802
CustServ Calls	3.416%	6
Intl Charge	2.913%	0.0007694
Night Mins	2.641%	0.008467737

Churn? Prediction [Copy](#)

False.

Average prediction

True. 0.584% ⓘ

False. 99.416% ⓘ

本番適用へ

Select Build **Analyze** Predict

Model status

97.794%

The model predicts the correct Churn? 97.794% of the time. ⓘ

Predict

Share with SageMaker Studio

Launcher Canvas model SageMakerAutopilotDataExpl

Model overview

Read only V1: Created May 27, 02:59 PM

Canvas user ID gokazu	Shared versions 1 Version	Domain ID d-azbf54xop50l	Problem type Binary	Notes from Canvas
Model name my_second_churn	Current version Version 1	Input dataset train.csv	Best model Canvas303693	

Input Dataset Auto ML Job

Data Flows

Open data exploration notebook ⓘ

```
graph LR; Source[S3: train.csv] --> Transform1[Transform]; Transform1 --> Transform2[Transform]; Transform2 --> DropColumn[Drop column];
```



Amazon SageMaker Autopilot Data Exploration Report

This report contains insights about the dataset you provided as input to the AutoML job. This data report was generated by **Canvas303693** AutoML job. To check for any issues with your data and possible improvements that can be made to it, consult the sections below for guidance. You can use information about the predictive power of each feature in the **Data Sample** section and from the correlation matrix in the **Cross Column Statistics** section to help select a subset of the data that is most significant for making predictions.

Note: SageMaker Autopilot data reports are subject to change and updates. It is not recommended to parse the report using automated tools, as they may be impacted by such changes.

Dataset Summary

Dataset Properties

Rows	Columns	Duplicate rows	Target column	Missing target values	Invalid target values	Detected problem type
4000	20	0.00%	Churn?	0.00%	0.00%	BinaryClassification

Detected Column Types

	Numeric	Categorical	Text	Datetime	Sequence
Column Count	16	2	1	0	0
Percentage	84.21%	10.53%	5.26%	0.00%	0.00%

Report Contents

1. Target Analysis

目次

1. 機械学習は突然に

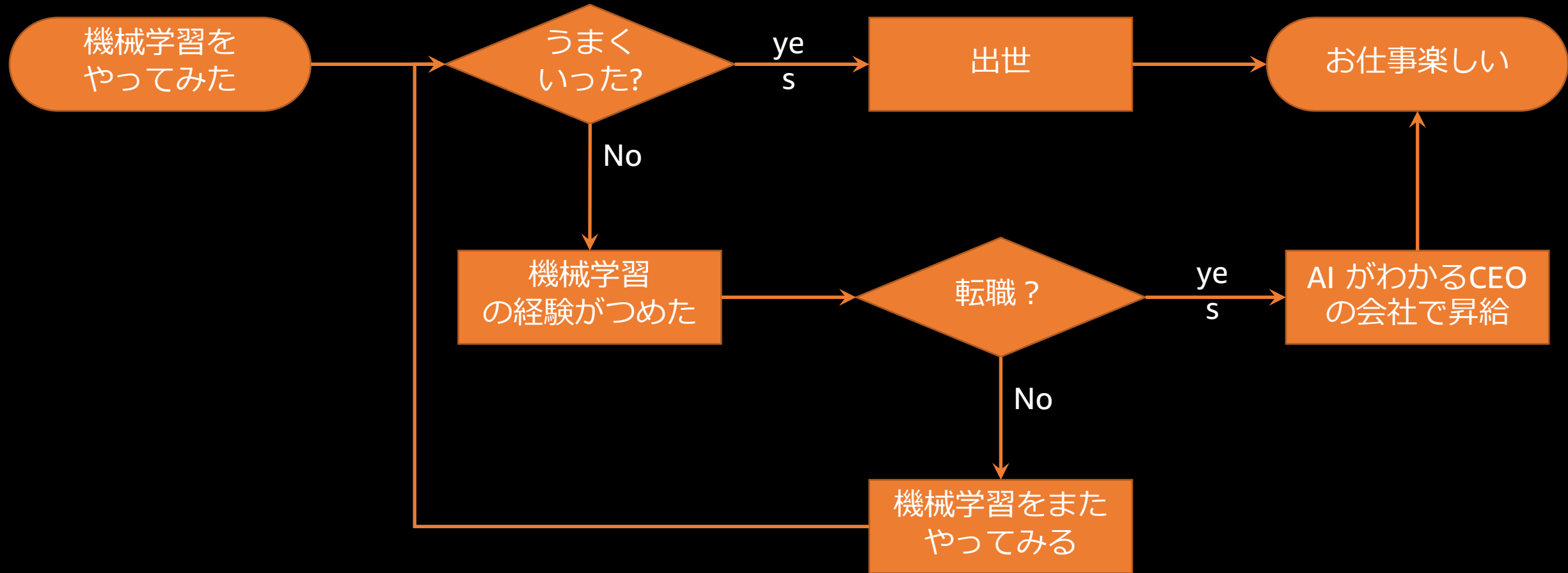
2. 機械学習、その前に

3. 私も機械学習ひとりでできるもん

4. 完全に理解した私とその後

幸せな未来へ

どのルートを通ってもよいので、まずは手を動かして試みるのが大切



本セッションのまとめ

- ML comes to you unexpectedly. Be prepared.
 - 特にデータを取得する経路は早めに確保しておきましょう
- 機械学習は目的ではなく、ビジネス課題を解くための1手段という観点を忘れずに
- 機械学習をやったことがなくても Amazon SageMaker Canvas がある程度肩代わりしてくれるのでまずはやってみましょう

More Info

- 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>

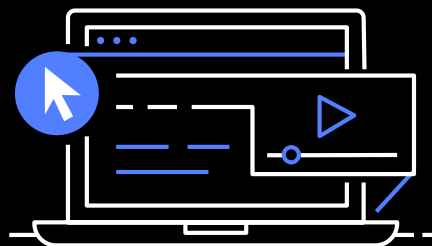


- (Dark Part) 機械学習モデルの開発や運用をマネージドサービスで効率的に行う方法
<https://bit.ly/3927PCN>



- Amazon SageMaker Canvas の Documents
<https://docs.aws.amazon.com/sagemaker/latest/dg/canvas.html>

AWS デジタルトレーニング



実力、自信、信頼性を
高め、業界で認められ
た資格で差をつけよう

デジタル学習

- [スキルビルダー](#) – AWS のエキスパートが開発した数百のデジタルトレーニングを自分のスケジュールで学習できます
- [Cloud Quest](#) - AWS Cloud Quest は、実践的なクラウド経験を積み、AWSクラウドのスキルを身につけることができる、初めてで唯一のロールプレイングゲームです

認定試験準備ためのリソース

- [Cloud Practitioner](#) - AWS Certified Cloud Practitioner 取得に役立つリソースをご紹介します
- [Developer – Associate](#) – AWS Certified Developer – Associate 取得に役立つリソースをご紹介します

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がございましたら、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!