

今こそデータドリブンなDX推進を！ 成功する機械学習/AIプロジェクトの進め方

株式会社DATAFLUCT
CTO 原田 一樹

DATAFLUCT



原田 一樹

株式会社DATAFLUCT
CTO / プラットフォーム事業技術責任者

伊藤忠テクノソリューションズに入社後、先進技術実証・デジタルビジネスを推進する組織で、アプリ開発・クラウド・AI・IoT技術の実証実験に幅広く携わってきたフルスタックエンジニア。

技術コミュニティの牽引経験や数多くの技術開発コンテストの受賞経験を持つ。

2016年6月、日本マイクロソフトにコンサルタントとして参画後、インフラ案件・AIアプリ開発・IoT/BigDataアーキテクチャ策定等のプロジェクトを推進。その後テクニカルトレーナーとして、同社クラウド技術を全方位にコーチング。

2020年11月、DATAFLUCTのCTOとして参画し、現在はノーコード機械学習プラットフォーム「Comler」のリードアーキテクト、複数のDXプロジェクトのテクニカルアドバイザーとしてデータ活用ビジネス創出をビジネス面・技術面でリード。

DATAFLUCT

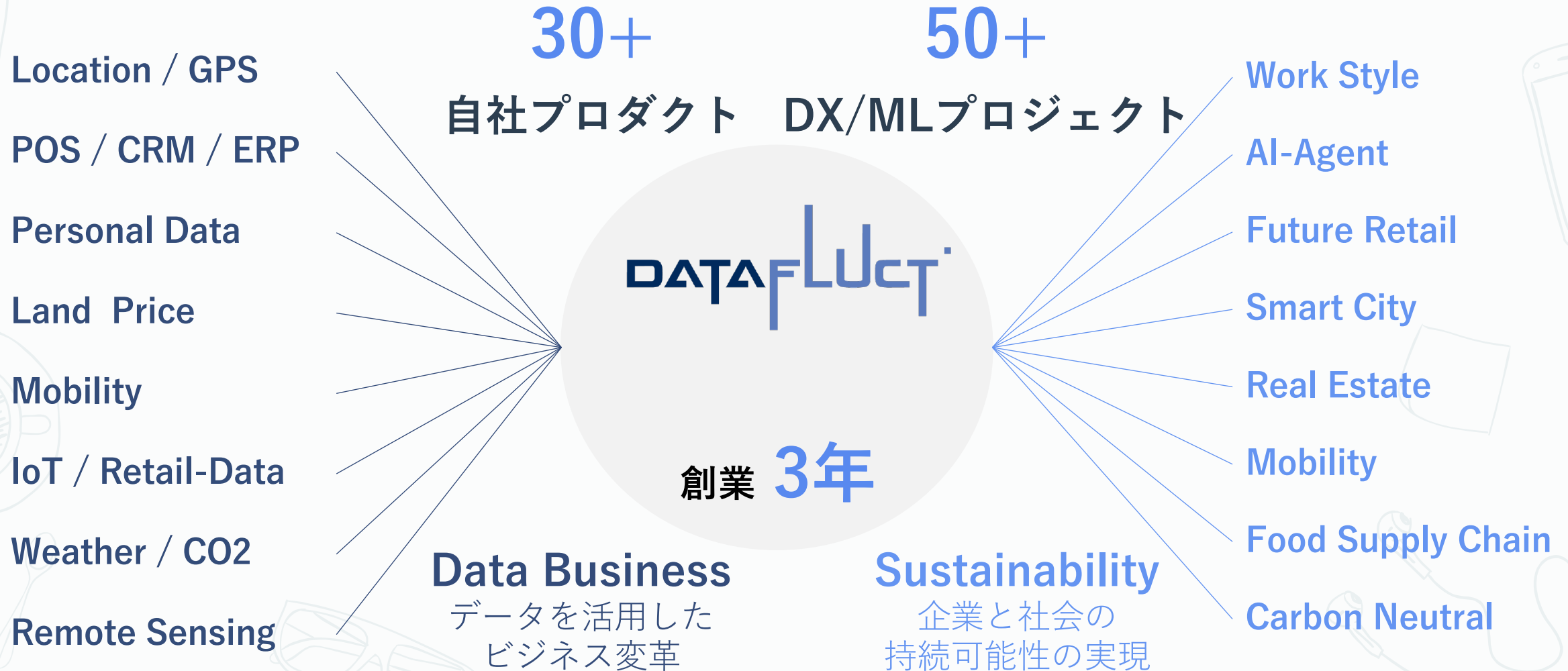
会社概要

会社名	株式会社DATAFLUCT (データフラクト)
資本金	4億4,712万円（資本準備金含む）
設立日	2019年1月29日
住所	東京都渋谷区道玄坂一丁目19番9号 第一暁ビル6階
代表	代表取締役 久米村隼人
従業員数	24名（業務委託合わせると約155名）

沿革

2019年1月	DATAFLUCT創業
2019年6月	JAXA ベンチャー認定
2019年8月	増資、取締役CFOに細田 雄吾氏が就任
2019年10月	エンジェルラウンドで4,600万円調達
2020年9月	デットとエクイティで約6,000万円の追加調達
2020年10月	社員第一号採用
2020年11月	CTOに原田 一樹氏が就任
2021年3月	シリーズAでUTECから3億円の資金調達
2021年11月	日本公庫からの融資により2.5億円の資金調達

“データ活用の民主化”で 社会と企業のサステナビリティを可能にするスタートアップ



DATAFLUCT とは

モビリティ領域



2019年12月

都市監視/リテール領域



2020年2月

データ分析技術領域

TOSHIBA

2020年6月

スマートシティ/不動産テック領域



2020年8月

食品廃棄ロス領域



2020年10月

スマートシティ領域



2021年2月

FinTech領域



2021年2月

デジタルクローン

TOSHIBA

2021年7月

FinTech / Insurtech



2021年8月

機械学習プラットフォーム



2021年11月

多業界・多分野でのデータ活用事業創出するプロ集団
10のアクセラレーションプログラム・Awardに採択

DXを成功させるには？



DX とは

デジタル技術

AI

5G

AR/MR/VR

BigData

...



目的：
競争優位性の構築

ビジネス変革

ビジネスモデル変革

オペレーション変革

組織変革

業界変革

...

DXとはデジタル技術を活用して、ビジネスを変革すること



DXの本質とは（＝DATAFLUCTの解）

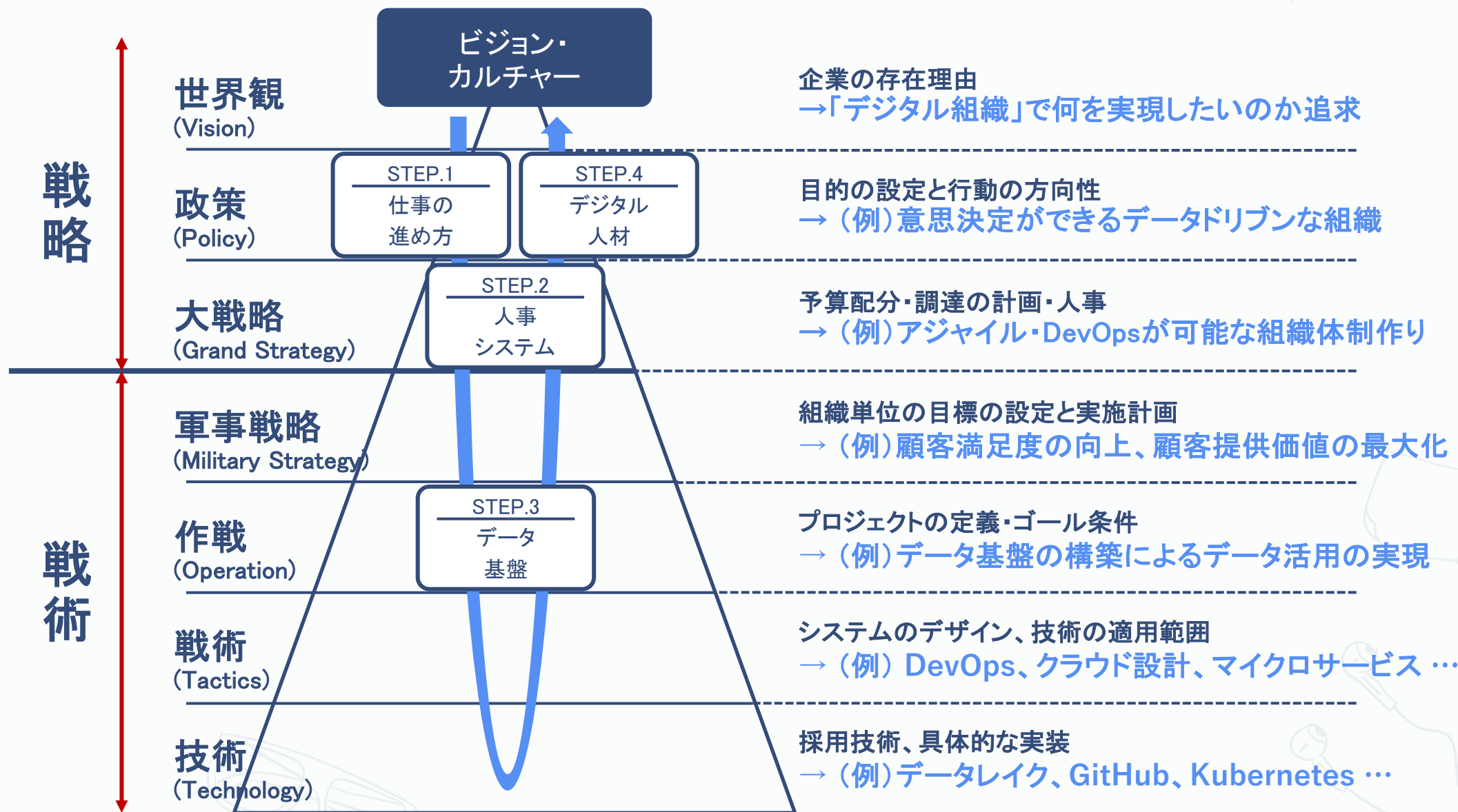
D(B)X

II

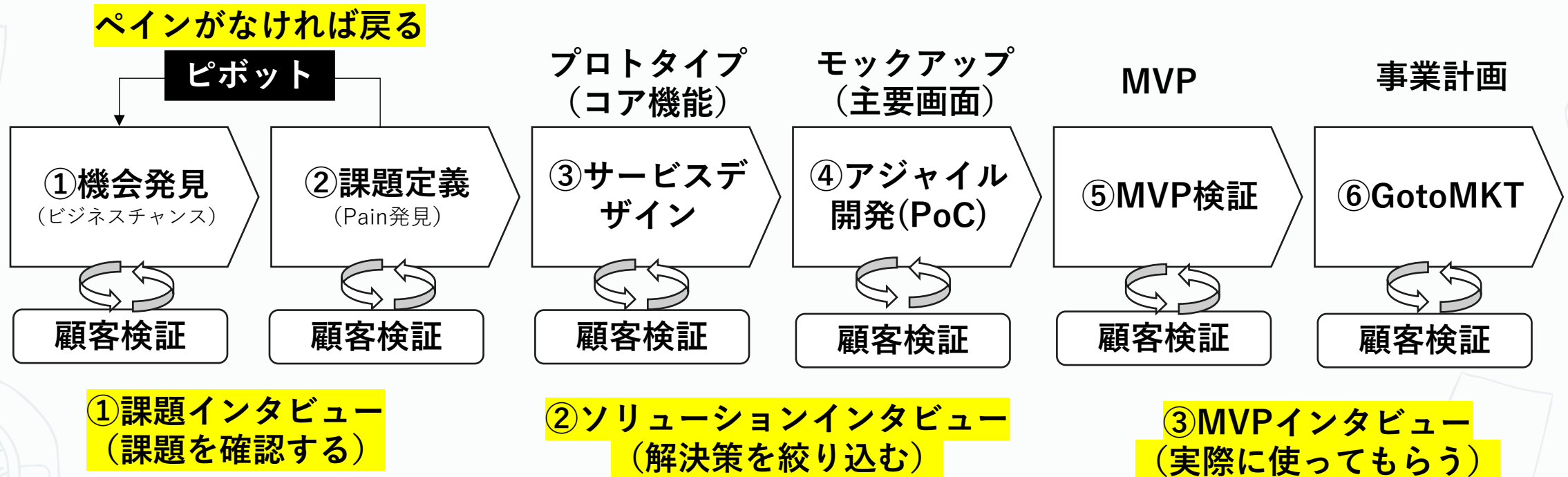
Digital **Business** Transformation

デジタル活用は手段で、ビジネス変革が真の目的

DX推進のための戦略ピラミッド



成功するDX推進アプローチ：「顧客検証×アジャイル開発」



- どのような状況の人の課題を解決するのか（顧客課題の仮説）
- 顧客課題に対する必要な機能は何か（コア価値の仮説）
- 正しくない仮説を採用しないように「検証・データに基づく意思決定」を前提に。

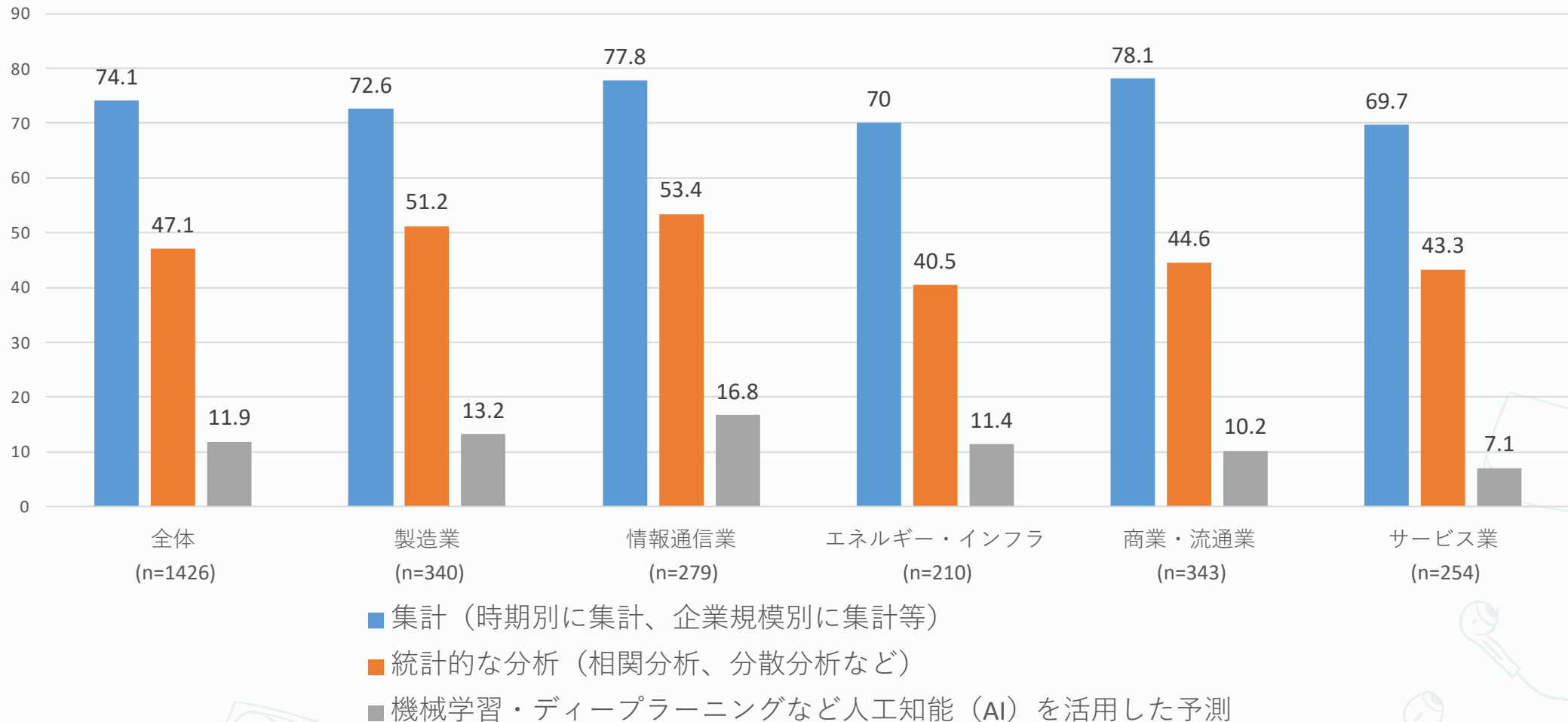
DXを成功させるには、
データドリブンな
意思決定が不可欠



日本国内におけるデータ活用の現状

統計的な分析 = **47.1 %**

機械学習/AIによる予測型の分析 = **11.9%**



(出典)総務省(2020)「デジタルデータの経済的価値の計測と活用の現状に関する調査研究」

DX推進を阻むデータ活用の「壁」

統計的データ
活用レベル

Level 5
全社レベルで
高度な分析

Level 4
横断的な
データ活用

Level 3
データ基盤を
整備済

Level 2
特定の
部署のみ

Level 1
ほぼできて
いない

非デジタル
企業

デジタル
企業

GAFA
レベル

デジタル人材は日本の
就業者人口の**約12.3%**

(出典：デロイト トーマツ
「デジタル人材思考性調査」)

Level 1
聞いたことはある

Level 2
概要は知っていて
関心がある

Level 3
既に検証したが、
課題がある

Level 4
一部のチームで
活用できている

Level 5
全社的に
体制あり

機械学習
活用レベル



4% の機械学習モデルのみ、
本番運用にたどりつくことができる
(Dimensional Research)

機械学習プロジェクトが抱える多くの課題

ノウハウ 不足

×

スキル 不足

Problem1

×

データ準備・データ理解

連携 不足

- ・データ準備・データ理解に時間がかかる
- ・社内のデータにアクセスできない
- ・BIツールをうまく使いこなせない
- ・部署の壁で内部データを収集できない
- ・データ基盤をうまく構築できない



Problem2

モデル構築・モデル評価

- ・データサイエンティストがいない
- ・人件費や外注費が高価
- ・適切なアルゴリズムの選択が難しい
- ・チューニングが難しい
- ・予測精度が実用レベルに到達しない

低い成功確率

Problem3

モデル展開・運用管理

- ・モデル運用基盤を構築する知見がない
- ・モデル再学習に都度データエンジニアやデータサイエンティストに依頼
- ・モデルを利用するために必要なパラメーターが多すぎて使いにくい

機械学習/AIプロジェクトの進め方

機械学習が最適な
解決手段か判断

問題・要求の定義

ビジネス上の効果測定し、
新たな問題・要求を定義

ビジネス
目標の評価

機械学習ライフサイクル

精度監視と再学習を継続的に行う（MLOps）

十分な精度になるまで繰り返す（アルゴリズム選定、チューニング…）

データ準備

モデル構築

モデル評価

モデル展開

モデルの
運用管理

- ・データインポート
- ・データ定義
- ・環境構築
- ・データ加工
- ・欠損値対応
- ・異常値対応
- ・変数確認

- ・ライブラリの準備
- ・モデル設定
- ・チューニング
- ・アルゴリズム予測

- ・モデル指標選択
- ・スコアモデル

- ・環境構築
- ・実装
- ・コーディング
- ・テスト

- ・安定稼働
- ・モデルの精度監視
- ・モデルの再学習
- ・モデルの再配置



データエンジニア



プロジェクトリーダー・
マネージャー



データサイエンティスト



運用エンジニア



ビジネスユーザー
意思決定者



アプリエンジニア



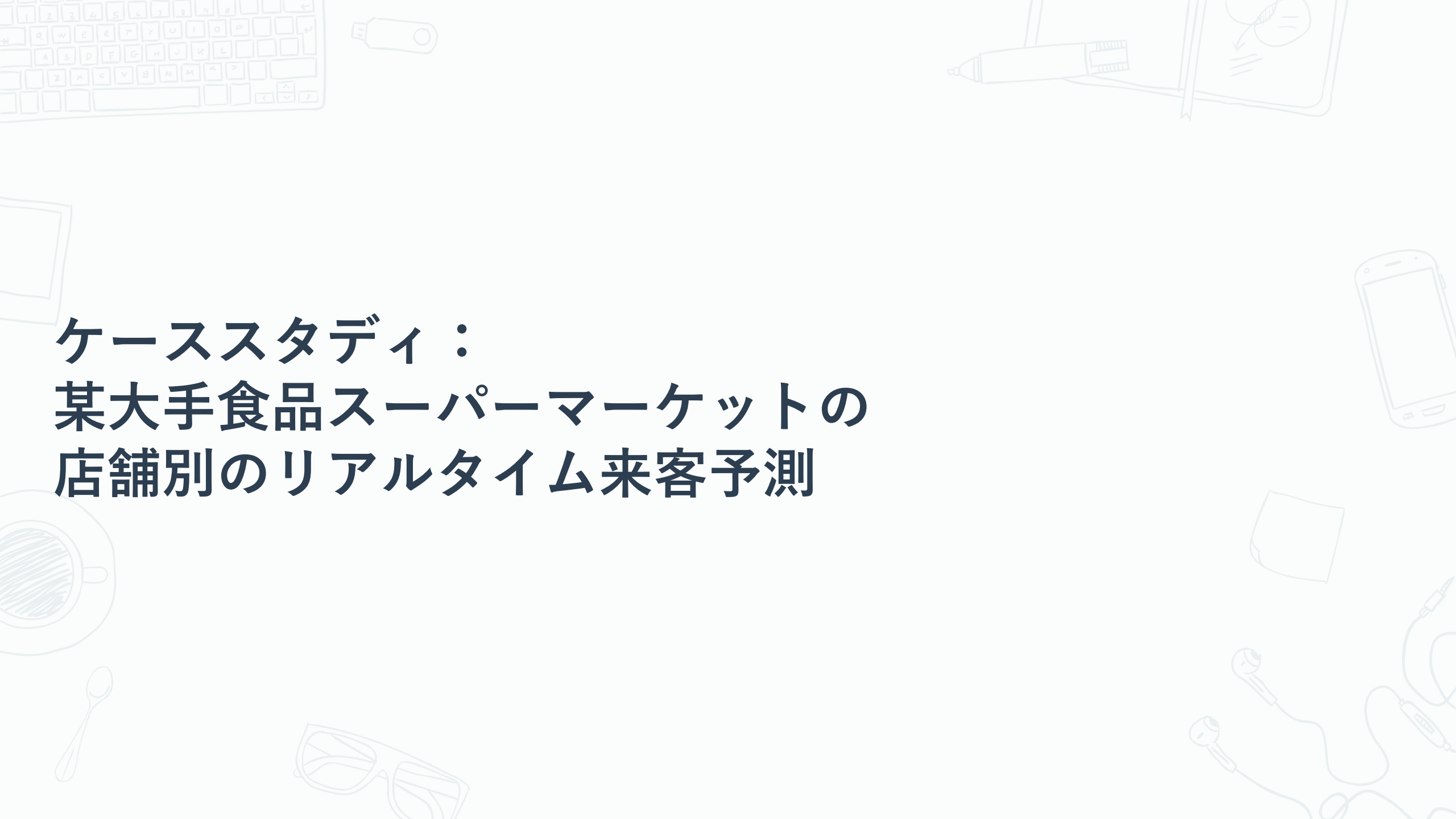
業務担当者

問題・要求の定義を正しく行うことが重要

問題・要求の定義

以下の事柄が明確・共有されている状態が望ましい

- どのような問題を解決する必要があるか？
- ビジネス上で測定すべき事象は何か？（品質、コスト、NPS、CS …etc）
- 機械学習は適切なアプローチか？（分類、予測、クラスタリング…etc）
- どのようなデータを利用できるか？（利用可能なデータで十分か？）
- 何を目標とするか？（ビジネス目標、顧客価値の最大化、予測精度…etc）



ケーススタディ： 某大手食品スーパーマーケットの 店舗別のリアルタイム来客予測

プロジェクトの目的：食品廃棄物の発生量を半減させよ！

悩みを抱えている人

データ カツヒロ
出板 克洋さん（42歳）

- ・現在、大手食品スーパーマーケットの本部で発注業務を担当。
- ・食品廃棄物の発生量を数年以内に半減させるプロジェクトにアサイン中。
各店舗のID-POSデータを分析し、日々対策を検討している。
- ・その一方で、惣菜市場の成長に合わせた売上向上施策の立案も期待されており、両立に頭を悩ませている。



どのような問題を解決する必要があるか？

- └ 各店舗で発生している食品廃棄物の発生量を50%削減と惣菜市場における売り上げ20%アップ

ビジネス上で測定すべき事象は何か？（品質、コスト、顧客満足度…etc）

- └ 各店舗の生鮮食品・惣菜の「販売数」「売り上げ」「値引きした商品数」「廃棄した商品数」

機械学習は適切なアプローチか？（分類、予測、クラスタリング…etc）

- └ 時系列予測により、1時間単位の来客数予測する（ID-POSは導入済）

どのようなデータを利用できるか？（利用可能なデータで十分か？）

- └ 店舗毎のID-POSデータ、気象データ、地域イベント情報、人流データ…

何を目標とするか？（ビジネス目標、顧客価値の最大化、予測精度…etc）

- └ 当日内の来客数予測精度90%以上

(補足) 本質的な目標設定をすることが重要

→精度よりも重要視したいポイントを正しく共通認識化する



店舗Aの店長

予測精度: 89%
(2日後の来客数)



店舗Bの店長

予測精度: 94%
(2日後の来客数)



店舗Cの店長

予測精度: 84%
(2日後の来客数)



店舗Dの店長

予測精度: 91%
(2日後の来客数)

- 大事なのは本当に予測精度？
- 上振れ予測と下振れ予測はどちらが困る？
- 店舗毎のバラつきが解消される方が良いのでは？

Step1. データ準備

まずはすぐに利用可能なID-POSデータを利用して来客数の予測モデルの作成に着手
特性の異なる以下の店舗を対象に、ID-POSのデータから時間帯別の来客数情報を抽出

■ 店舗データ（店舗ごとの特性情報を含むデータ）

店舗	気象観測の位置	常住人口	昼間人口比率	駅からの距離	客タイプ	店舗面積
店舗1	東京	多い	増	普通	近所の住民、高級志向	中
店舗2	東京	普通	増	近い	若い主婦、学生	小
店舗3	東京	普通	減	普通	主婦	中
店舗4	東京	少ない	減	遠い	近所のお年寄り、買い溜め	大
店舗5	大阪	普通	減	普通	主婦	中

■ ID-POS データ（2021年2月の1ヶ月分のデータを時間帯別に抽出）

date	time	customerID	storeID	itemCategory	itemName	price
2021/2/16	9	294	1001	6	非食品3	200
2021/2/16	9	294	1001	3	惣菜4	360
2021/2/16	9	294	1001	5	一般食品5	600
2021/2/16	9	294	1001	5	一般食品5	600
2021/2/16	9	294	1001	4	日配品4	130

Step2. モデル構築

データ準備

モデル構築

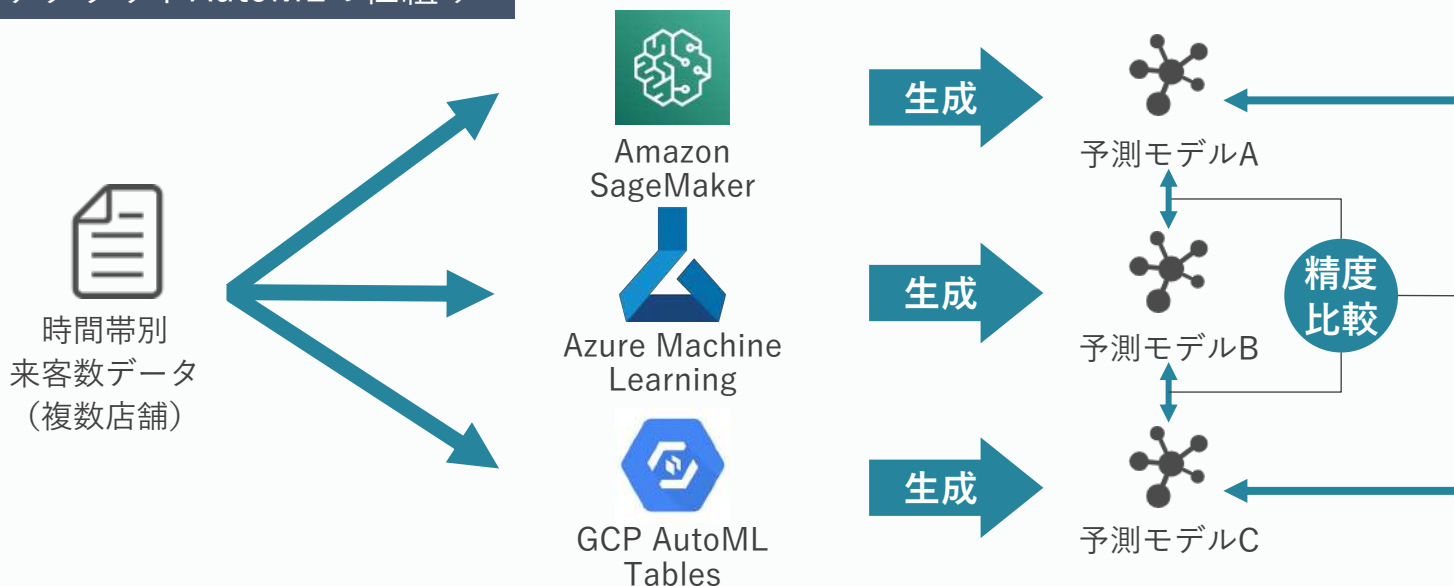
モデル評価

モデル展開

モデルの
運用管理・更新

マルチクラウドAutoMLサービス『Comler for AutoML』に
データセットをアップロードし、来客数を予測するモデルを作成

Comler – マルチクラウドAutoMLの仕組み



作成モデルを使って“時間帯別の来客数を予測した結果”と“実際の来客数データ”を比較してみた結果、
人の経験則と比べても見劣りするレベルの精度（**予測誤差：±30%～50%**）に。

なぜ、このような結果になってしまったのか？

Step3. モデル評価

データ準備

モデル構築

モデル評価

モデル展開

モデルの
運用管理・更新

予測精度が低かった理由について考察し、精度改善に向けた対策を検討

担当者の考察

- 1日単位・1週間単位で分析した場合、予測誤差は $\pm 10\%$ 程度だった。
- 時間帯別の来客数は平日でも曜日によって異なる上、店舗によっても傾向が異なる。
- 人の店舗前の通行量は、天候によって変動する。
- イベントやSNSなどの外部要因による影響も考慮した方がよい。

データサイエンティストからの助言

- データ量と特性が異なる店舗データ混ぜてしまったことによって、教師データに分散と偏りが生じている可能性がある。
- 東京の店舗のデータ量が多すぎるため、大阪の店舗の予測が影響を受けている。
- 同じ年代の顧客でも駅近の店舗の方が人数が多いため、郊外の店舗の予測に悪影響が出ている。
- 2021年2月のデータで作成したモデルは、季節性の違いや新型コロナなどの社会情勢の違いによって、3ヶ月後、半年後、1年後には更に精度が下がる可能性が考えられる。

機械学習の予測精度を高く維持し続けるためには、以下の対応が必要に。

- 「自社のデータと外部のデータを組み合わせた必要なデータの準備」
- 「季節性・社会性を常時反映するための継続的なデータの収集・加工・蓄積できるデータ基盤」

Step1. データ準備（改善）

データ構造や更新頻度が異なるデータをデータ基盤に集約し、
店舗ごとに最適なモデルを構築できるようにデータを整形

データ集約・データ整形

自社データ

ID-POS、店舗データ

外部データ

気象データ、人流解析データ

店舗データ
date
store ID
area code
address
store area
...

・店舗の特性データ

気象データ
date
address
temperature
rain
...

・店舗周辺の天気

人流データ
date
day of week
area code
population
...

・店舗周辺の流動人口

ID-POS
date
store ID
customer ID
Item Category
price
...

・店舗の顧客ごとの
購買データ加工・
蓄積教師データ
(整形済)予測用データ
(整形済)

Step2. モデル構築/評価（改善）

データ準備

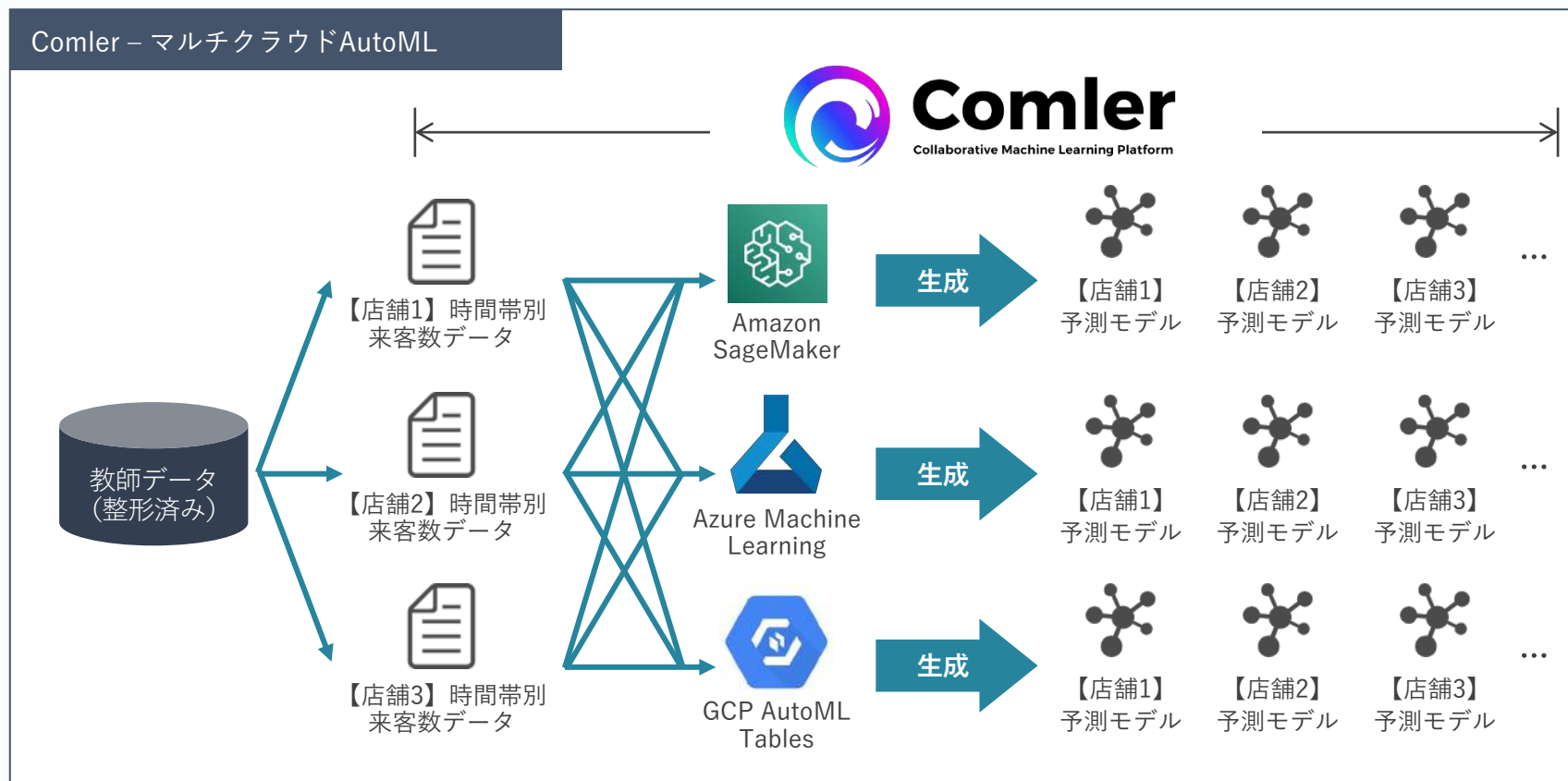
モデル構築

モデル評価

モデル展開

モデルの
運用管理・更新

データ基盤に用意した店舗ごとに
整形済のデータセットを使い、AutoMLで予測モデルを作成



各店舗毎に最適化したモデルを用意することで、当日内の来客数予測精度90%以上を達成する精度に

Step3. モデル展開・運用管理

データ準備

モデル構築

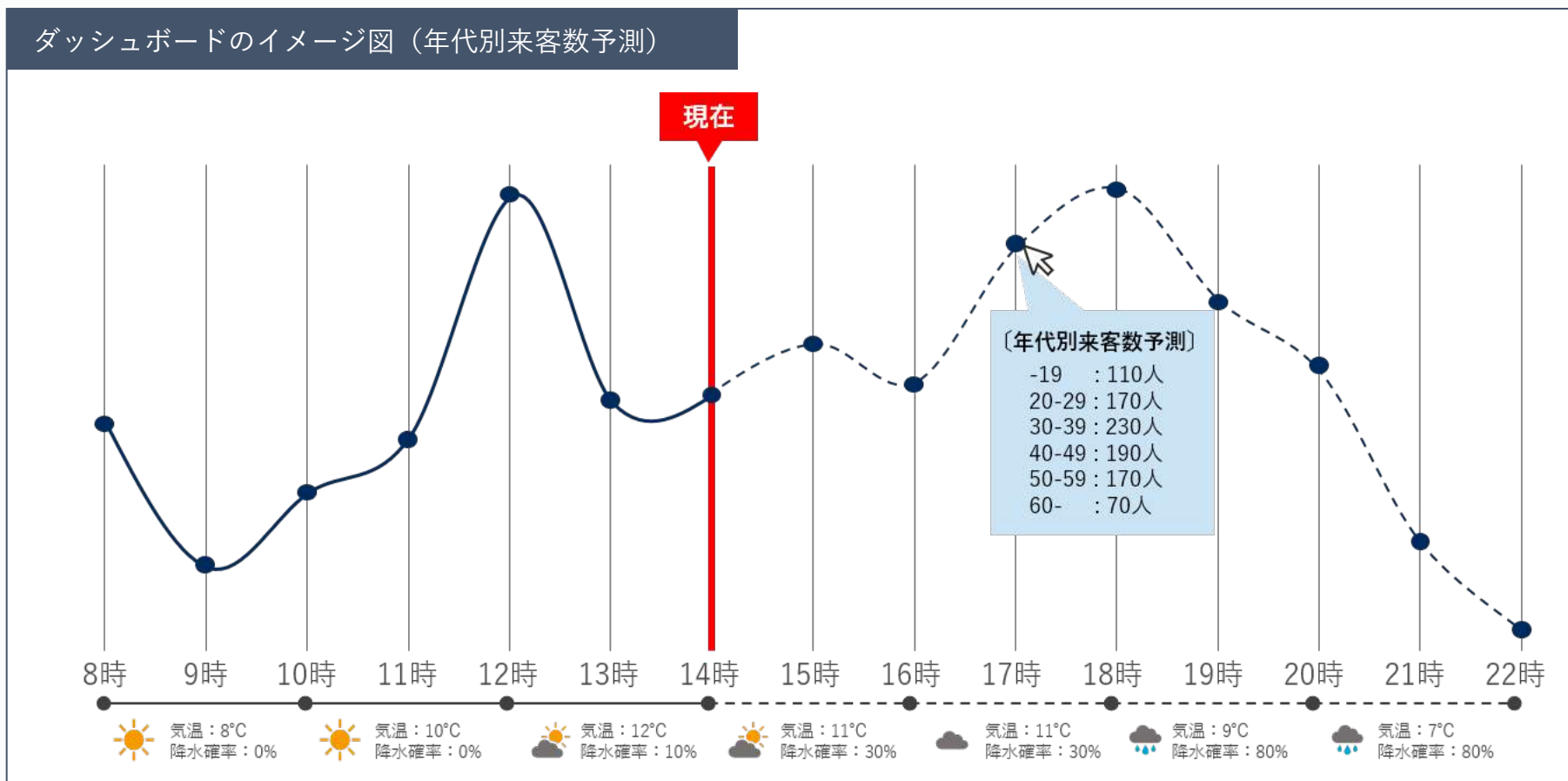
モデル評価

モデル展開

モデルの
運用管理・更新

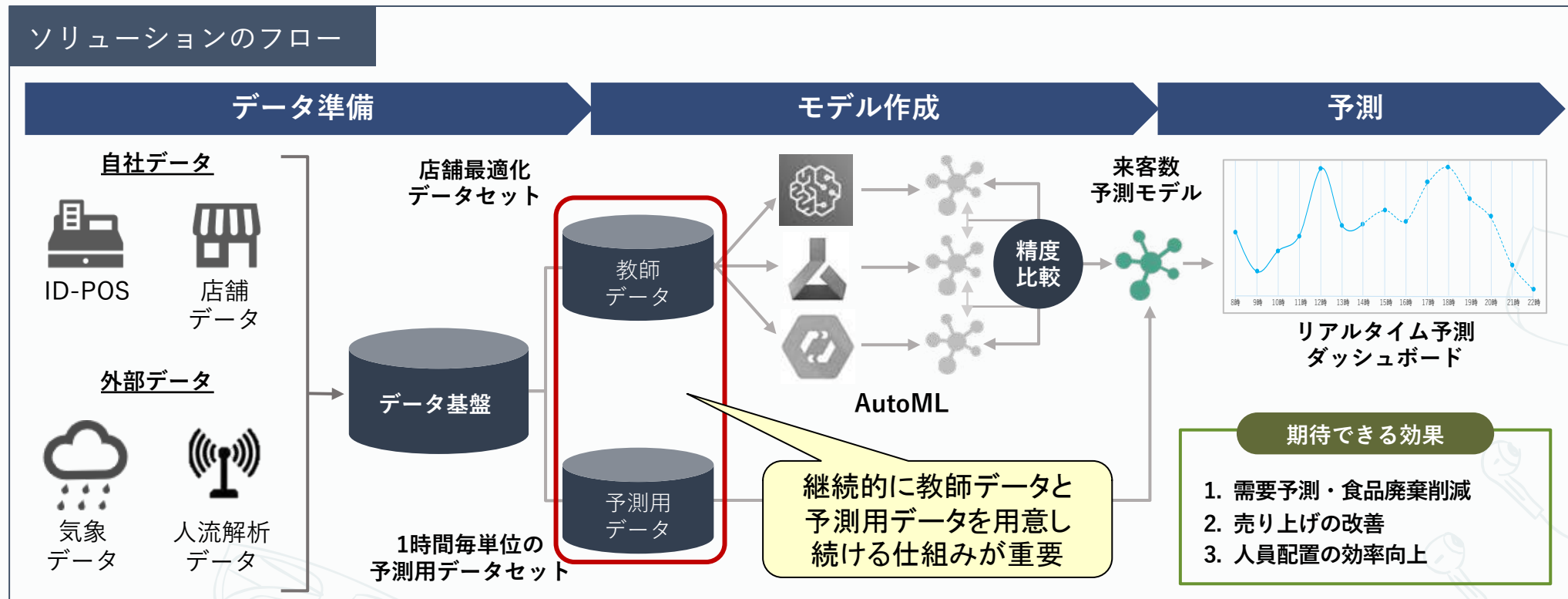
店舗毎に最適化したモデルをアプリケーションから利用できるようにデプロイ。
年代別来客数の予測結果をリアルタイムに表示するダッシュボードを用意し、
惣菜品の調理タイミングや棚出し計画のリアルタイムな意思決定が改善に成功。

ダッシュボードのイメージ図（年代別来客数予測）



本ケーススタディのポイント

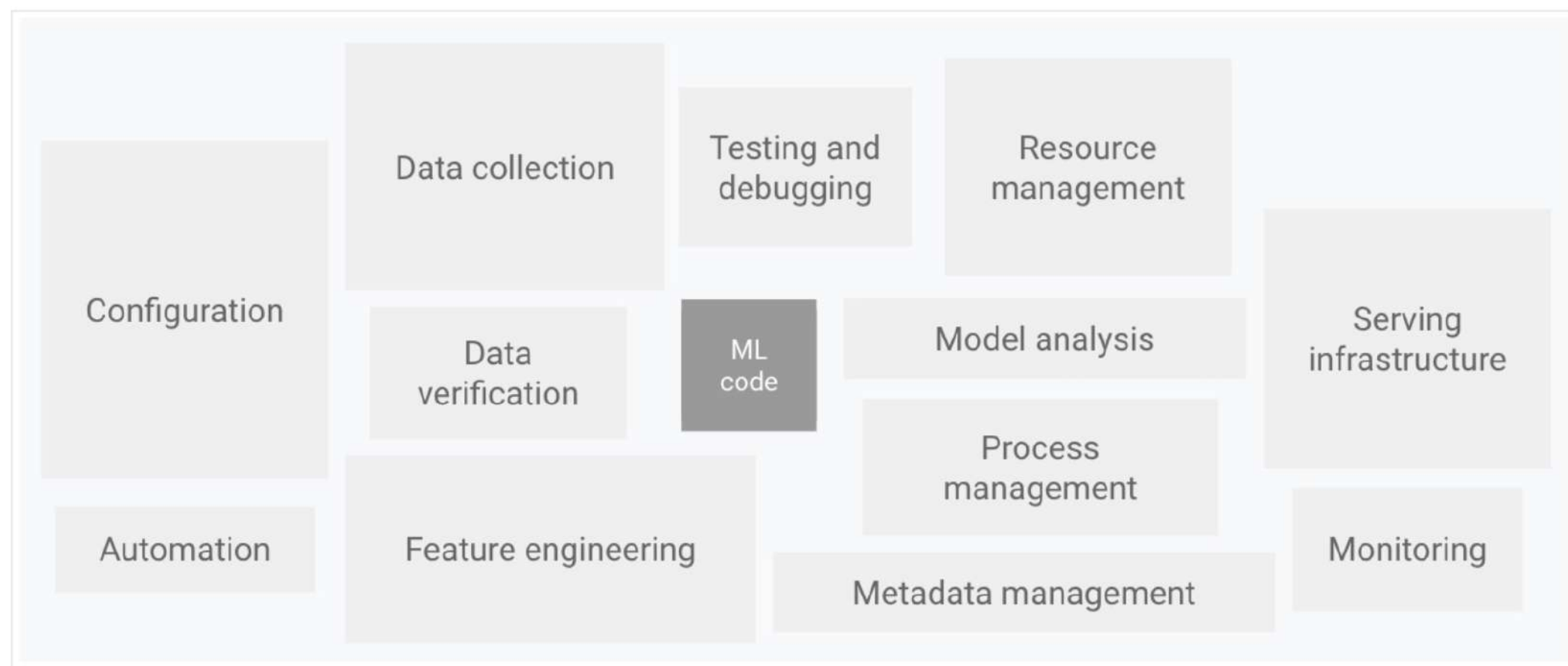
- ・ 各店舗に導入しているID-POSデータを集約するデータ基盤を整備
- ・ 外部要因を考慮するために外部データ（気象データ、人流データ）をリアルタイムに収集・蓄積・加工
- ・ 収集した店舗別データと外部データを基に、店舗ごとに最適な予測モデルを構築



(参考) 機械学習の本番運用の難しさ

本番環境で継続的モデル更新を行うには、MLコード以外の周辺機能を含む全体の設計が必要

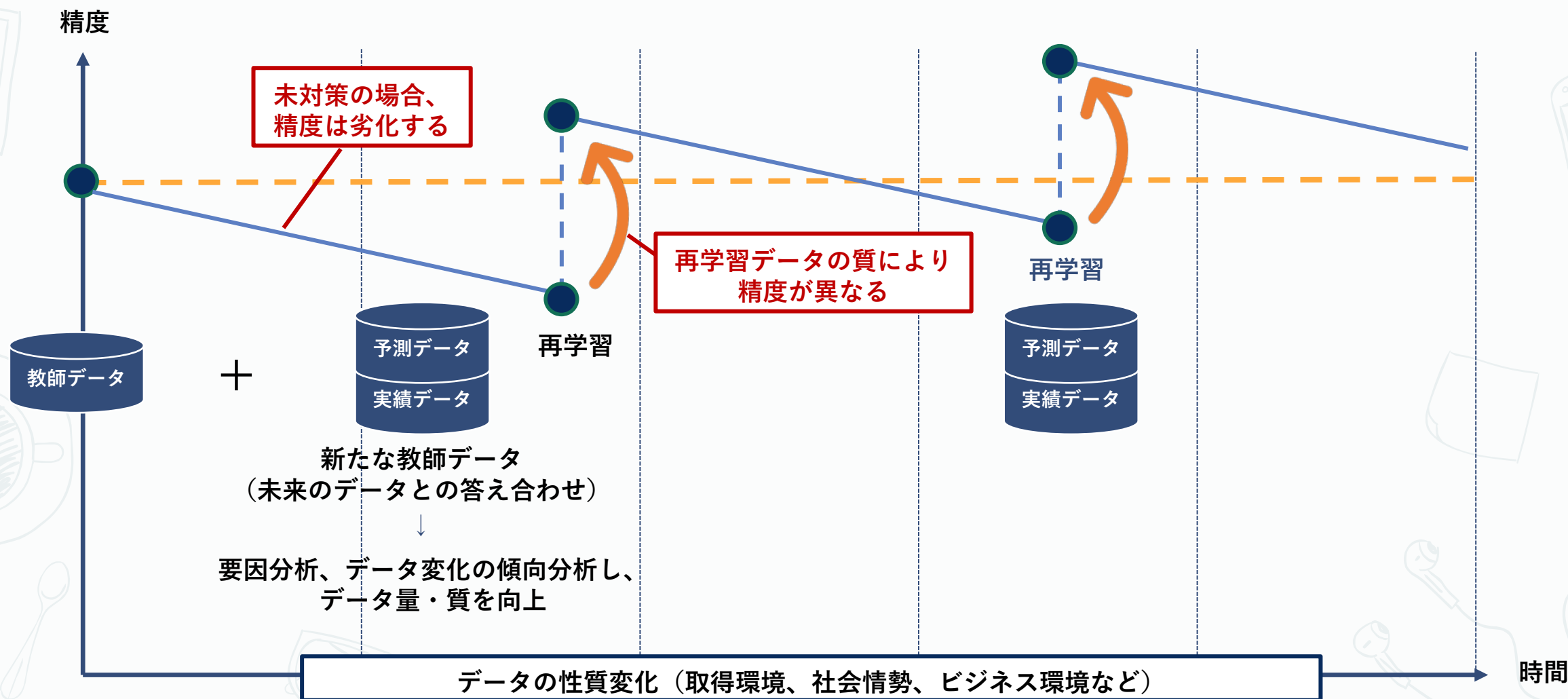
(構成管理、自動化、データ収集、データ検証、テストとデバッグ、リソース管理、モデル解析、プロセス管理とメタデータ管理、サービス インフラストラクチャ、モニタリング…)



出典：機械学習における継続的デリバリーと自動化のパイプライン (Google Cloud)

(参考) モデルの精度は本番デプロイ後、劣化していく

機械学習モデルは継続的に更新し続ける必要がある



機械学習プロジェクトに必要なスキル

機械学習が最適な
解決手段か判断

問題・要求の定義

ビジネス上の効果を測定し、
新たな問題・要求を定義

ビジネス
目標の評価

機械学習ライフサイクル

ビジネススキル（問題設定能力など）

精度監視と再学習を継続的に行う（MLOps）

十分な精度になるまで繰り返す（アルゴリズム選定、チューニング…）

データ準備

モデル構築

モデル評価

モデル展開

モデルの
運用管理

データサイエンススキル（ドメイン知識・高等数学・
Python・アルゴリズム・SQL・問題解決能力など）

- ・環境構築
- ・実装
- ・コーディング
- ・テスト

- ・安定稼働
- ・モデルの精度監視
- ・モデルの再学習
- ・モデルの再配置

エンジニアスキル
(MLOps + DevOps)

コミュニケーションスキル

データエンジニア

プロジェクトリーダー・
マネージャー

データサイエンティスト

運用エンジニア

ビジネスユーザー
意思決定者

アプリエンジニア

業務担当者

皆さんの会社は、
“**あらゆる組織**”で
このような体制を作り・
維持することはできますか？



“データ活用の民主化ツール”を利用し、 あらゆる組織をデータ分析組織へ。

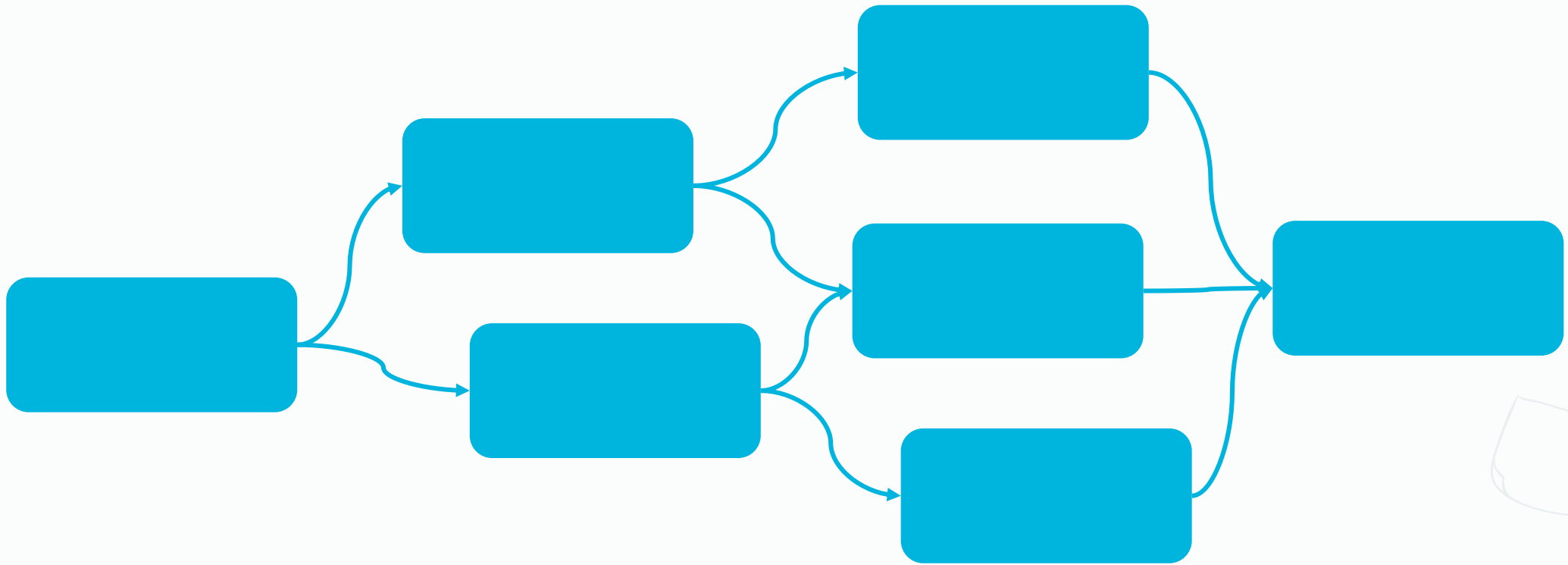




ノーコード = 民主化 ？



一般的なノーコードツール



“**ユーザーの知識を前提**”に自由に組み合わせ



ノーコード



テンプレート



自動化

データ
インサイト

自動
データ加工

ノーコード
ML開発

AutoML

モデル評価
レポート

予測シミュ
レーション

MLOps

Collaboration

データエンジニア

プロジェクトリーダー・
マネージャー

データサイエンティスト

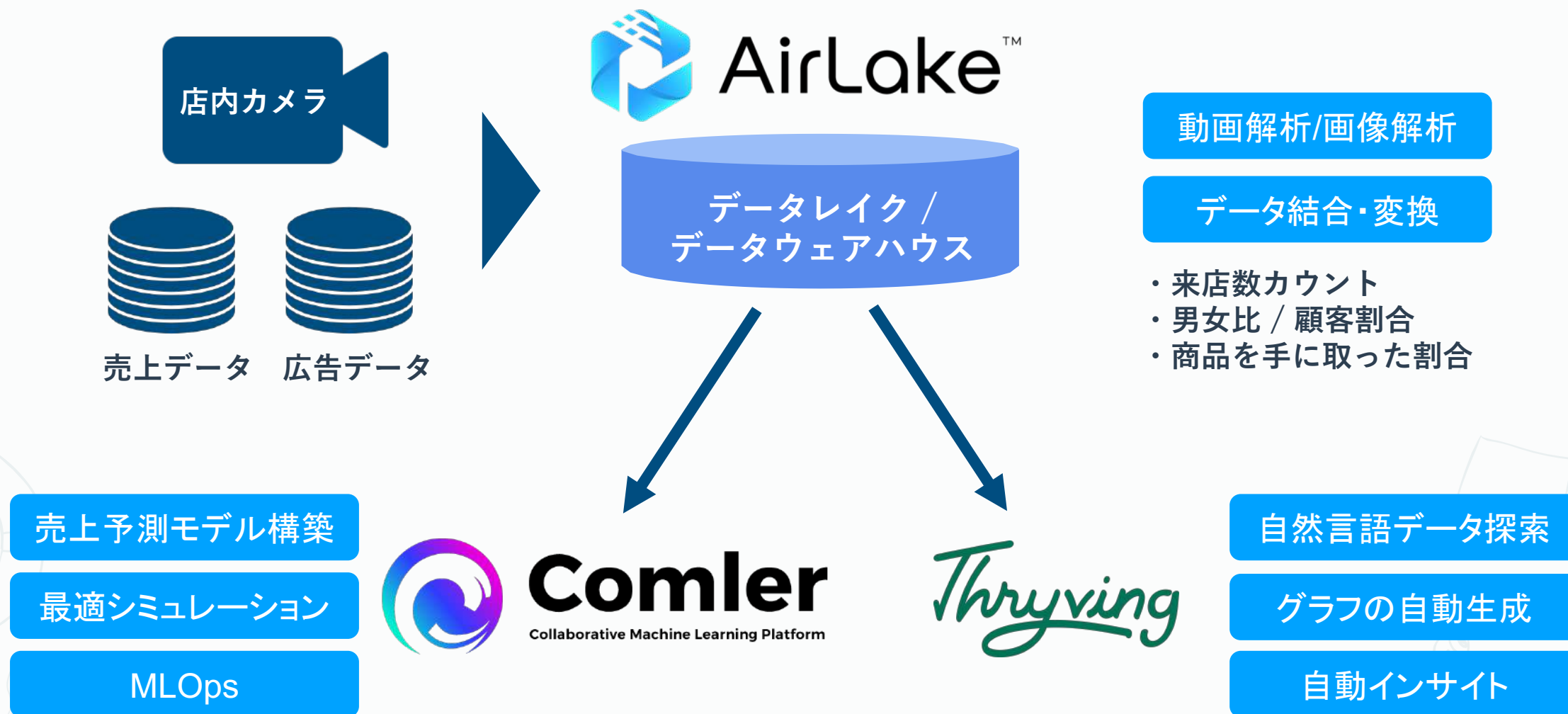
運用エンジニア

ビジネスユーザー
意思決定者

非デジタル人材でも使いこなせる真の民主化ツール

特許出願中

Demo：小売店におけるデータ活用アプローチ例



データ基盤構築、モデル構築・活用、データ可視化を専門家でなくても可能に

データ基盤構築ツール (DataLake&DWH)



AirLakeTM

AIとの対話を通じて示唆を
手に入れられるBIツール

Thryving

機械学習/AIプロジェクトを成功に導く
ノーコード機械学習プラットフォーム

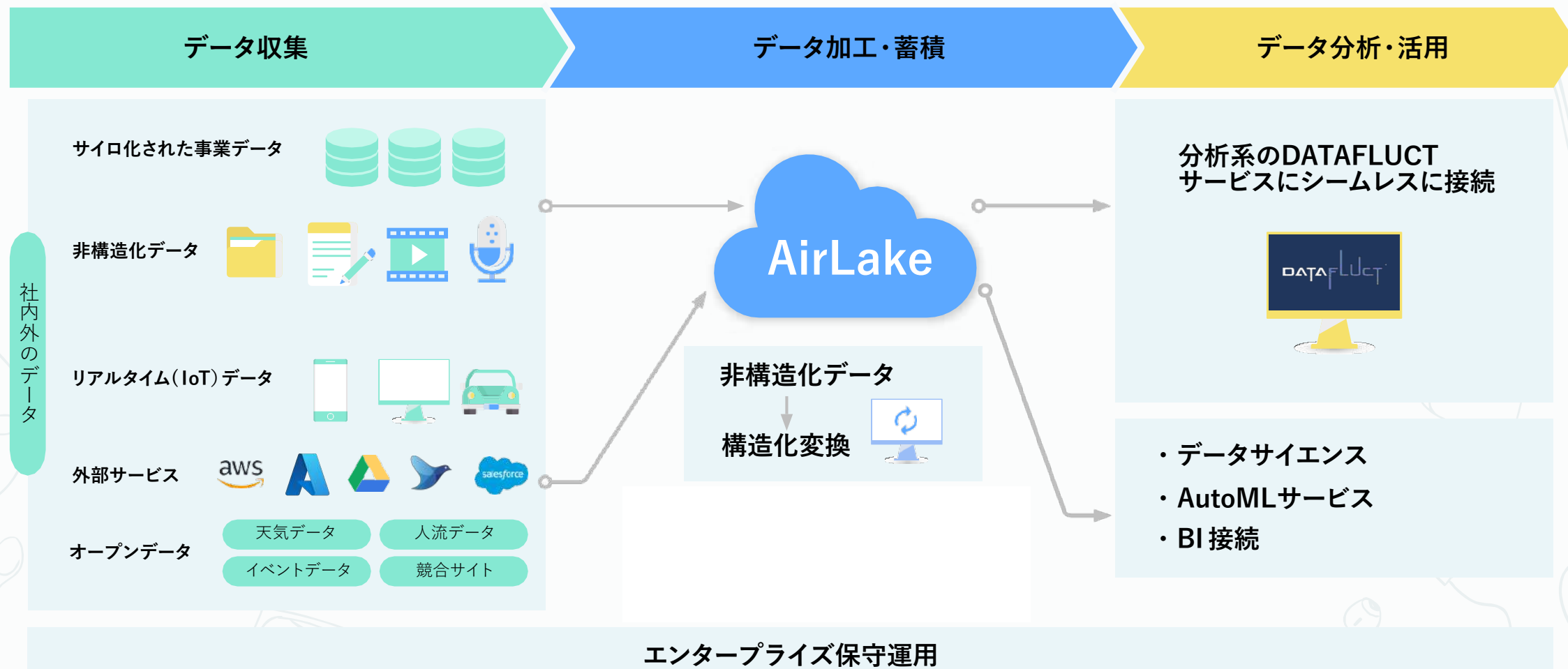


Comler

Collaborative Machine Learning Platform



社内外のあらゆるデータを分析しやすい形で蓄積



多種多様な非構造データの前処理に対応

対象データ

前処理の例



画像・動画

- ・ アノテーション（抽出対象、異常検知対象…）
- ・ メタ情報（上下幅、色、サイズ、カテゴリ…）



音声

- ・ 形式変換(フーリエ変換による周波数領域変換)
- ・ テキスト化、不要後削除



文章

- ・ クリーニング(ノイズ除去)、ストップワード除去
- ・ 単語の分割、正規化、文字種の統一…

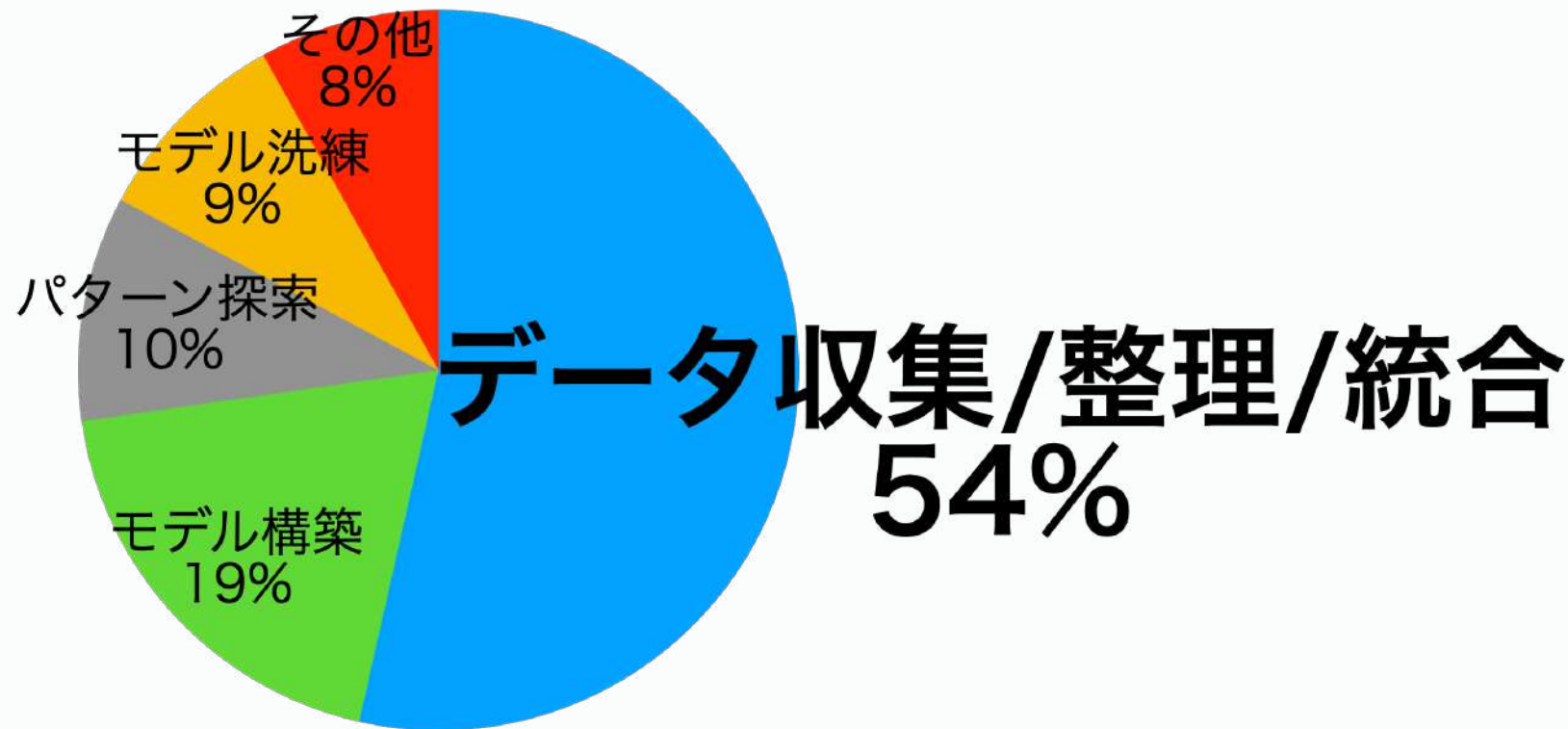


センサー

- ・ 半構造化データ(json、xml)の構造化変換
- ・ ダウンサンプリング、データの結合分割、正規化…



データ活用するための下準備に多くの手間が掛かる



データサイエンスの業務時間内訳

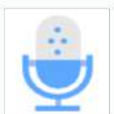


テンプレート自動変換技術で分析しやすい形へ

非構造化データ



画像・動画



音声



文章



センサー

整理されていない

動画解析/画像解析

テキストマイニング

自然言語処理 (NLP)

パーソナルデータフィルタ



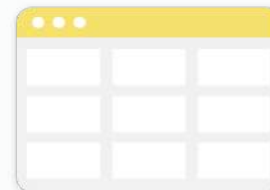
AirLake™

アノテーション付与

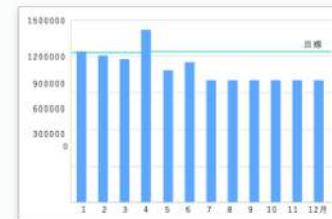
メタ情報付与

データカタログ化

構造化データ



分析しやすい

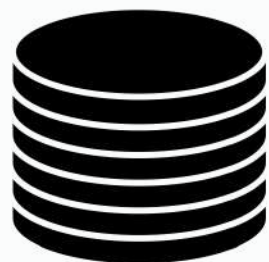


	A	B	C
1		実績	目標
2	1月	1,236,848	1,250,000
3	2月	1,203,783	1,250,000
4	3月	1,174,932	1,250,000
5	4月	1,414,519	1,250,000
6	5月	1,081,239	1,250,000
7	6月	1,148,582	1,250,000
8	7月	1,000,000	1,250,000
9	8月	1,000,000	1,250,000
10	9月	1,000,000	1,250,000
11	10月	1,000,000	1,250,000
12	11月	1,000,000	1,250,000
13	12月	1,000,000	1,250,000
14		13,259,903	15,000,000

表形式で整理

Thryving

AI技術を活用し、自分でBI構築が可能に



Thryving

AIによるデータ加工



経営指標



経営

配置分析



人事

広告効果分析



マーケ

店舗成績



営業

発注管理



業務

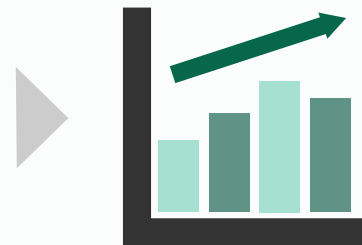


自然言語でデータを検索・可視化し、インサイトを提案

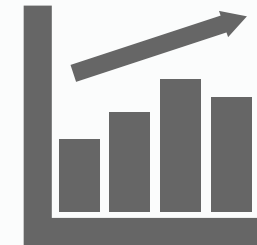


企業内外の膨大なデータの中から
必要なデータを瞬時に発見

売上
上位20
商品名



キーワードを基に
AIがグラフを自動生成



データから読み取れる
示唆をAIが自動明示



さらに深掘りすべき
データをAIが自動明示



組織内の他者のコメントや
示唆を自動でレコメンド



新たな示唆を学習し続けるAIが
見るべきデータを自動通知



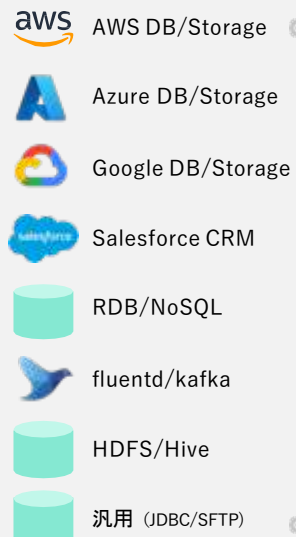
機械学習による予測モデルの構築・運用をカバー

データ接続・加工

モデル構築・評価・活用

モデル展開・運用管理

データ連携・統合 (ETL/EAI機能)



前処理

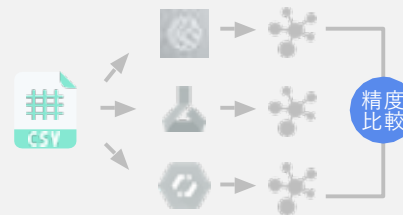
- データ加工・変換
- データ結合・分割
- 集計・フィルタリング
- 前処理・データ補完
- 正規化・標準化
- サンプリング

ノーコードML開発

- 回帰・分類
- 時系列分析
- 異常検知
- 画像解析

- 精度評価
- 特徴量分析

マルチクラウドAutoML



可視化・シミュレーション

モデル評価レポート

データインサイト

実験ノート
モデル比較

予測シミュレーション

モデルインポート・エクスポート

インポート

エクスポート

エンドポイント管理

Webリクエスト

CI/CD機能

Webレスポンス

アクセス分析

精度維持・監視

精度監視

バージョン管理

データ作成

モデル比較

自動教師データ作成

プロフェッショナルサポート

- データ準備支援
- モデル作成支援
- 本番デプロイ支援
- パイプライン構築
- ML評価支援
- 精度監視
- データ更新
- モデルの再学習
- 基盤運用

外部データ連携

- 統計情報
- 交通情報
- 各種店舗情報
- 農業・天気情報

Data Marketplace

データパイプライン/ MLOps パイプライン

Connect

Prep

Modeling

Evaluation

Deployment

AirLake (データ基盤)

**Comler**

Collaborative Machine Learning Platform

機械学習向けデータ加工を自動化 (データ整形、外れ値除去、欠損値補完…)



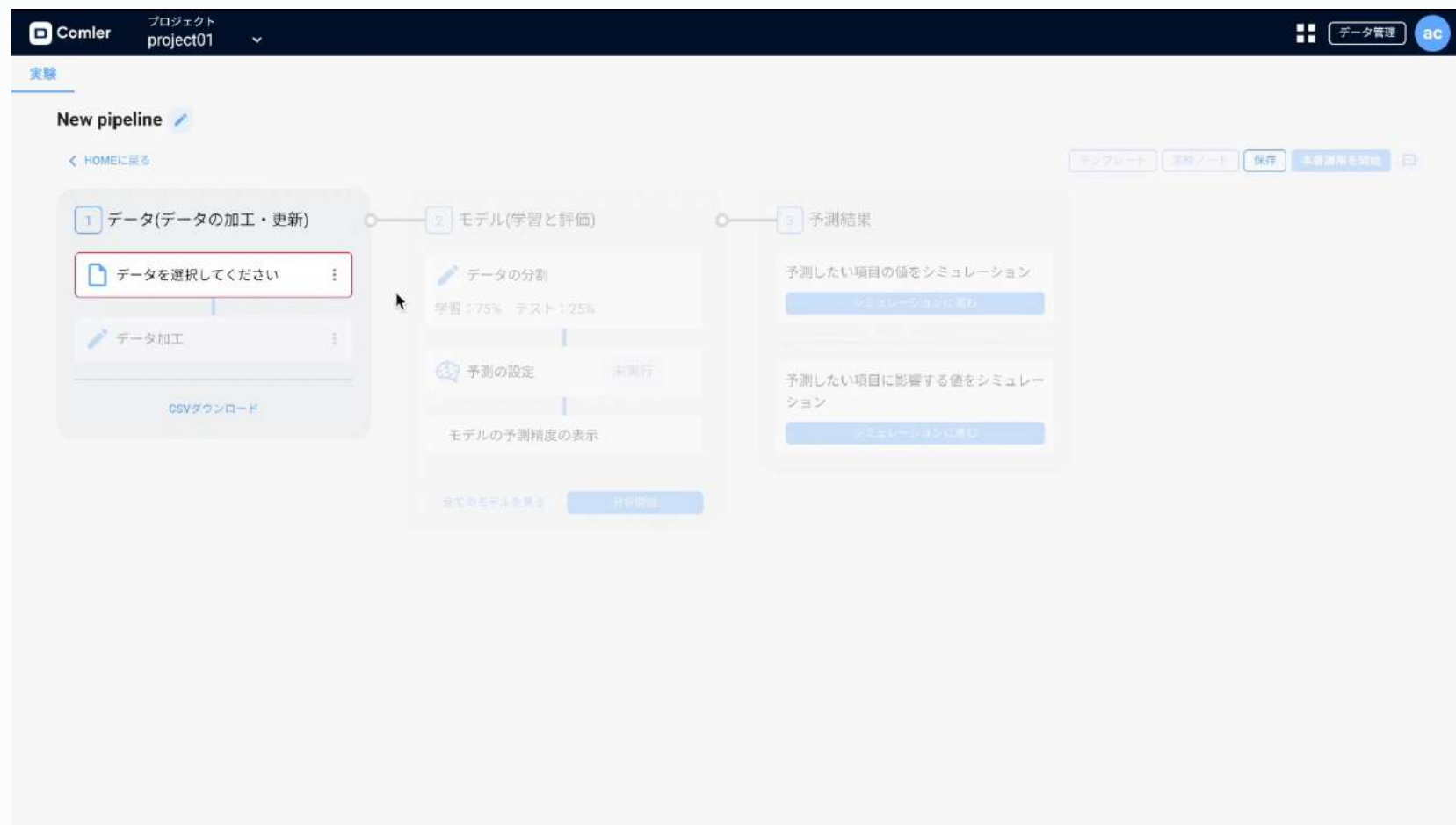
データ可視化・
インサイト機能



データ加工の
自動化



データ加工内容の
差分比較





Comler
Collaborative Machine Learning Platform

モデル構築は数クリックで完了 評価レポート・モデル比較を参考に選別



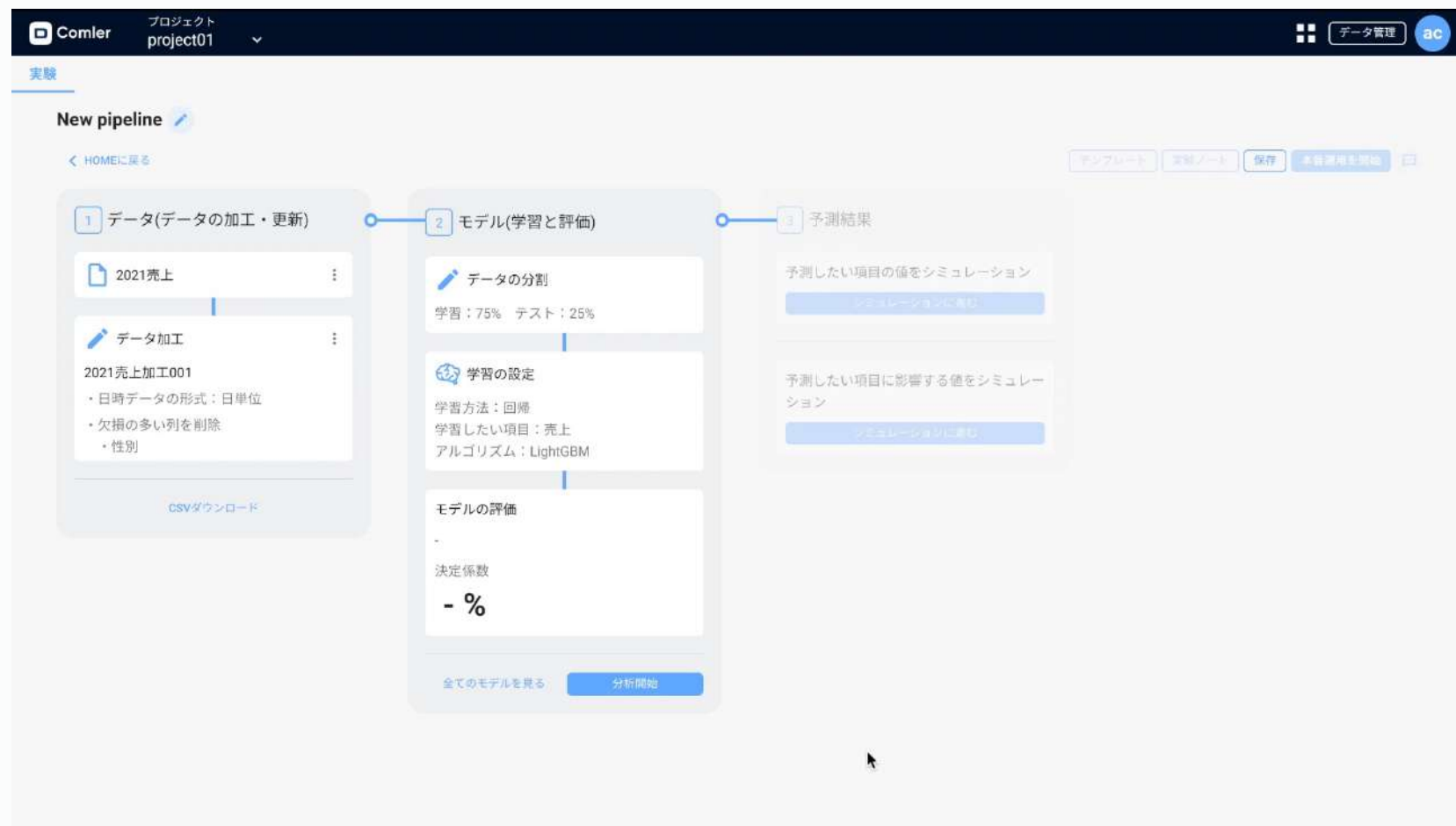
モデル構築機能
(回帰・分類・時系列)



**モデル評価
レポート**



モデル比較





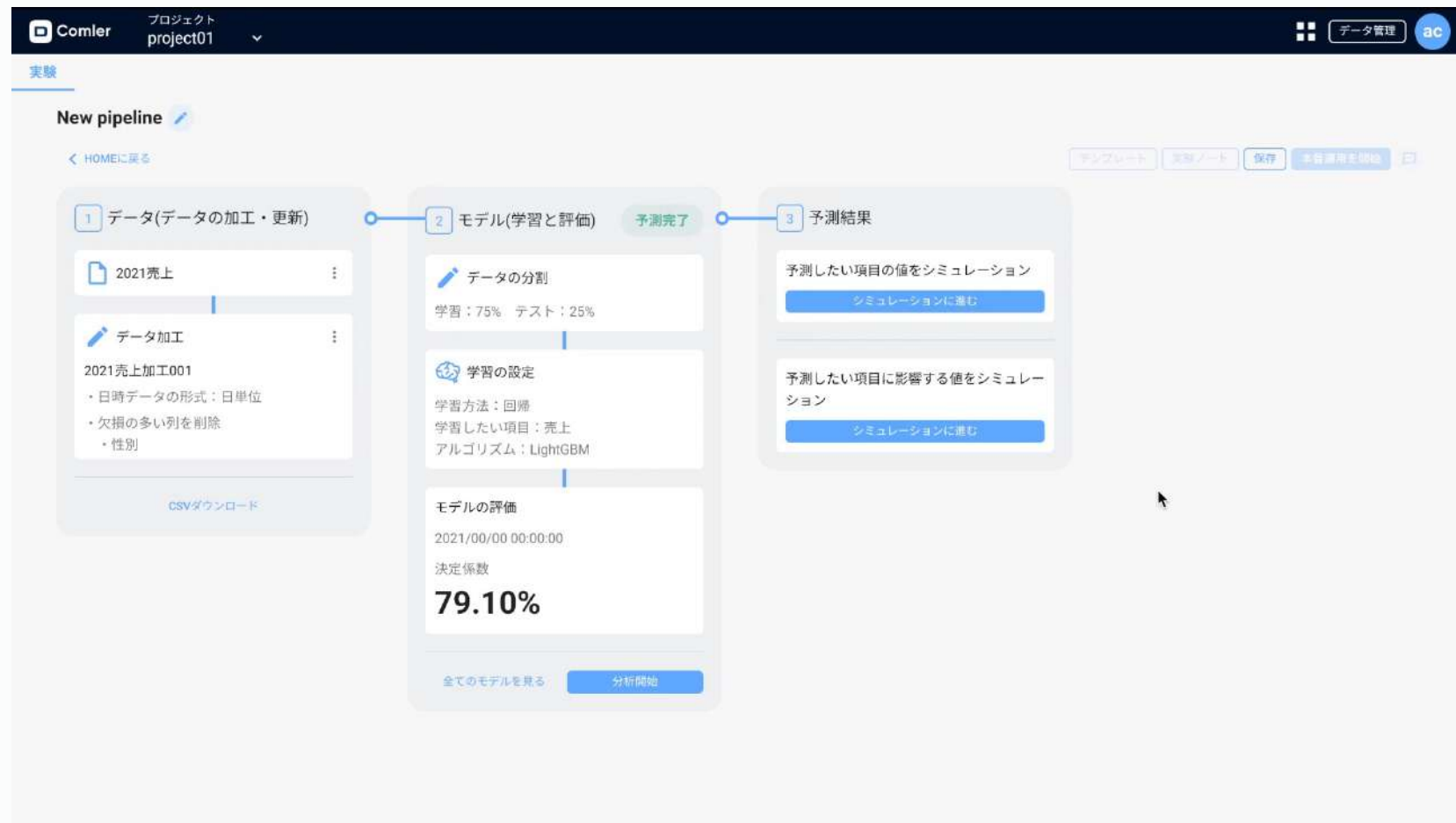
Comler
Collaborative Machine Learning Platform

構築したモデルはすぐにビジネス観点で活用が可能 最適化シミュレーションにより施策検討を支援

✓ 予測テスト

✓ 最適化
シミュレーション

✓ 影響度
シミュレーション



特許出願中

* 現在開発中のUIのためリリース時に変更になる可能性がございます。

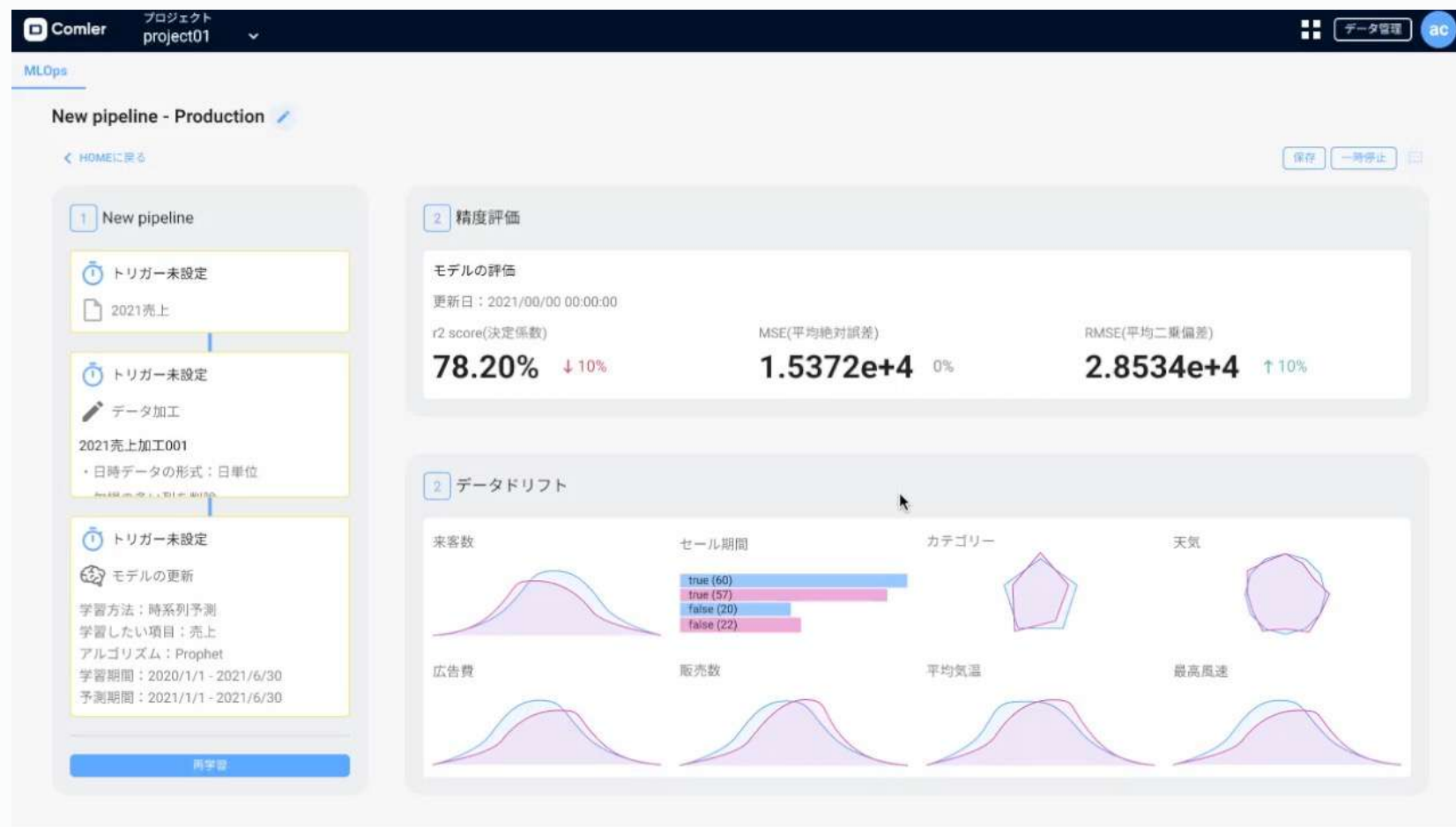


モデル運用のためのMLOpsパイプラインを簡単に用意 モデル作成時と実際に使われているデータの差分分析が可能

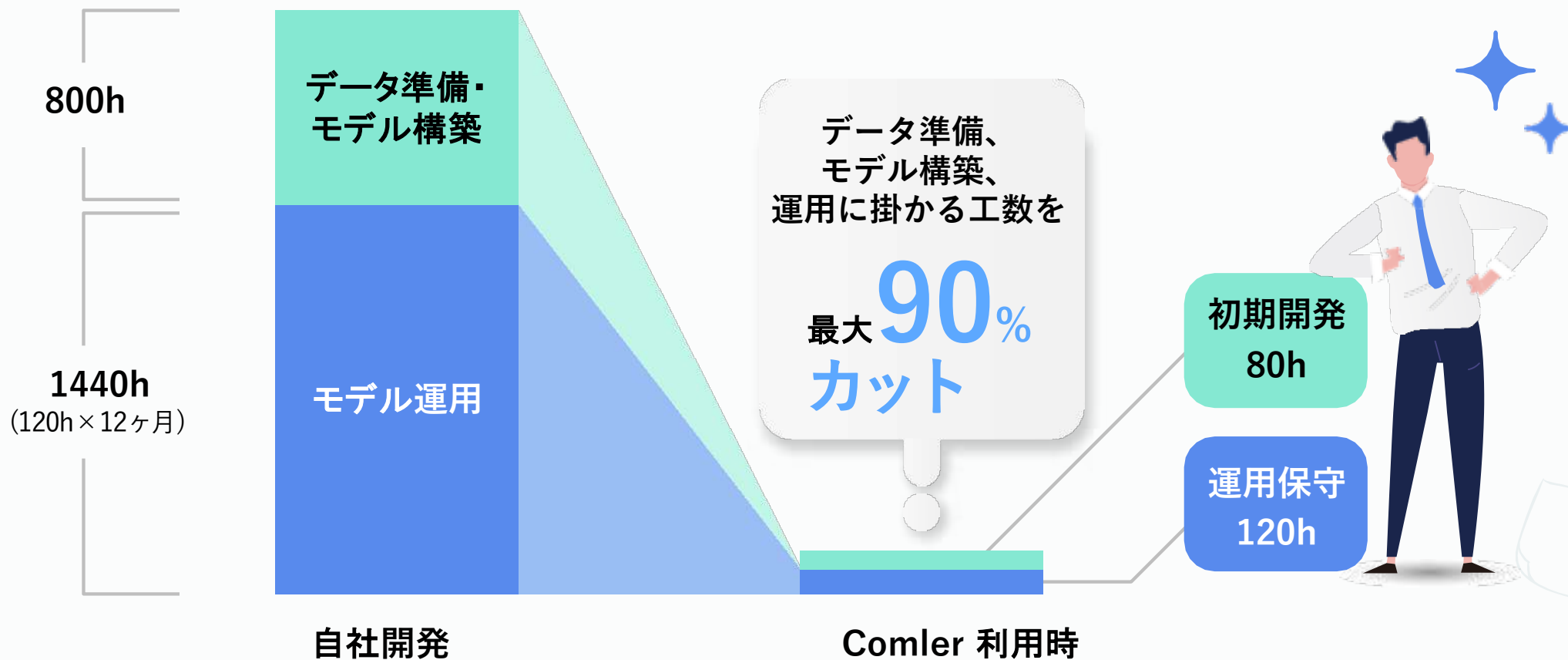
✓ MLOps
パイプライン

✓ データ差分解析・
再学習要否判定

✓ モデル精度監視



モデル構築・運用工数を削減し、プロジェクトを成功へ



最大 **90%** の工数を削減

初期開発だけでなく、毎月 **120** 時間もかかる運用が毎月 **10** 時間に

業界別ユースケース例

小売	来客数予測・売上予測・需要予測
流通・農業	仕入れ最適化・収穫量予測
金融	経済指標予測・個人支出予測
環境・エネルギー	発電予測・蓄電予測・プライシング最適化
物流・工場	ルート最適化・故障予兆検知
デジタルマーケティング	コンバージョン率予測
モビリティ	渋滞予測・都市交通最適化
都市開発	人口増減予測・災害予測
スマートシティ	健康リスク判定・診断要否判定
エンターテインメント	視聴率予測
ソーシャルメディア・ゲーム	顧客離反リスク予測

“データ活用の民主化ツール”を利用し、 あらゆる組織をデータ分析組織へ。

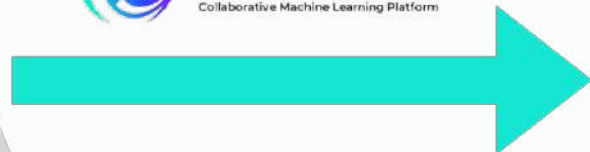
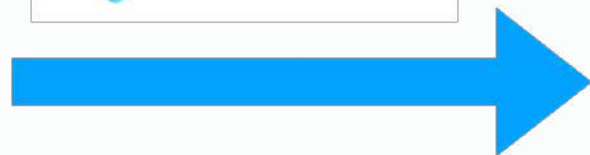


非デジタル人材をデータ活用人材へ。

高度データ活用人材

潜在データ人材

非デジタル人材

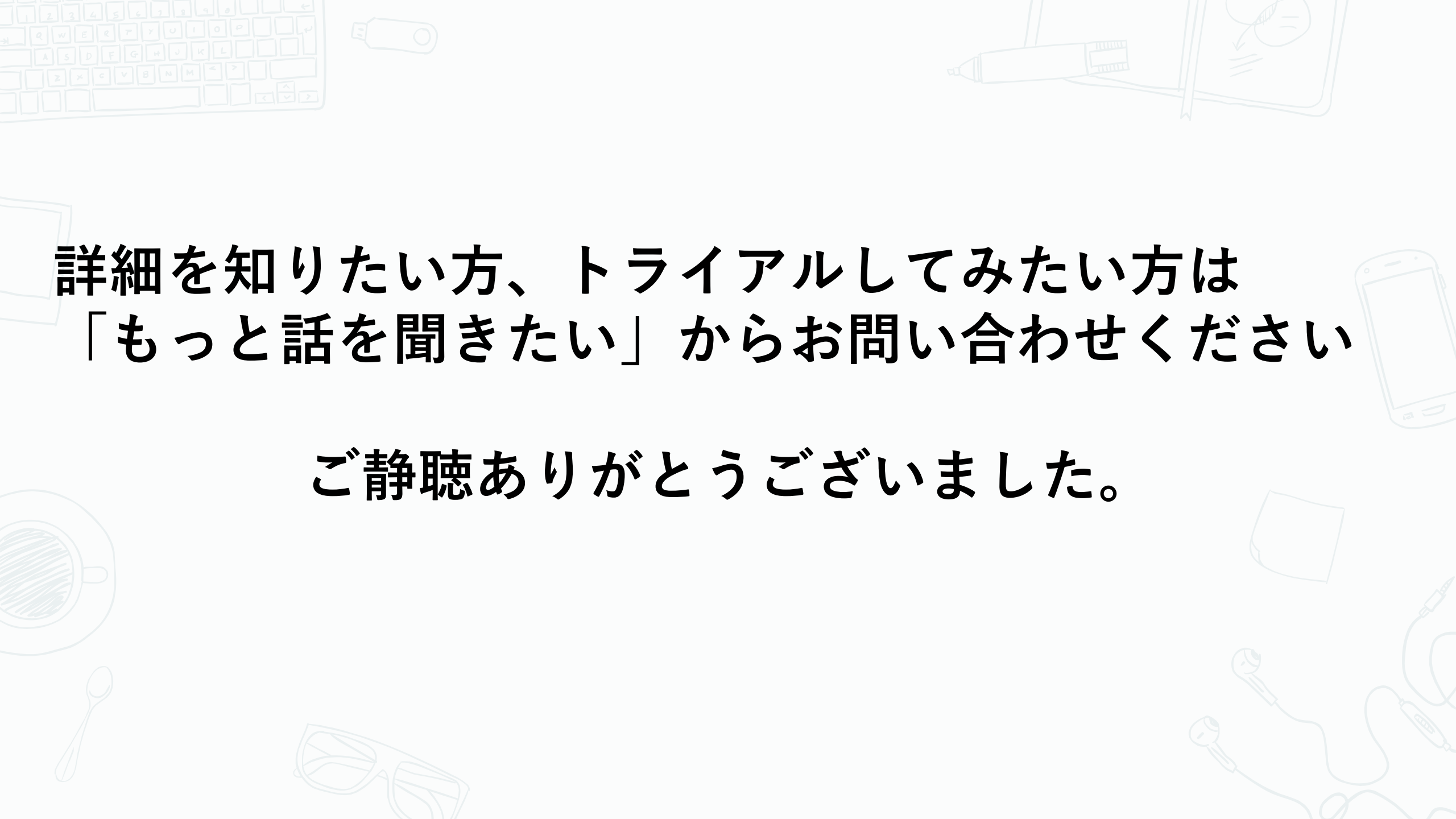


高度データ活用人材

ノーコードデータエンジニア
(基盤構築可能)

ノーコードアナリスト
(モデル構築可能)

ノーコードBIデザイナー
(データ可視化)



詳細を知りたい方、トライアルしてみたい方は
「もっと話を聞きたい」からお問い合わせください

ご静聴ありがとうございました。