



ML Enablement Workshop #1 : 理解編

ベストプラクティスの理解

久保 隆宏

Developer Relation
Machine Learning

ML Enablement Workshop: 開会

ワークショップの受講を決めたCXO 〇〇様より、
ワークショップを通じ得たい成果について一言お願い致します。

ML Enablement Workshopのプロセス

理解: ベストプラクティスの理解

機械学習の「勝ちパターン」を理解し、活用どころを見極める。



応用: 顧客体験改善への応用

勝ちパターンが成立するプロセスを可視化し解決すべき課題を特定する。



開始: 顧客体験の改善を開始する

機械学習による勝ちパターン成立まで、段階的に顧客体験の改善をする。



「理解」プロセスの対象者とゴール

誰のためのセッション?

**実装機能の決定権を持つ
プロダクトマネージャー。**

また、機械学習を扱う
プロジェクトに参加され
る方は誰でも。

何ができるようになる?

機械学習が価値を生む
勝ちパターンを理解し、
ユーザーストーリーの
中から発見できるよう
になる。

何ができるようにはな
らない?

機械学習アルゴリズム
の実装。

疑問：なぜ機械学習に取り組む必要があるのか？

回答:顧客体験を迅速に改善するため。

詳細は本編で！

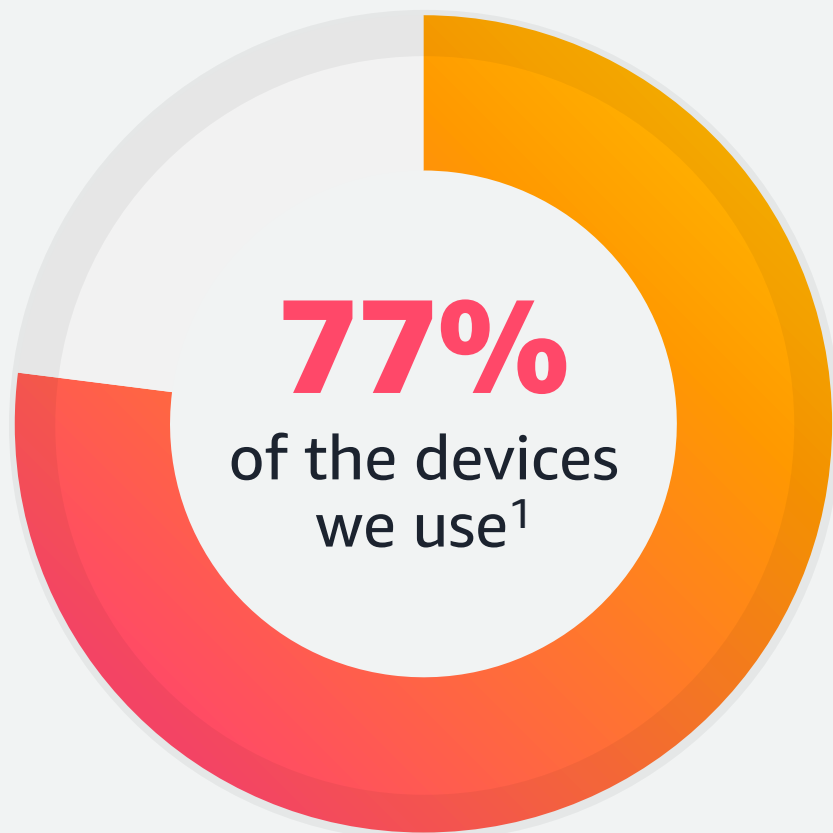
1. 機械学習の仕組み
2. 機械学習の勝ちパターン
3. ベストプラクティス事例

機械学習はだれが使っている？

唐突な短いクイズ！

今日、朝起きてから今の時間までに、
機械学習を使用したアプリケーションを使った方！

あなたが機械学習の利用者です！



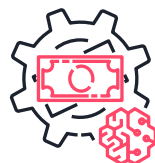
ライドシェア (例: Uber, Lyft)

待ち時間、需要予測、価格設定



オンラインショッピング

検索結果、商品レコメンド



金融機関

書類（小切手等）の認識
クレジット・デビットカードの不正取引検知



2023年の世界

80億個以上のAI/ML搭載デジタルアシスタント

¹ <https://techjury.net/stats-about/ai/> ² <https://www.juniperresearch.com/press/press-releases/digital-voice-assistants-in-use-to-8-million-2023>

わたしです！

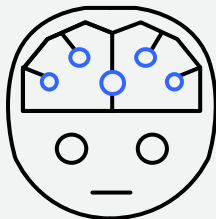


一家に一台

iRobotの導入事例

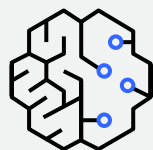
[AWS re:Invent 2020: Running machine learning workflows at enterprise scale using Kubeflow](#)

AI/MLとは?



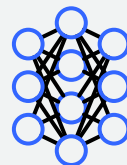
人工知能 (Artificial Intelligence, AI)

人間の知的判断をコンピュータ上で実現するための技術全般
(ロジック、if-then文、機械学習 (深層学習を含む))



機械学習 (Machine Learning, ML)

AIの一種であり、知的モデルを構築するためにデータの中の傾向を学習する技術



深層学習 (Deep Learning, DL)

MLの一種であり、音声・画像認識などのタスクを深い複数レイヤー構造のニューラルネットワークで実現する技術

従来型ソフトウェア vs. 機械学習

人手で設計されたルール ➤

データ ➤



➤ 答え

演繹的

過去の回答 ➤
過去のデータ ➤



データ ➤

➤ 答え

帰納的

機械学習のメリット

1. 実装容易性

プログラミング言語の知識がなくても計算処理を実装できる。

2. 高速な演算

行列演算で処理が構成されるため、GPUによる高速化が可能。

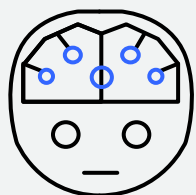
3. 高精度

画像認識や音声認識など一部領域で既存ソフトウェアを上回る。

4. 可搬性

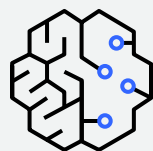
モデルの実行環境があればどこでも動かせる。

生成系AIは、深層学習の用途の一つ。大規模なデータから追加学習なしに人間のようなデータを生成するモデルを構築。



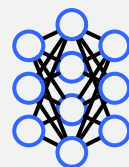
人工知能 (Artificial Intelligence, AI)

人間の知的判断をコンピュータ上で実現するための技術全般
(ロジック、if-then文、機械学習 (深層学習を含む))



機械学習 (Machine Learning, ML)

AIの一種であり、知的モデルを構築するためにデータの中の傾向を学習する技術



深層学習 (Deep Learning, DL)

MLの一種であり、音声・画像認識などのタスクを深い複数レイヤー構造のニューラルネットワークで実現する技術



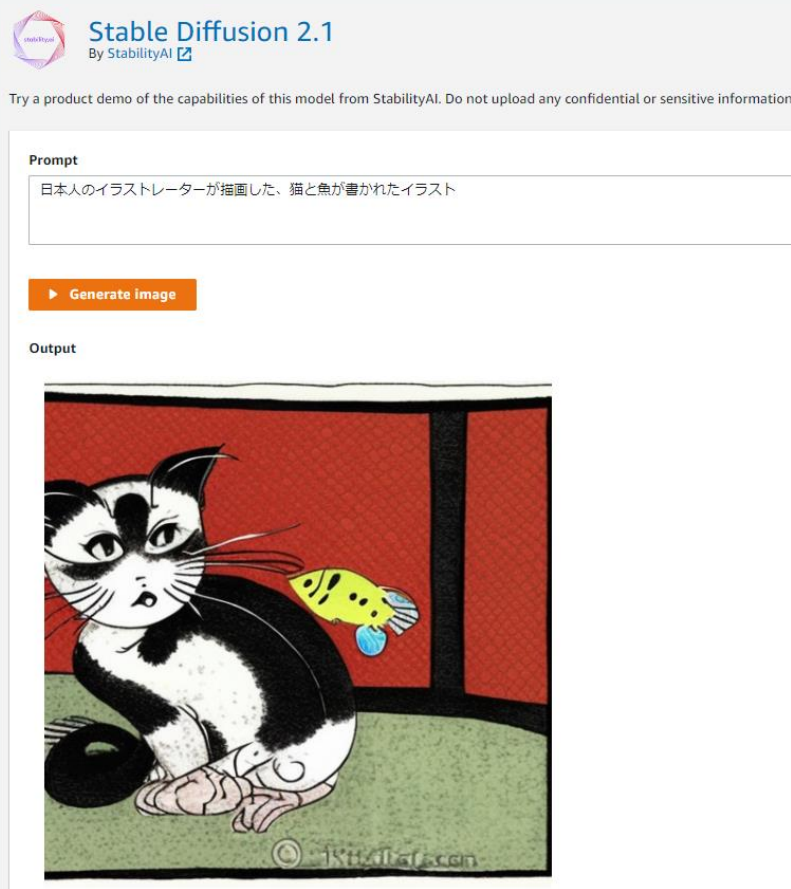
生成系AI (Generative AI)

テラバイト規模のデータで数千億規模のパラメーターのモデルを学習することで、**追加学習なしに**人間のような生成を実現する技術。

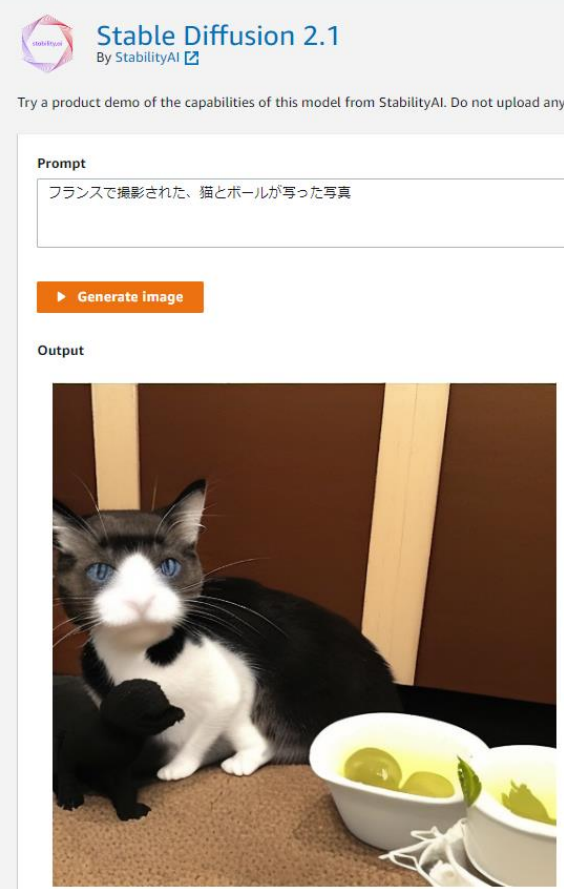
15

生成系AIの生成例

“日本人のイラストレーターが描画した、猫と魚が書かれたイラスト”の生成例



“フランスで撮影された、猫とボールが写った写真”の生成例



AWSのSageMaker JumpStartで、テキストから画像を生成するStable Diffusionのモデルを実行した例

生成系AIの生成例

“日本人のイラストレーターが描画した、猫と魚が書かれたイラスト”の生成例



Stable Diffusion 2.1
By StabilityAI

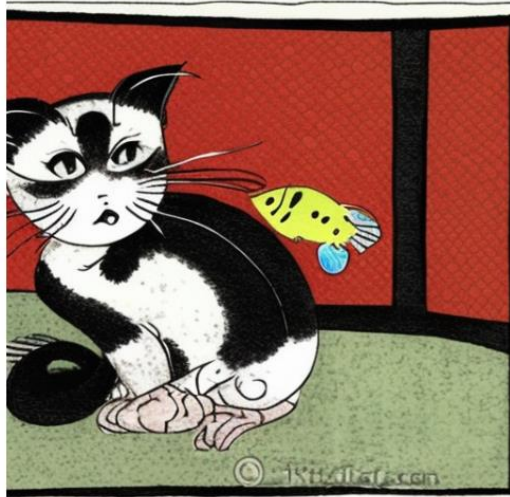
Try a product demo of the capabilities of this model from StabilityAI. Do not upload any confidential or sensitive information.

Prompt

日本人のイラストレーターが描画した、猫と魚が書かれたイラスト

▶ Generate image

Output



“フランスで撮影された、猫とボールが写った写真”の生成例



Stable Diffusion 2.1
By StabilityAI

Try a product demo of the capabilities of this model from StabilityAI. Do not upload any

Prompt

フランスで撮影された、猫とボールが写った写真

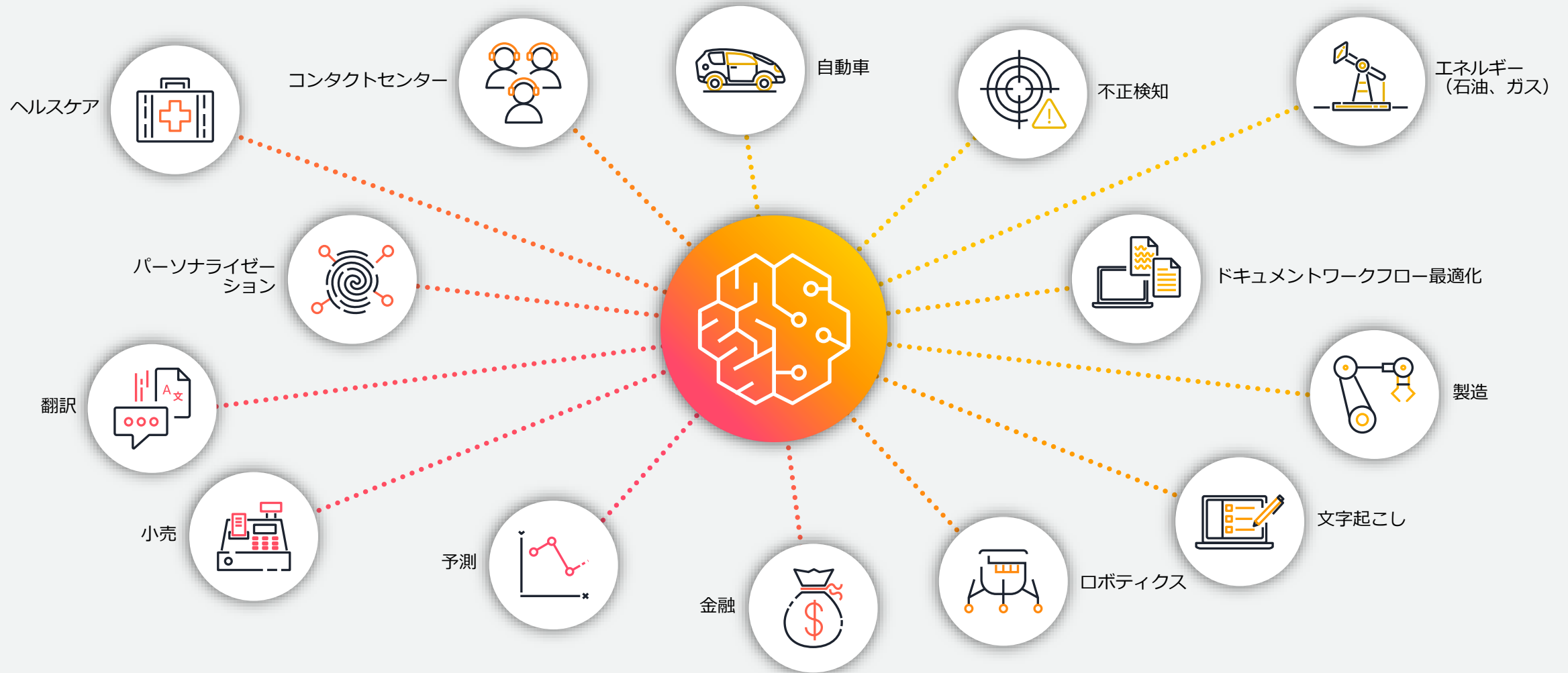
▶ Generate image

Output



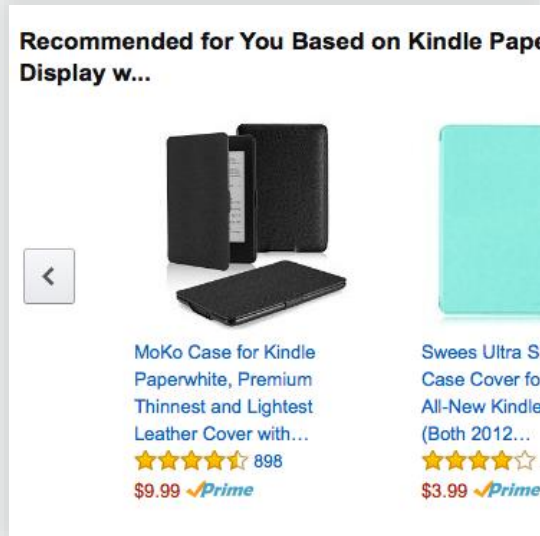
追加の画像データを学習させることで、プログラミングの知識がなくても出力をカスタマイズできる。

機械学習はあらゆる場面に

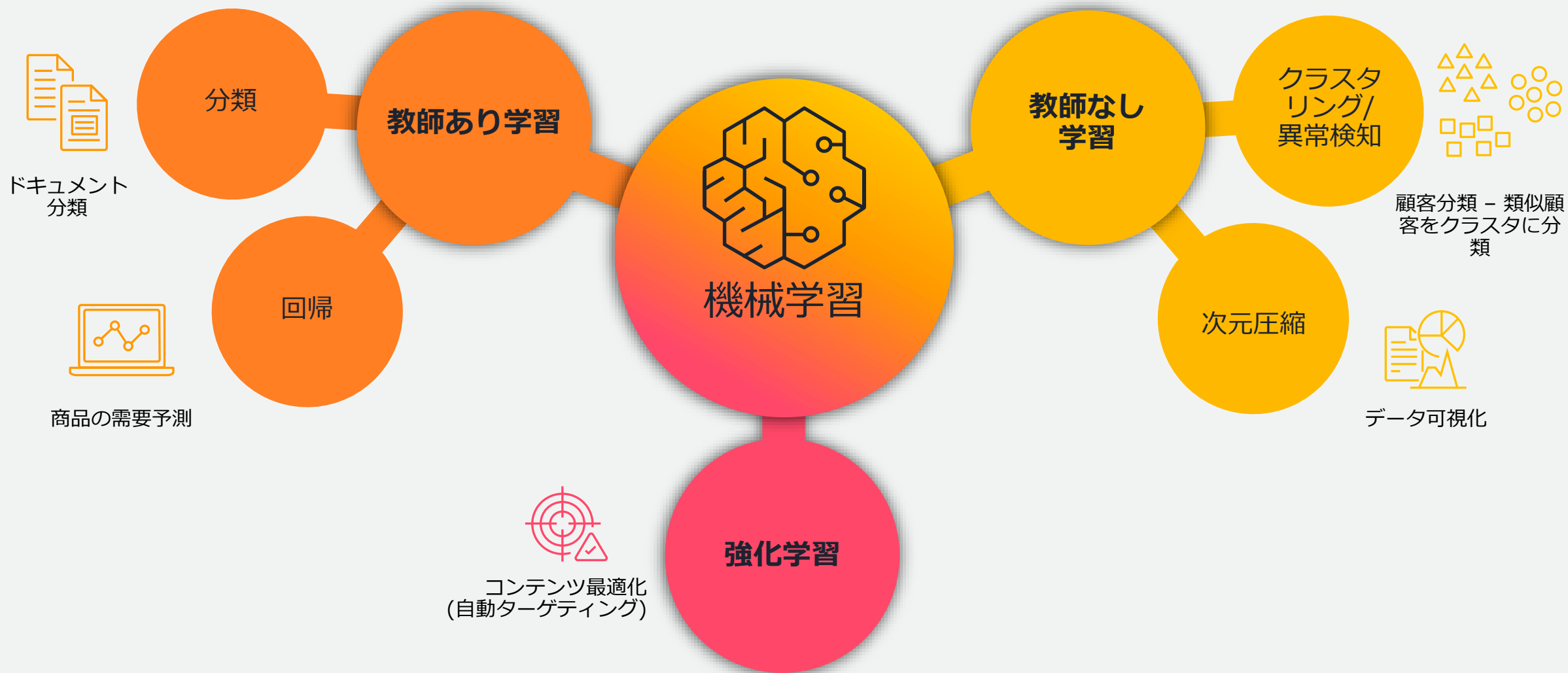


Amazonでの機械学習の実用例

AMAZON'S MACHINE LEARNING INNOVATION



モデルの学習方法とタスク



MLモデルの一般的なタスクとユースケース (1/6)

タスクの種類

ユースケース

事例

回帰

物件価格の予測

物件価格を予測

Zillow

Zestimate

\$1,496,442



ZESTIMATE RANGE

\$1.38M - \$1.62M



LAST 30 DAY CHANGE

+\$24,008 (+1.6%)

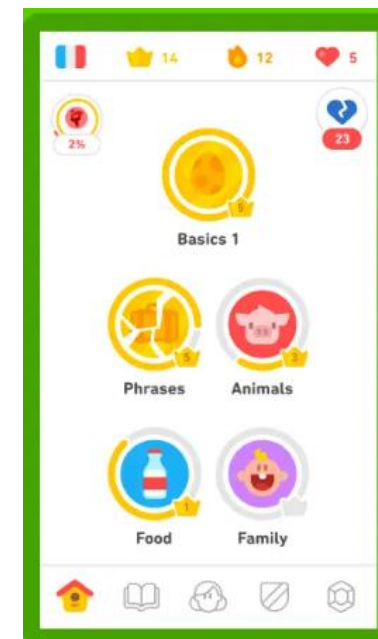
MLモデルの一般的なタスクとユースケース(2/6)

| タスクの種類 | ユースケース |
|--------|---------|
| 回帰 | 物件価格の予測 |
| 分類 | 単語の分類 |

事例

見直しが必要な単語か否かを予測

duolingo

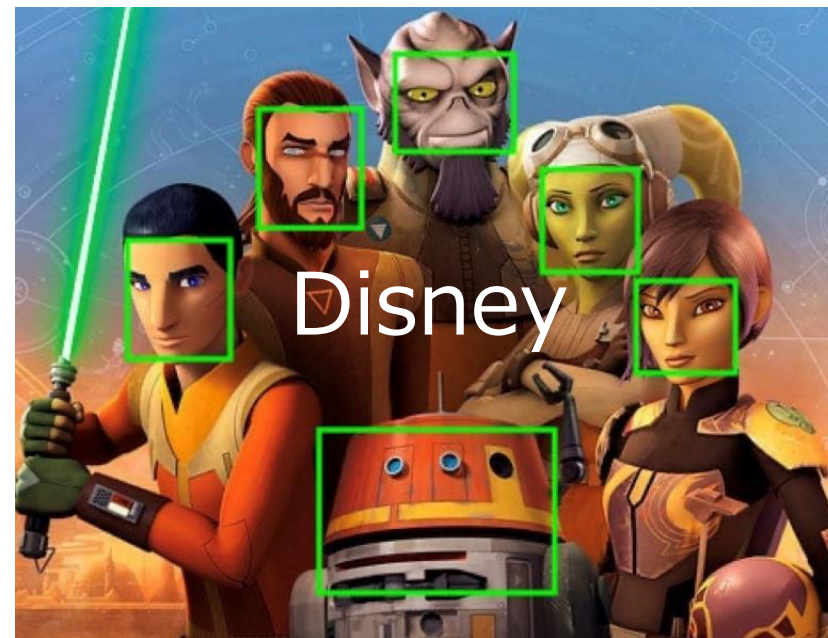


MLモデルの一般的なタスクとユースケース(3/6)

| タスクの種類 | ユースケース |
|---------|--------------|
| 回帰 | 物件価格の予測 |
| 分類 | 単語の分類 |
| クラスタリング | 画像内の顔パターンの発見 |

事例

アニメーションの自動タグ付け



MLモデルの一般的なタスクとユースケース(4/6)

| タスクの種類 | ユースケース |
|---------|--------------|
| 回帰 | 物件価格の予測 |
| 分類 | 単語の分類 |
| クラスタリング | 画像内の顔パターンの発見 |
| 異常検知 | 例外的パターンの発見 |

事例

新たな不正行為の特定



MLモデルの一般的なタスクとユースケース(5/6)

タスクの種類

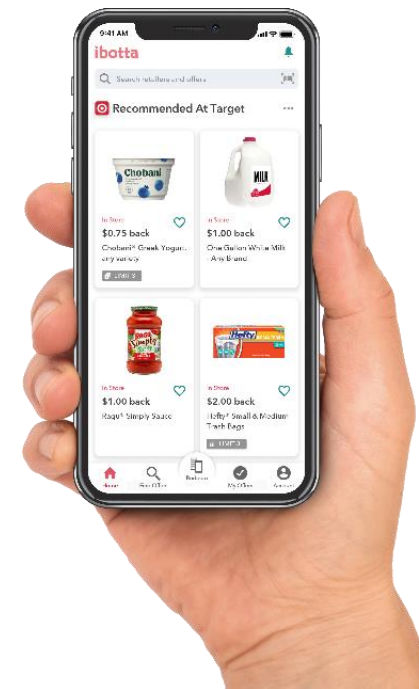
ユースケース

事例

| | |
|---------|--------------|
| 回帰 | 物件価格の予測 |
| 分類 | 単語の分類 |
| クラスタリング | 画像内の顔パターンの発見 |
| 異常検知 | 例外的パターンの発見 |
| ランキング | 選択されやすい順序の予測 |

検索結果の並び替え

ibotta

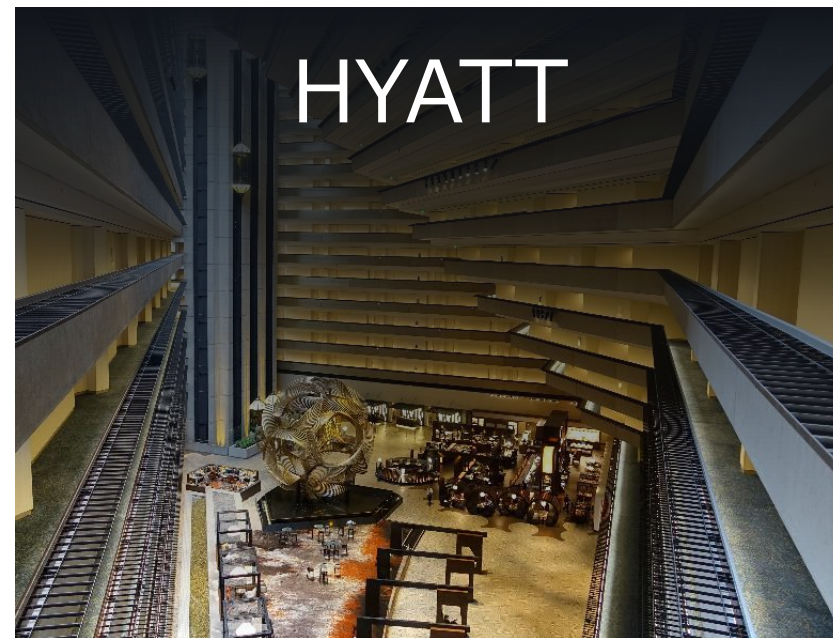


MLモデルの一般的な用途 (6/6)

| タスクの種類 | ユースケース |
|---------|----------------|
| 回帰 | 物件価格の予測 |
| 分類 | 単語の分類 |
| クラスタリング | 画像内の顔パターンの発見 |
| 異常検知 | 例外的パターンの発見 |
| ランキング | 選択されやすい順序の予測 |
| レコメンド | 選択されやすいアイテムの予測 |

事例

大規模イベント用ホテルの推薦



ナレッジチェック: MLのタスク

次のユースケースに適用できるMLのタスクは何?

回帰・分類・クラスタリング・異常検知・ランキング・レコメンド

- a) 人手による商品価格設定の効率化
- b) 偽造品の特定
- c) 誤って分類された商品の識別
- d) マーケティングセグメント特定のための顧客のグループ化
- e) コホート別顧客支出の予測

ナレッジチェック: MLのタスク

次のユースケースに適用できるMLのタスクは何?

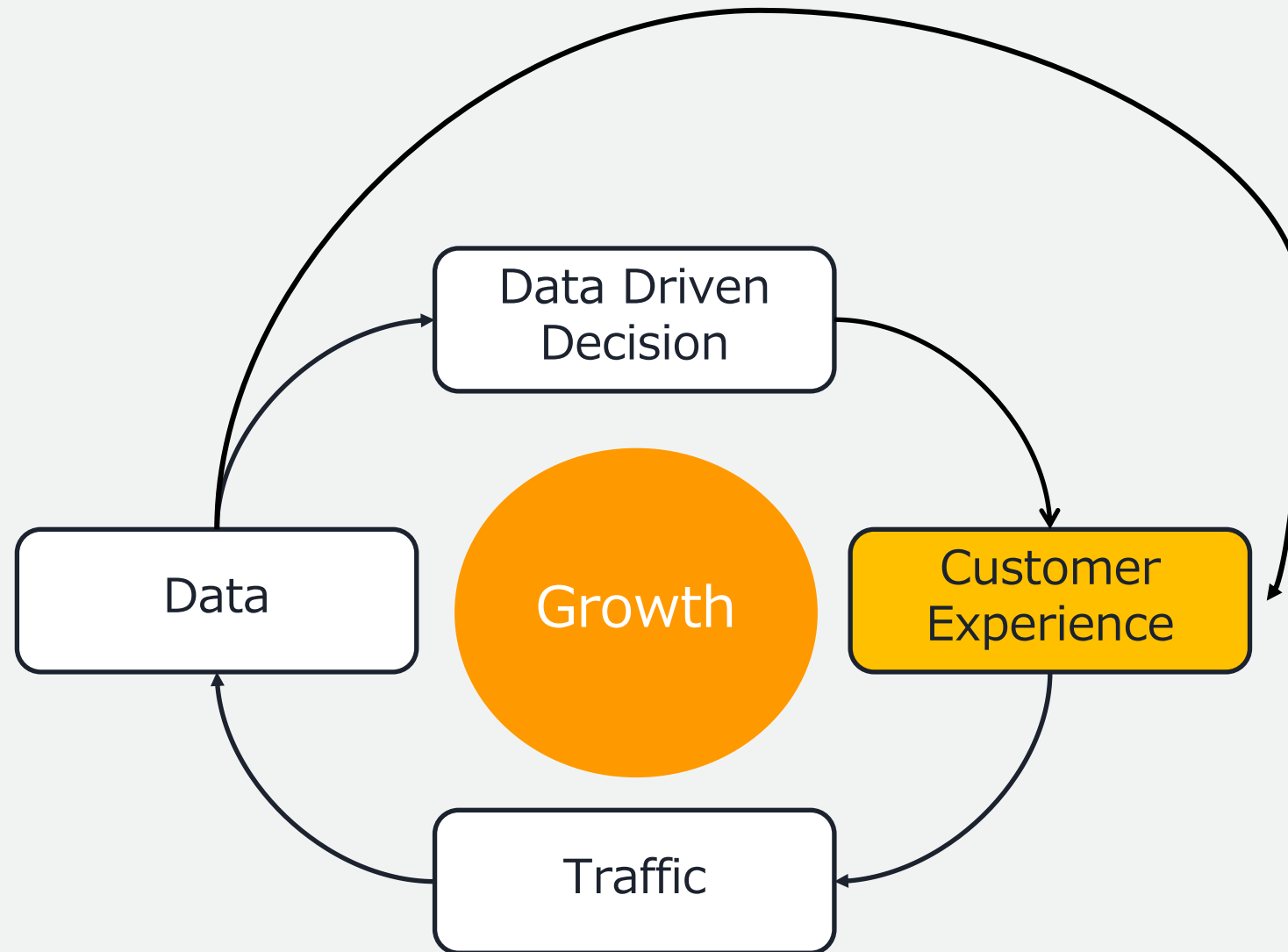
回帰・分類・クラスタリング・異常検知・ランキング・レコメンド

- a) 人手による商品価格設定の効率化 – 回帰
- b) 偽造品の特定 – 分類
- c) 誤って分類された商品の識別 – 異常検出
- d) マーケティングセグメント特定のための顧客のグループ化 – クラスタリング
- e) コホート別顧客支出の予測 – 回帰

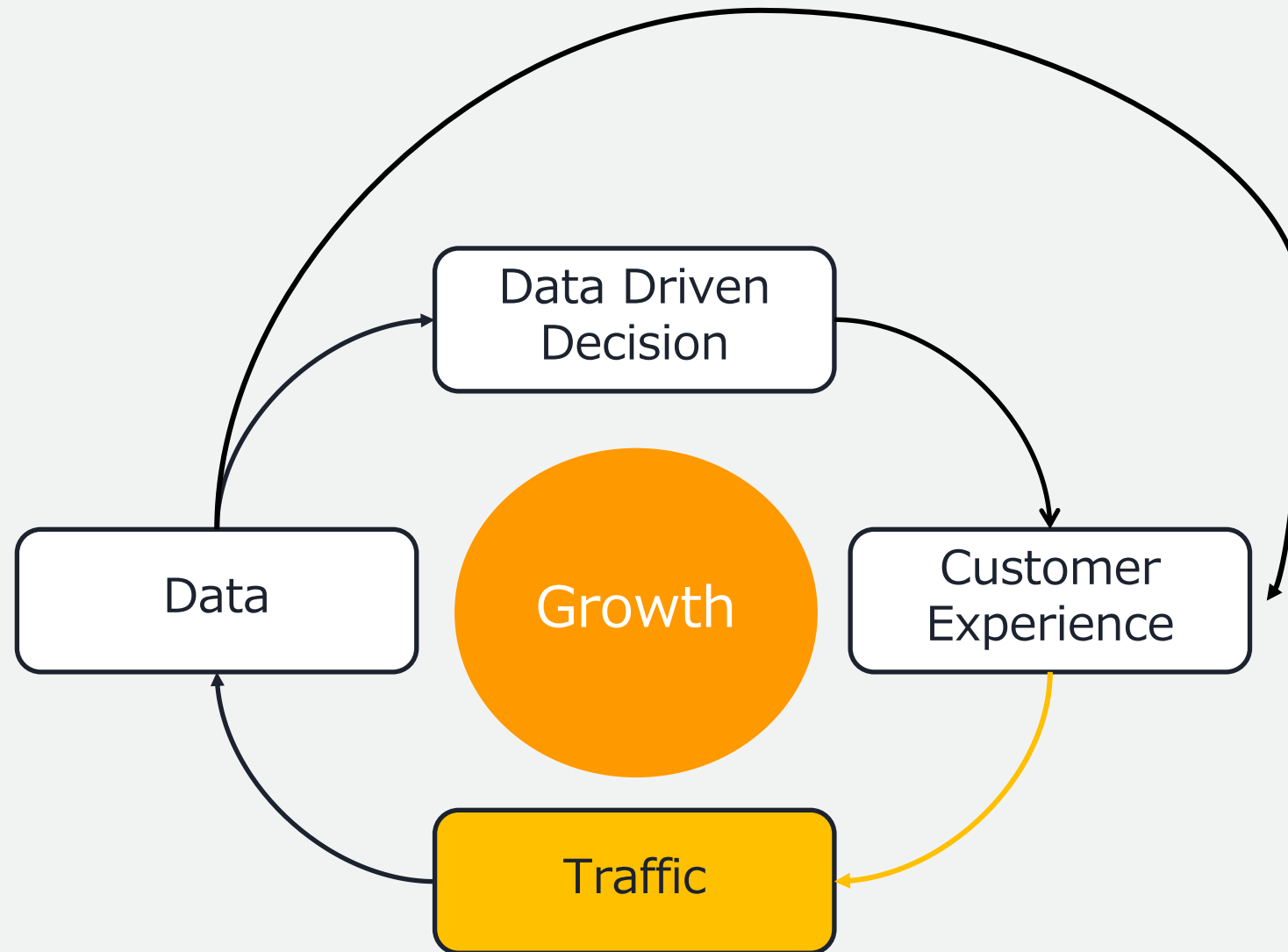
上記以外の方法がありますか？

1. 機械学習の可能性
2. 機械学習の勝ちパターン
3. ベストプラクティス事例

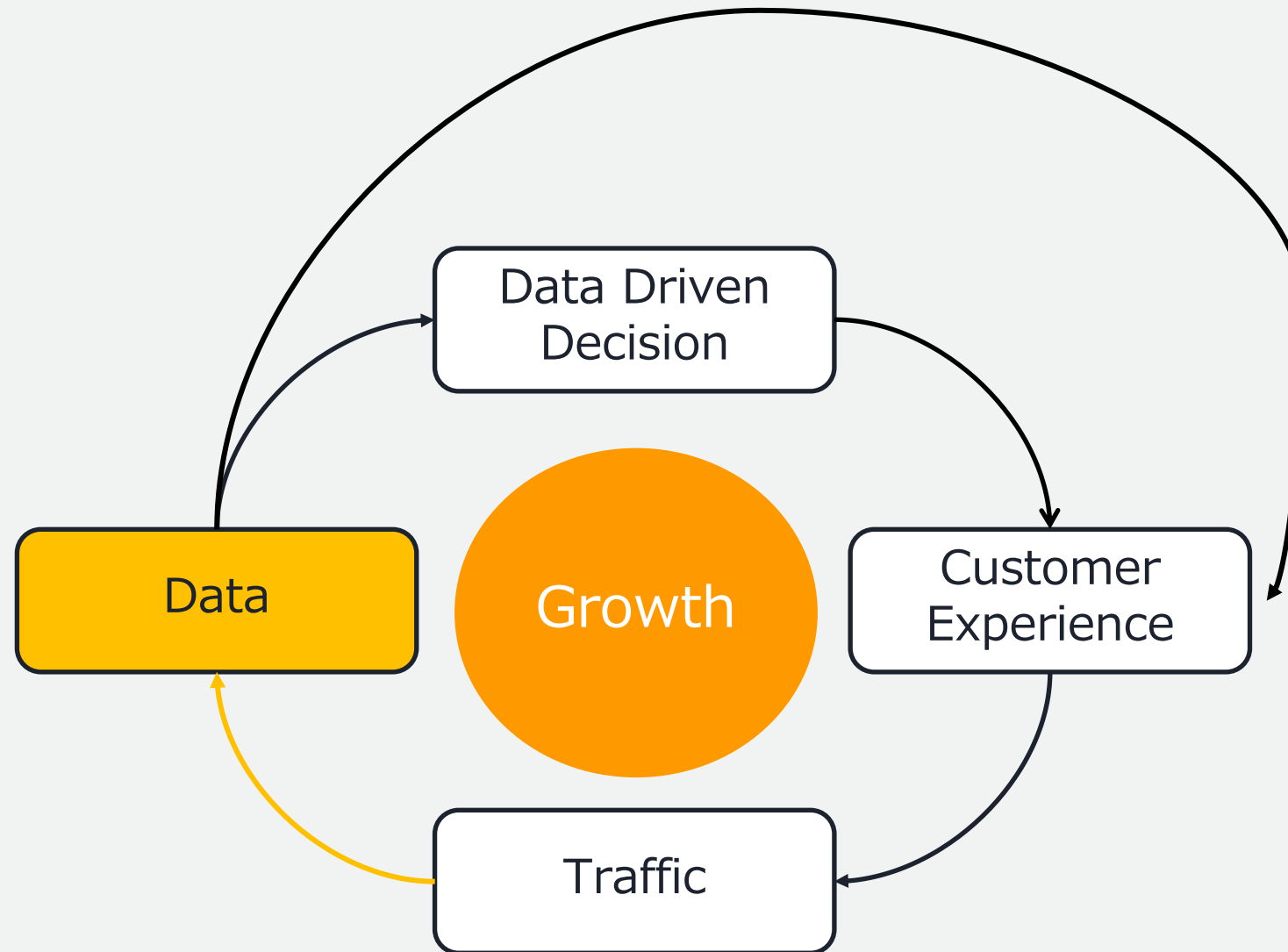
機械学習の勝ちパターン



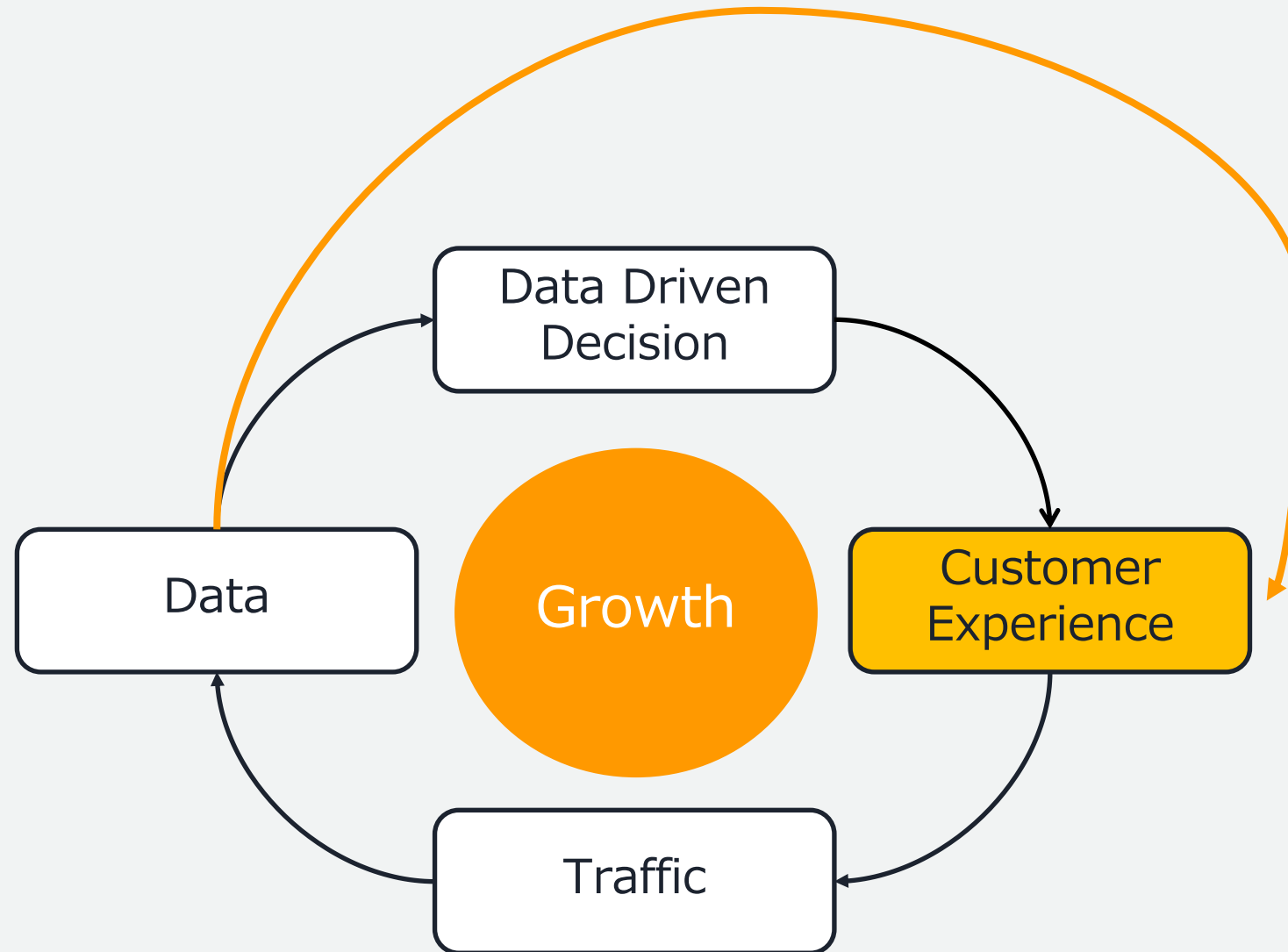
機械学習の勝ちパターン



機械学習の勝ちパターン



機械学習の勝ちパターン



機械学習の勝ちパターン

1. 機械学習がCustomer Experienceを改善する。
2. ユーザー数、Trafficが増える。
3. Trafficが増えるとDataが増える。
4. Dataが増えるとモデルが改善され・・・=> 1に戻る。

機械学習の勝ちパターン

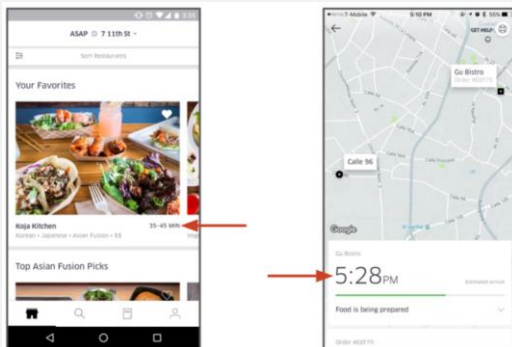
1. 機械学習がCustomer Experienceを改善する。
2. ユーザー数、Trafficが増える。
3. Trafficが増えるとDataが増える。
4. Dataが増えるとモデルが改善され・・・ => 1に戻る。

データが蓄積する速度は、一般的にソフトウェアをリリースする速度より速い。

機械学習を使うことで迅速にソフトウェアを改善できる。

回答:顧客体験を迅速に改善するため。

勝ちパターンを確立した事例



画像引用: [Meet Michelangelo: Uber's Machine Learning Platform](#)



画像引用: [Get Perfect Song Recommendations in the Playlists You Create with Enhance](#)

Uber Eats

機械学習で配達時間を予測。配達時間が正確になるほど顧客体験が改善され、顧客体験が改善されるほど利用者が増えてよりデータが集まり正確になる。

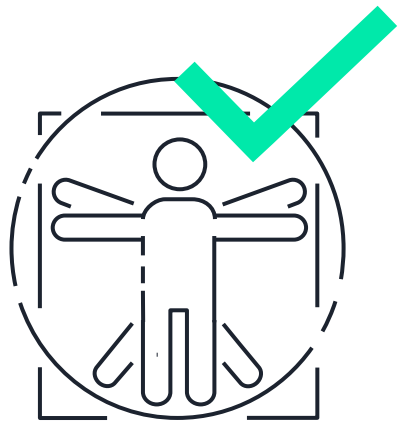
Spotify

機械学習によるプレイリストのパーソナライゼーションを実装。気に入った音楽を+でプレイリストに加えていくことで、より好みに合った音楽が推薦され、好みに合うほど機能を使う機会が増えデータが集まり正確になる。

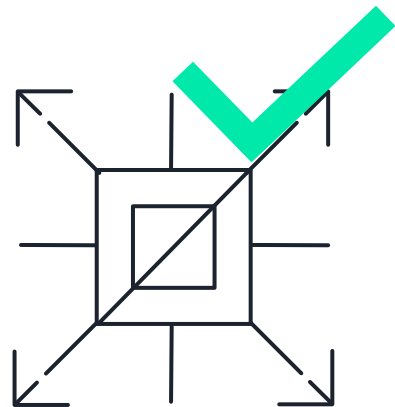
掃除ロボットを通じデータが集まるほど新製品・機能で顧客体験を改善できる。

<https://www.irobot-jp.com/policy/>

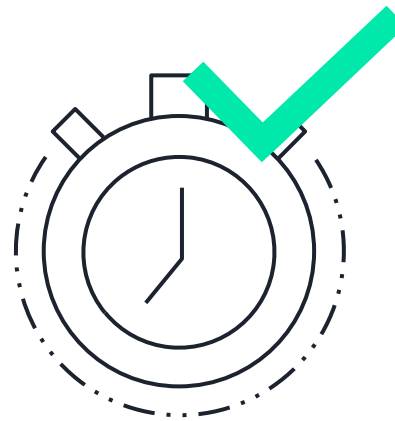
勝ちパターンに繋がりやすい条件



複雑なロジック



スケーラビリティ



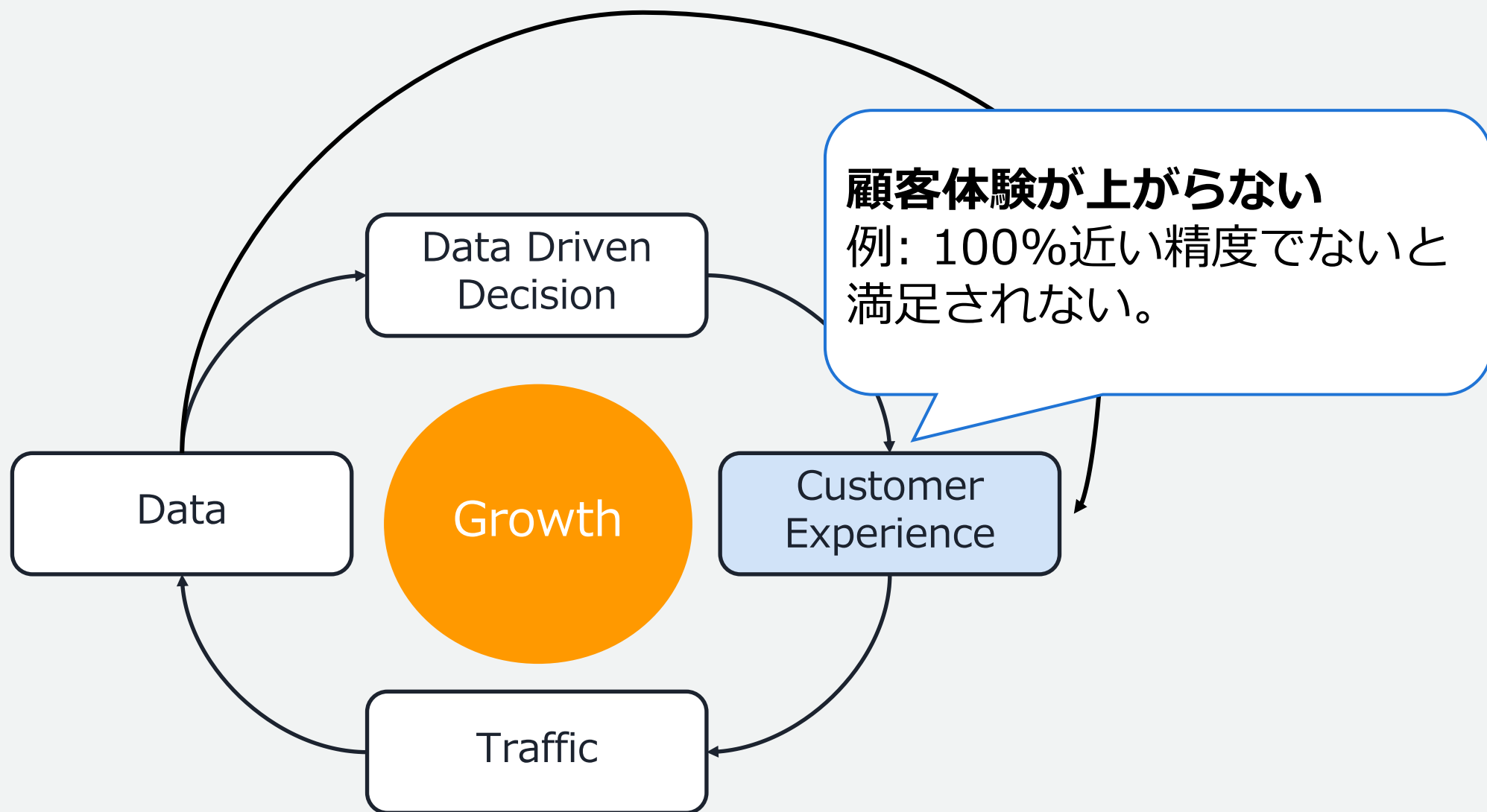
即時的な
レスポンス



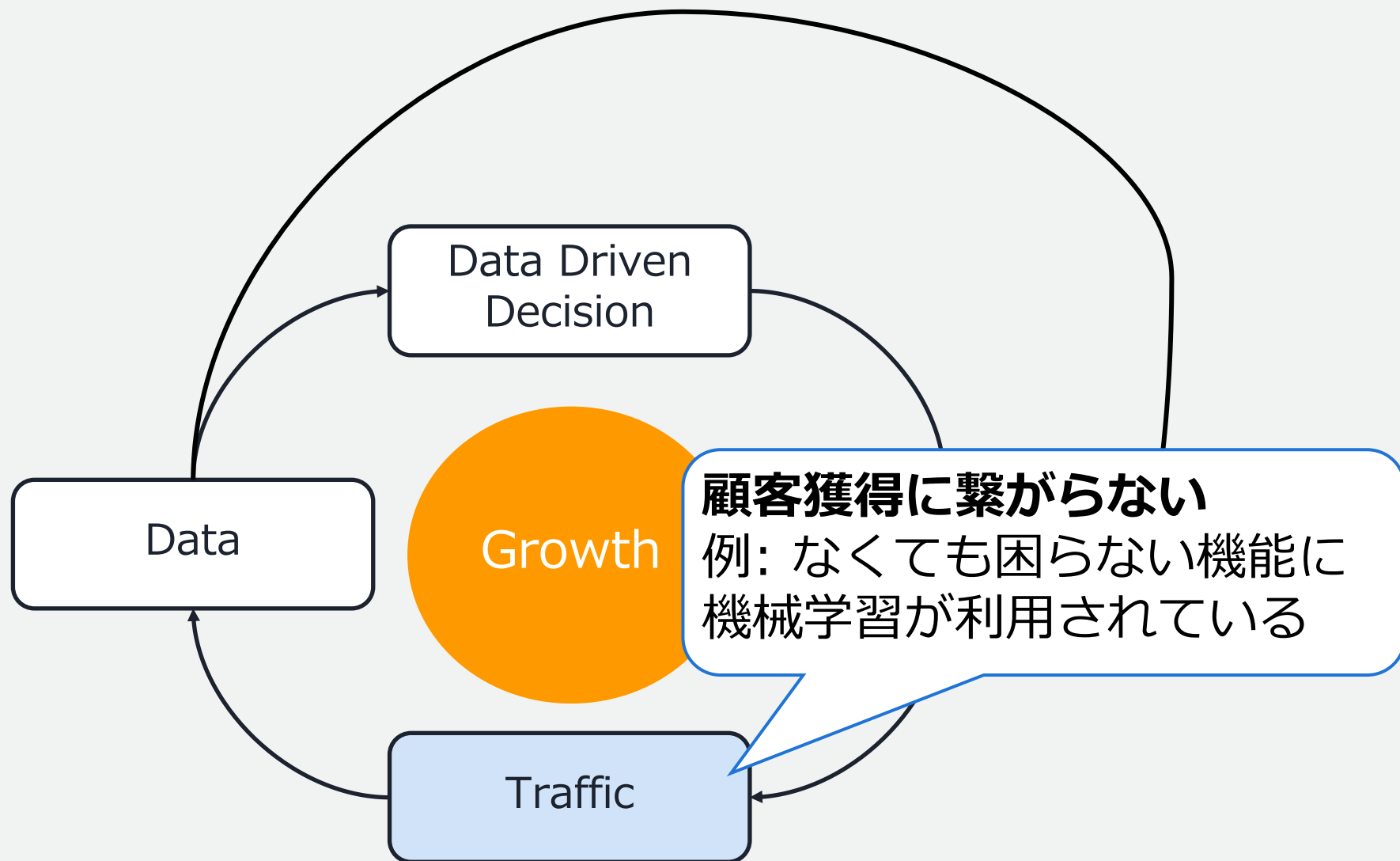
パーソナライゼー
ション

負けパターンとは?

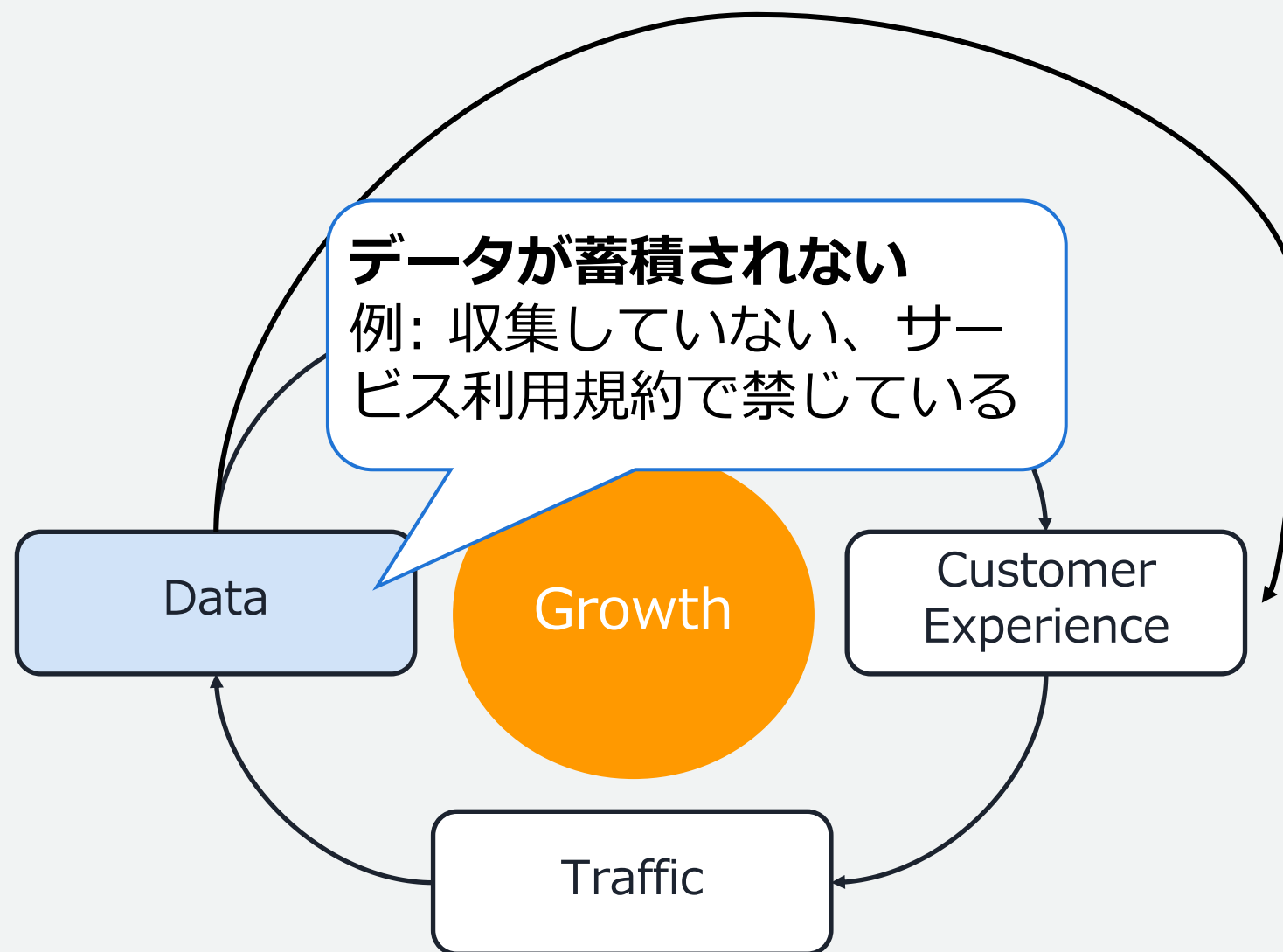
機械学習の負けパターン



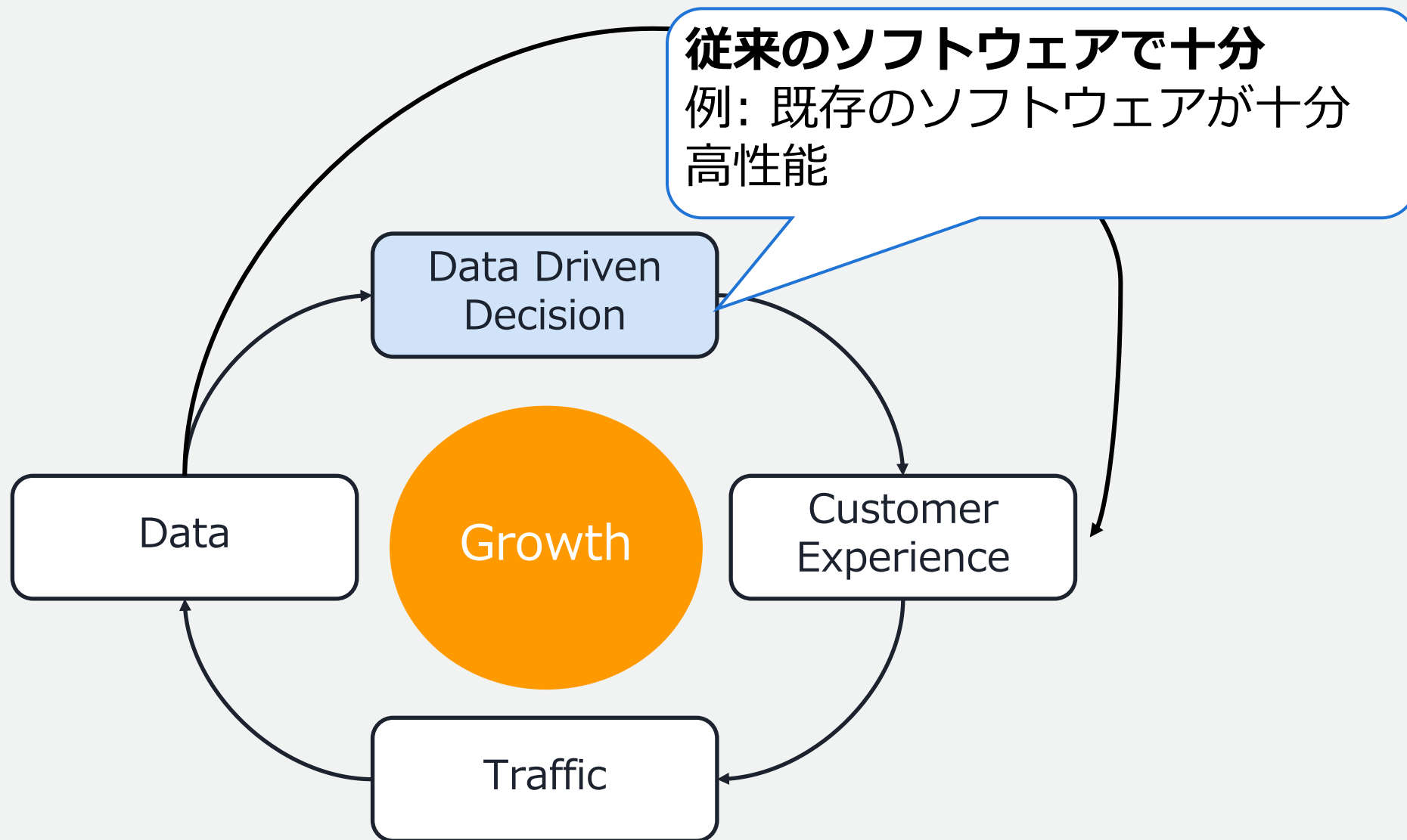
機械学習の勝ちパターン



機械学習の勝ちパターン



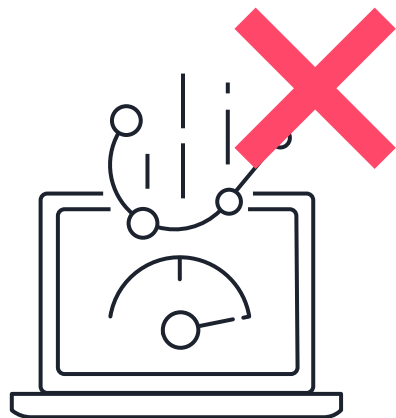
機械学習の勝ちパターン



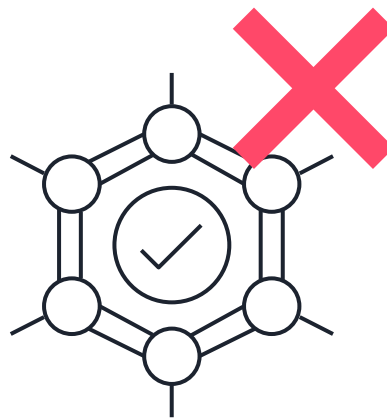
負けパターンに繋がりやすい条件



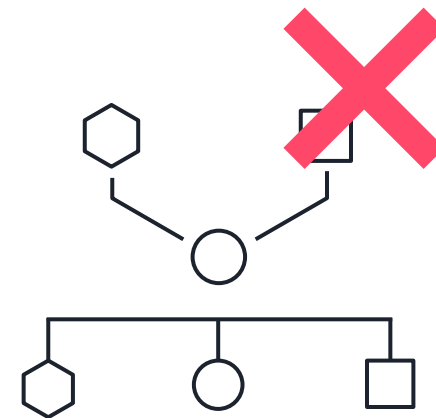
既存アルゴリズム
で解決可能



新たなデータへの
対応不要



精度**100%**が
要求される

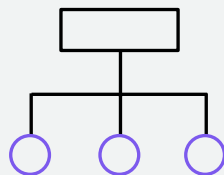
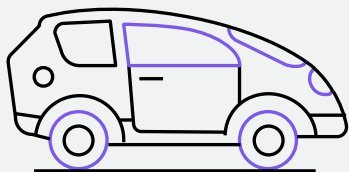


完全な説明性が
必要

MLの説明性

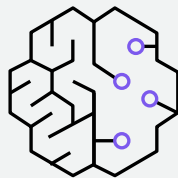
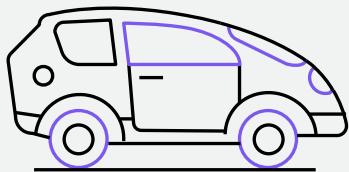
**MLアルゴリズムの中には説明性を有するものがある
その場合は、出力結果がどのように作られたのかを理解できる
ただし、全てのMLモデルがそうではない**

SIMPLE
MODEL



Tree

COMPLEX
MODEL



Car

説明性と精度との間には
トレードオフが見ら
れることが多い

シンプルなモデル vs. 複雑なモデル



モデルがどのように使われ、説明性がどの程度求められるのかを意識する

- › 初期段階ではシンプルなモデルを試す。結果が理解しやすく、早く開発することが可能。
- › ビジネスにとって意味のあるベースラインを設定する。最先端技術の性能の必要性を見極める。

説明性の不足

許容されない状況は？

- › 顧客に対し、ローン申請が拒絶された理由に対する説明が求められるとき
- › 取引が不正なものと判断された理由を説明する必要があるとき

許容される状況は？

- › 誤分類・誤判定のリスクが小さいとき（アイテムの表示順序など）
- › 最終判断が人間によって行われるとき

不確実性

モデルは過去のデータに基づき確率的に予測する。予測の傾向は、人間の判断と一致しないかもしれない。



どれが犬でどれがパンか？
人が間違える傾向とモデルが間違える傾向が一致するとは限らない

1. 機械学習の可能性
2. 機械学習の勝ちパターン
3. ベストプラクティス事例

Intuit: 控除対象の経費を探す時間を削減

PROBLEM

Intuit社の主力製品「TurboTax」を用いて米国の所得税控除の申告書を作成する際、ユーザが控除対象となる項目の抽出に多くの作業時間を要していた

SOLUTION

顧客の1年間の銀行取引データを解析し、控除可能な経費を自動的に「**分類**」。
モデルの構築に Amazon SageMaker(MLモデルの開発環境) を活用し、「ExpenseFinder」という商品を開発。

IMPACT

ユーザの作業時間を数時間から数分に短縮。また、SageMakerの導入によってAI/MLを実用化するまでの期間を90%短縮（6カ月→1週間）。

演習：Intuitの勝ちパターンの要素は？

1. なぜ顧客体験が改善するか？
2. なぜ利用者が増えるか？
3. なぜデータが増えるか？
4. なぜモデルの精度が向上するか？

回答例: Intuitの勝ちパターンの要素は?

1. なぜ顧客体験が改善するか?
機械学習により控除が自動で分類できるから。
2. なぜ利用者が増えるか?
数時間かかっていた作業が数分になるから。
3. なぜデータが増えるか?
利用者が増えるほど取引データが増えるから。
4. なぜモデルの精度が向上するか?
取引データと実際申請された結果のセットが手に入るから。

Pomelo: 嗜好に合う商品を探す時間の削減

PROBLEM

創業以来同じ形式でWebサイト上に商品が表示されており、販売データの反映も遅かった。そのため、顧客は好みの商品を見つけることもトレンドの商品を見つけることも困難だった。

SOLUTION

個々の顧客向けにパーソナライズされた商品**レコメンド**を実装。レコメンドに使用するデータを最新化し、顧客の好みを数分で反映できるようにした。実装には Amazon Personalize(レコメンド)を活用した。

IMPACT

商品カテゴリページから個別商品へのクリックスルー率を18%向上し、商品カテゴリページからの収益を15%向上した。これにより、1カ月以内にROI 400%増加を達成。

MORE INFO: [CASE STUDY](#) | [VIDEO](#)



演習：Pomeloの勝ちパターンの要素は？

1. なぜ顧客体験が改善するか？
2. なぜ利用者が増えるか？
3. なぜデータが増えるか？
4. なぜモデルの精度が向上するか？

回答例: Pomeloの勝ちパターンの要素は?

1. なぜ顧客体験が改善するか?
機械学習により購買履歴に沿った商品が推薦されるから
2. なぜ利用者が増えるか?
サイトの体験が良いと口コミで広まるため
3. なぜデータが増えるか?
利用者が増えるほど購買データが増えるから
4. なぜモデルの精度が向上するか?
購買データあるほど推薦モデルの精度が向上するため

Fraud.net: 新たな詐欺の手法を検知し不正利用被害を防止

PROBLEM

新たな詐欺の手口がどんどん開発されるため、顧客の不正利用被害を防ぐには多様な手口それぞれについて高い精度で検知する必要がある。しかし、現状では1つのモデルで**異常検知**しているため個々の手口の精度を高めることが困難だった。

SOLUTION

単一のモデルですべての不正に対応する方式から、不正ごとにモデルを構築する方法に切り替えた。これにより、個々の不正検知の精度を高めることができる。一方で、モデルの管理が煩雑になるため管理、運用を行うためAmazon Machine Learningを採用した。

IMPACT

個別の不正を正確に検知できるようになったことで、Fraud.netの顧客は1週間で100万ドルの不正被害を防止できるようになった。200ミリ秒以内での応答が可能になり、アプリケーションのレスポンスも改善された。

演習: Fraud.netの勝ちパターンの要素は?

1. なぜ顧客体験が改善するか?
2. なぜ利用者が増えるか?
3. なぜデータが増えるか?
4. なぜモデルの精度が向上するか?

回答例: Fraud.netの勝ちパターンの要素は?

1. なぜ顧客体験が改善するか?
機械学習モデルが幅広かつ高速に不正利用を検知するから
2. なぜ利用者が増えるか?
増え続ける不正利用を自社で検知するのが困難なため
3. なぜデータが増えるか?
利用者が増えるほど口座明細が増えるため
4. なぜモデルの精度が向上するか?
口座明細が増えるほど幅広な不正のデータが手に入るため

他にどのようなベストプラクティスがあるでしょうか？

- データサイエンティスト・開発者の方から、チームで適用できそうなベストプラクティス事例や、自社内で実現済みの事例について紹介いただきます。

实践

実践

- プロダクトに応用できるベストプラクティス事例を2~6個リストアップしてください
- なぜ勝ちパターンが成立するのか、チーム内で議論してください。
- 上記の結果をシートにまとめ、結果を発表してください。

次回の応用編で、採用するベストプラクティスの優先順位付けと、自社で勝ちパターンが成立するかの確認を実施します。

理解編の復習

理解編の復習

Q: なぜ機械学習に取り組む必要があるのか?

A: 顧客体験を迅速に改善するため。

機械学習の勝ちパターン

顧客体験の改善が利用者の増加、利用者の増加がデータの増加、データの増加がモデルの精度向上、モデルの精度向上が顧客体験の改善につながるサイクルが回る。

機械学習の特性

データから帰納的に処理を実装する技術。複雑なロジックを高精度で解けるなどのメリットがあるが、説明性や精度100%が求められる状況には向かない。

TIME FOR A BREAK

See you in 5 minutes!

This material is based on the AWS Machine Learning Embark Program,
an immersive ML training program provided by Amazon Machine Learning Solutions Lab

