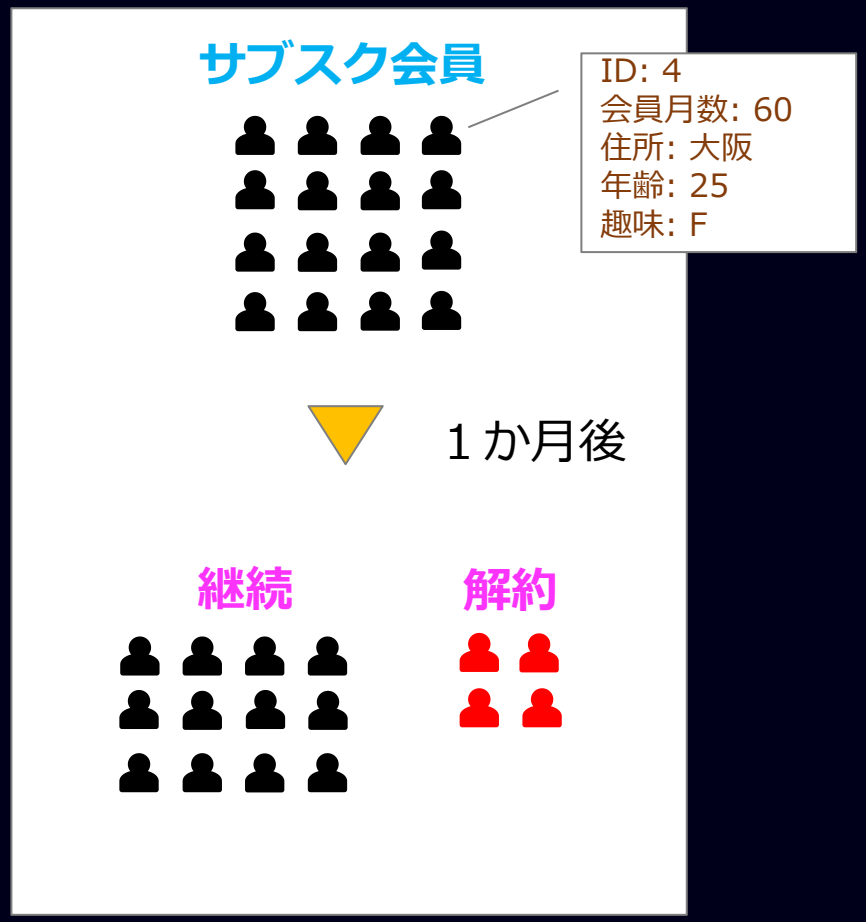
データサイエンスすいすい会

「アンケートと機械学習で効率的な顧客理解の
実践方法の共有」

ForecastFlow

表形式の自動機械学習（AutoML）とは

■ 一枚表さえ用意したら、自動で機械学習をさせることができるシステム



機械学習：　ターゲットを特徴量だけで予測すること

ID	ターゲット	特徴量１	特徴量２	特徴量３	特徴量４
1	Yes	120	東京	18	A
2	No	80	千葉	18	B
3	Yes	95	神奈川	20	B
4	Yes	60	大阪	25	F
5	Yes	86	東京	26	G
6	No	92	埼玉	22	C
7	No	105	兵庫	15	C
8	Yes	72	埼玉	19	D
9	Yes	88	東京	28	A

©GRI Inc.

本日の話題：　アンケートについて

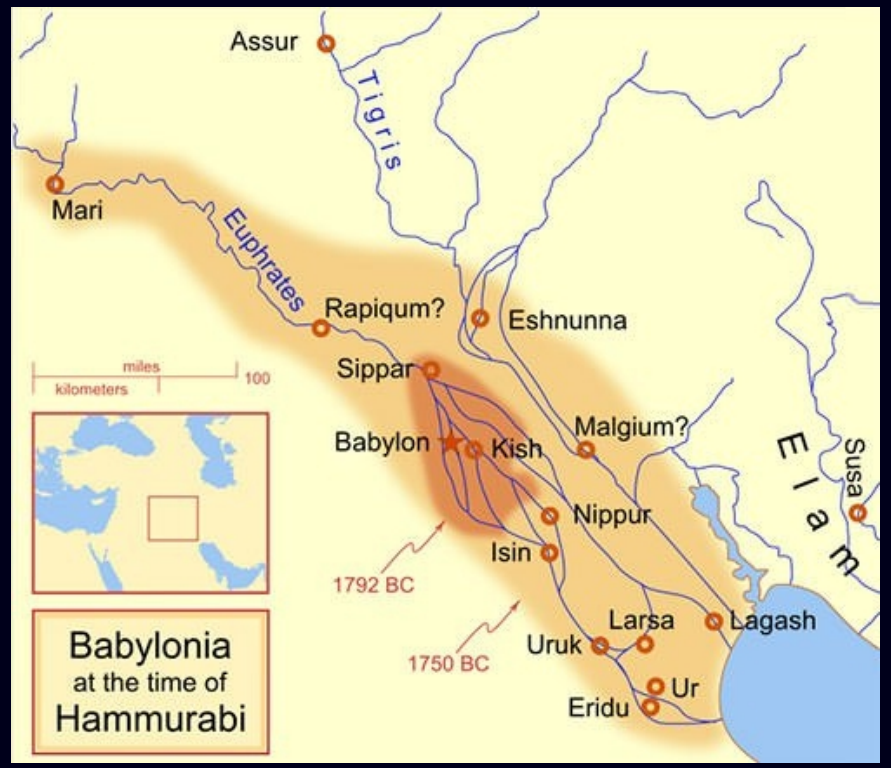
- アンケートは顧客の考えを理解できる

- ただしデータボリュームが少ないので、機械学習と相性が悪いと思われがち

©GRI Inc.

アンケートや国勢調査

■バビロン王朝時代の世界最古の顧客サービスの顧客クレーム表



一般的なアンケートの流れと使い方

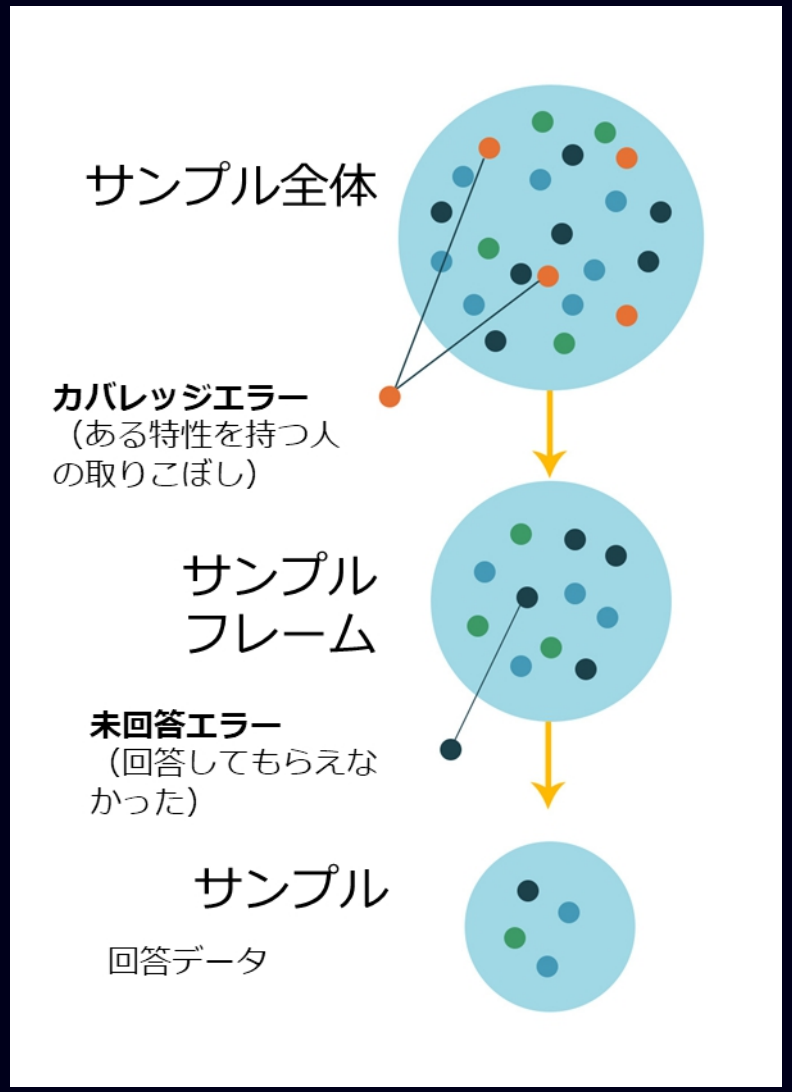
■アンケートを収集し、集計する

アンケートの難しさ

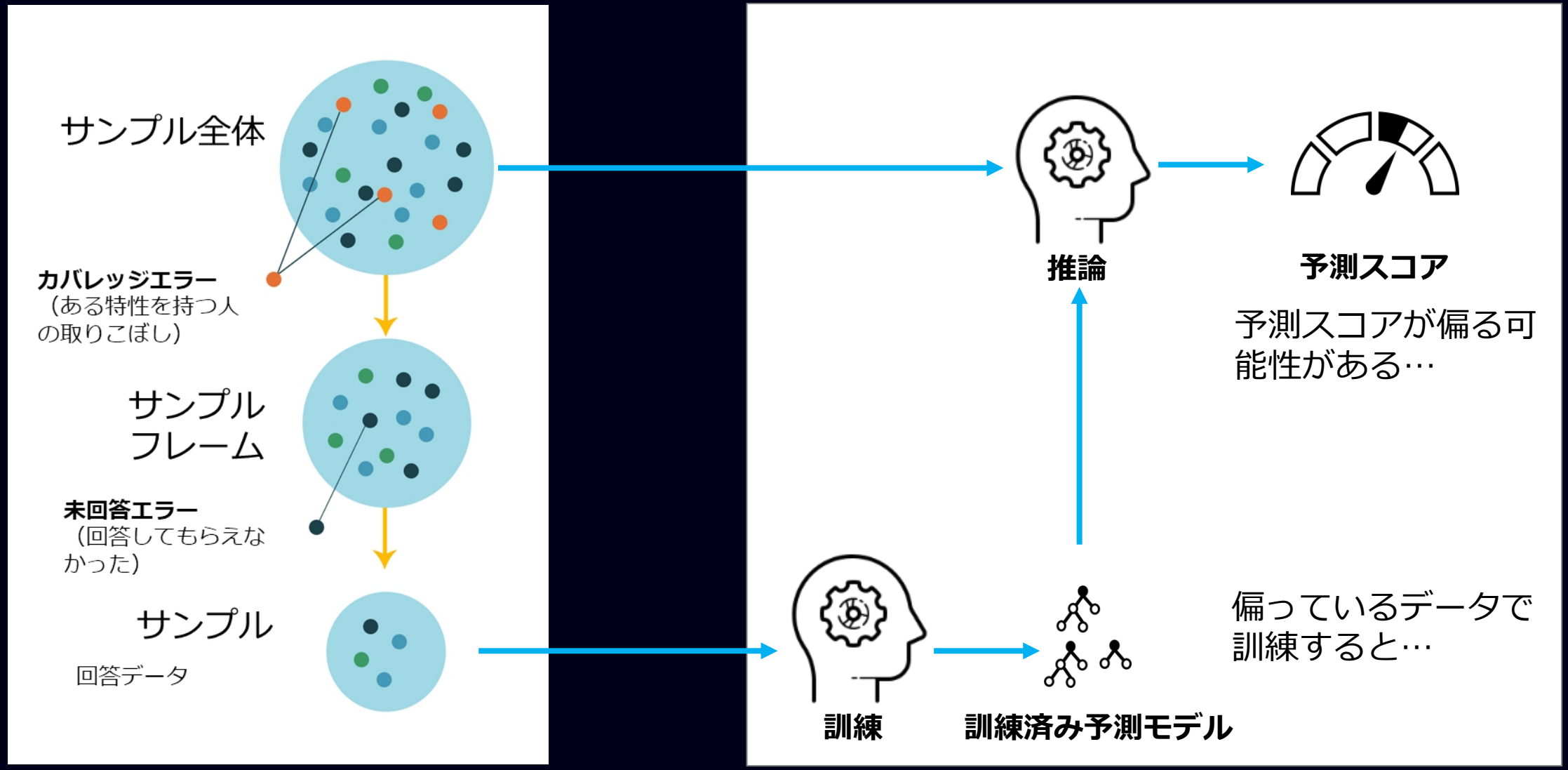
- 全数アンケート調査はコストがかかる

- サンプリングしようと、フレームを作ると
  - その時点で偏りの可能性

- サンプリングを実施すると
  - 未回答という偏りの可能性



アンケートと機械学習のチャレンジ（概算値として許容）



アンケートと機械学習を使った例

①アンケートを取ってない人に適用する

- 広告セグメント配信のカスタム・セグメント作りの事例

②顧客のテースト理解

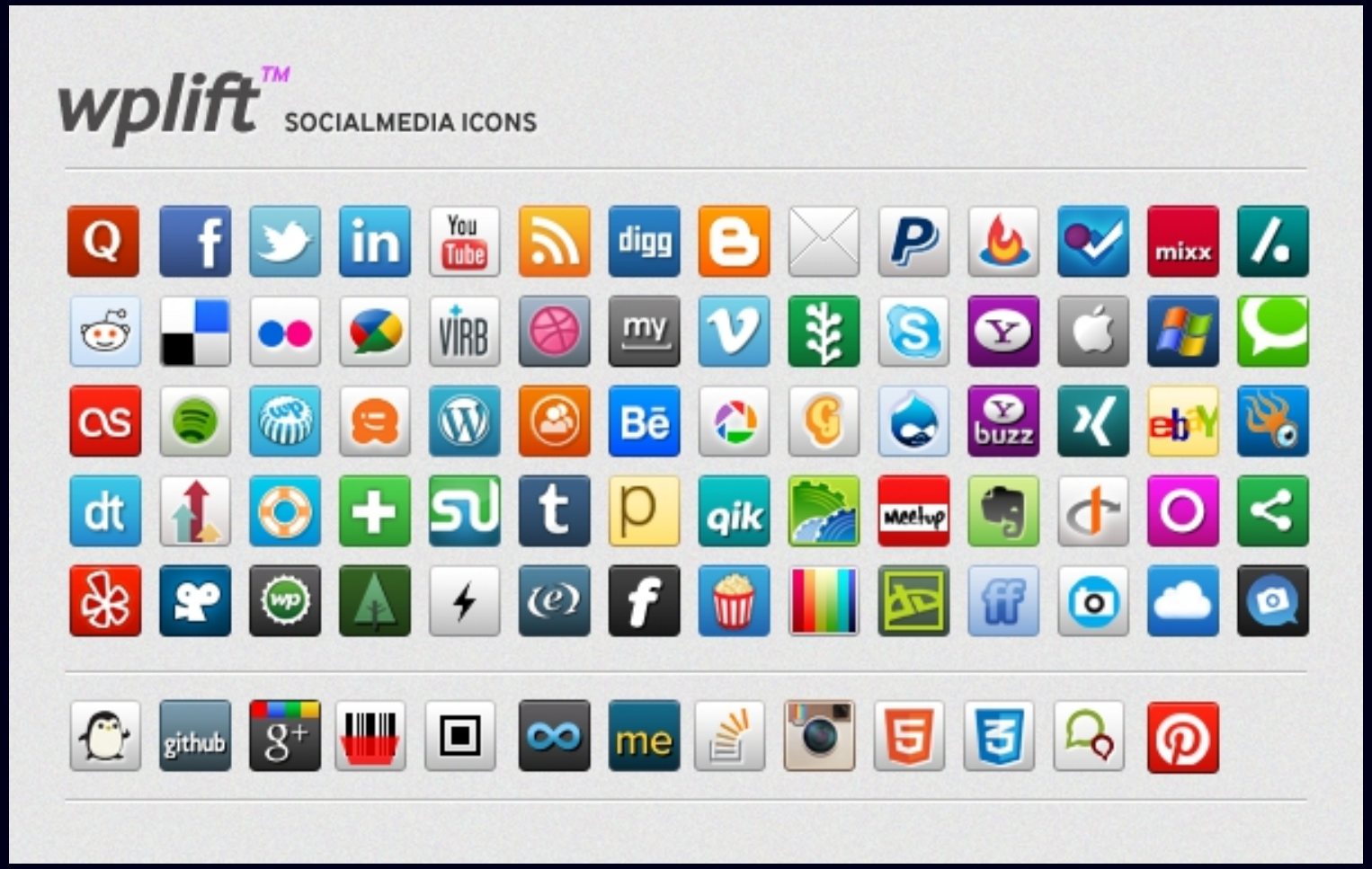
③アンケートを予測モデルの特徴量に加える

- サブスクの解約予測にアンケート結果を含める事例

©GRI Inc.

事例１：　広告セグメント配信のカスタム・セグメント作りの事例

■メディアサイトの広告営業を想定



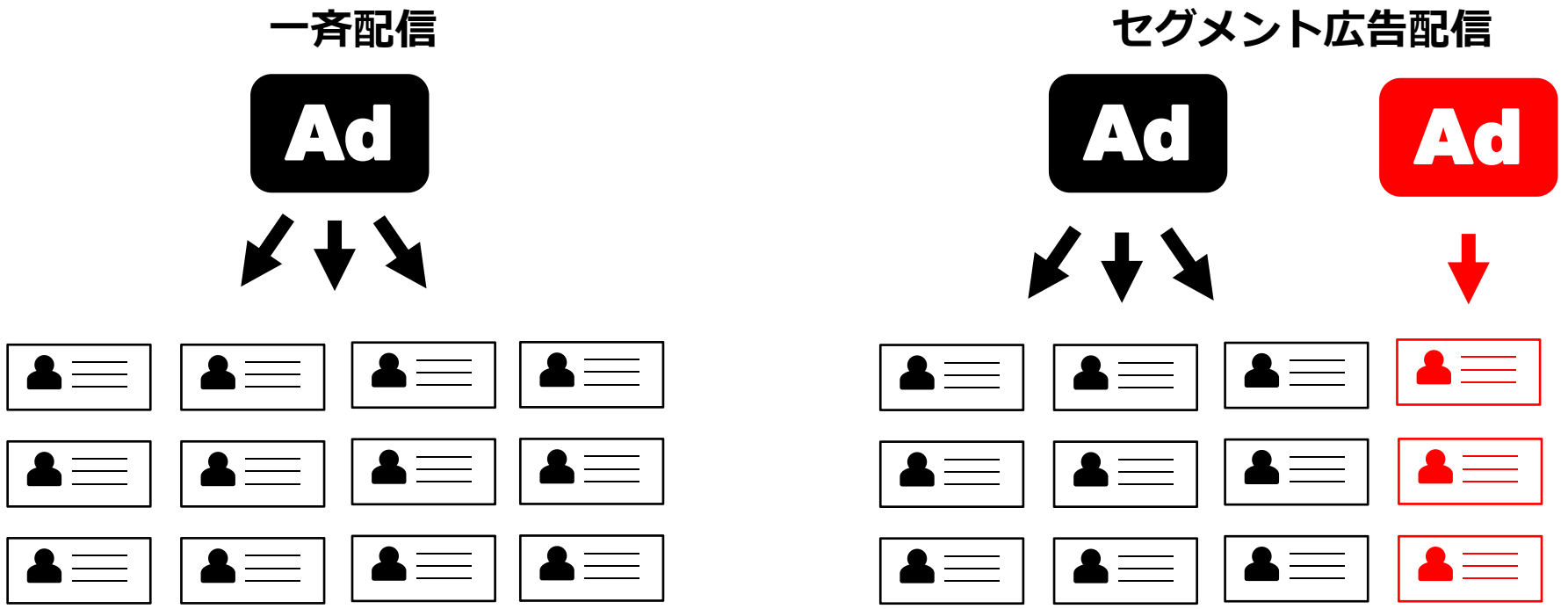
広告セグメント配信のカスタム・セグメント作りの事例

■純広告枠を売る

セグメント広告配信

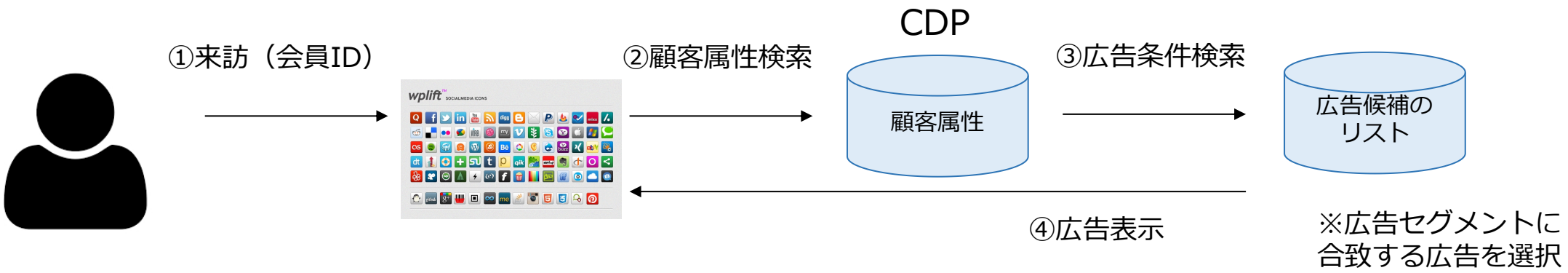
■各広告の訴求するターゲット・セグメントを選んで配信できれば、消費者は広告に反応しやすくなり、出稿主にとって価値のある広告配信になる

1st Partyデータ（会員ID、Cookie）と
そこに紐付くセグメントデータを利用



セグメント広告配信の流れ

1st Partyデータ（会員ID、Cookie）と
そこに紐付くセグメントデータを利用



典型的なセグメント

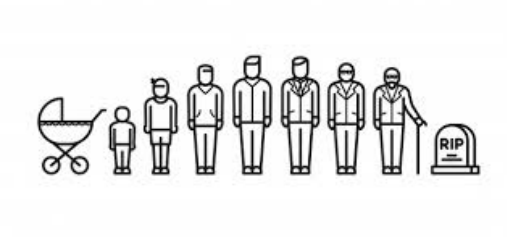
性別

収入

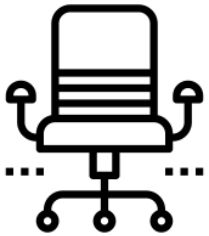
趣味・嗜好

意識

年齢層



職業／職位／業種

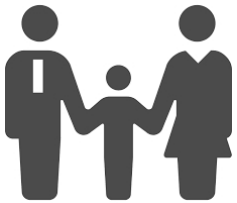


ライフスタイル

価値観

居住地

家族



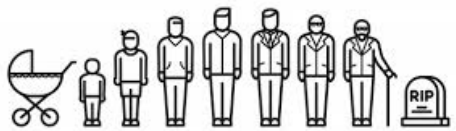
©GRI Inc.

典型的なセグメント

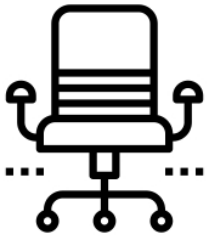
性別

収入

年齢層

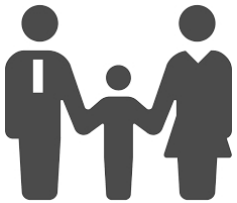


職業／職位／業種



居住地

家族



趣味・嗜好

ライフスタイル

メディアの閲読行動

意識

価値観

取得の難しいデータ

©GRI Inc.

純広告のセグメント配信の難しい所

クライアント

依頼 →

← 提案

メディア
広告部門

新製品（車）の良さを広めたい

車の記事を閲読している人に
配信しましょうか？

「走りを忘れた大人たちへ」
訴求したい

・・・

メディアは必ずしも広告主の訴求内容にぴったりのコンテンツを持っているわけではない
ぴったりのセグメントを持っていない

どんな要望にも応えるカスタムセグメントを作れないだろうか？

- 仮説：　アンケートをベースにすれば、カスタムセグメントを作れないだろうか

- 悩み：　アンケートを回収できる人数は限られてしまう

©GRI Inc.

どんな要望にも応えるカスタムセグメントを作る

■ ４つの要素技術を組み合わせ、新しいカスタム・セグメントを効率的に作成できる

セグメント配信

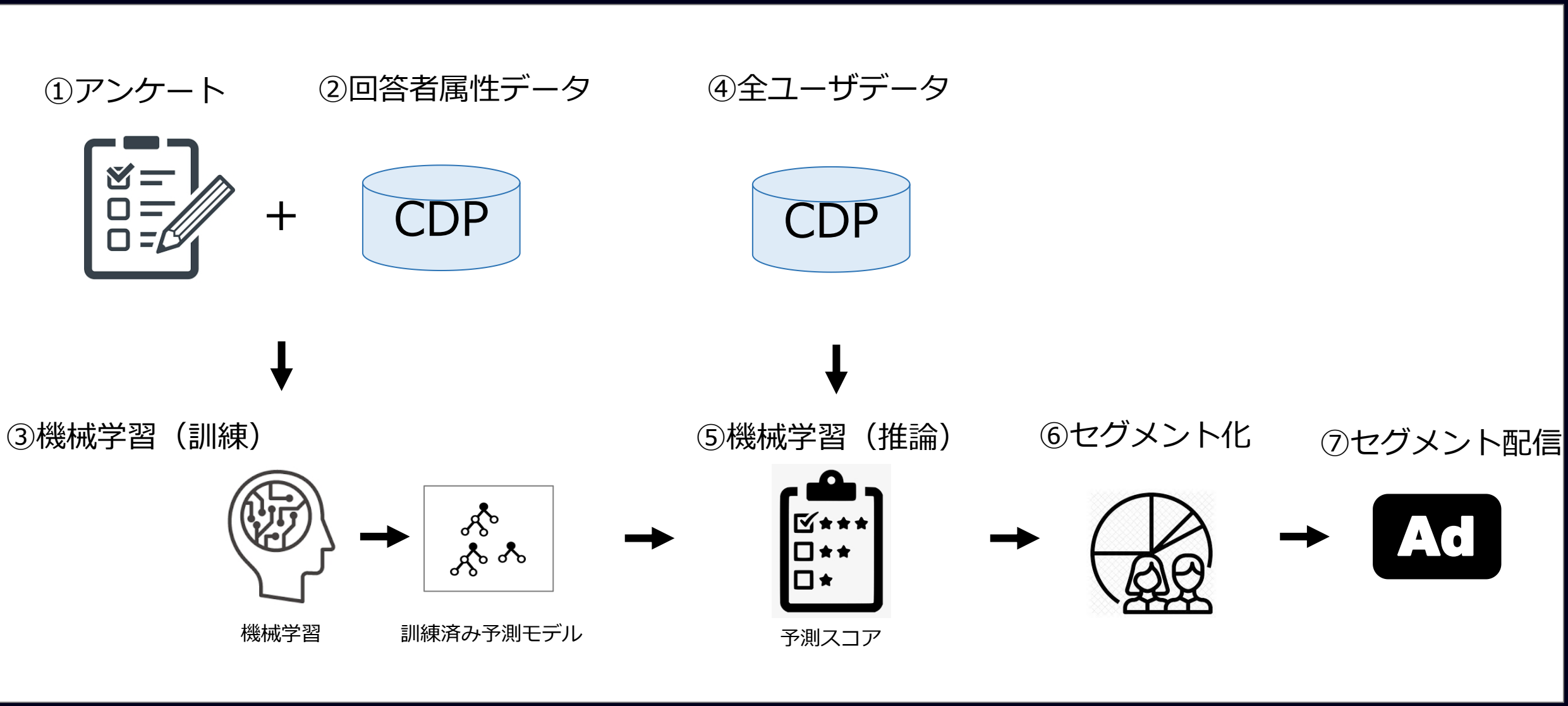
アンケート

CDP
（顧客データ基盤）

AI（機械学習）

©GRI Inc.

カスタムセグメント

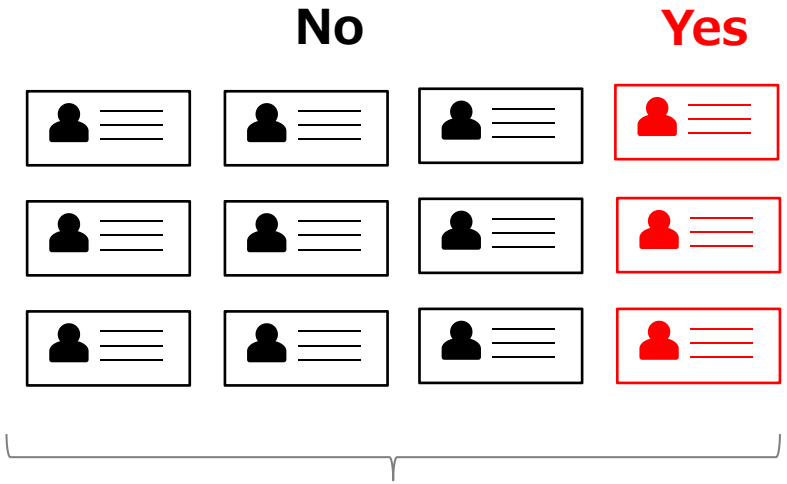

①アンケート
②回答者属性データ
④全ユーザデータ
+
CDP
CDP
③機械学習（訓練）
⑤機械学習（推論）
⑥セグメント化
⑦セグメント配信
Ad
機械学習
訓練済み予測モデル
予測スコア

アンケートから教師データを作成

メディアの会員に対して個別のアンケートをする

「走りを忘れた大人たちへ」
訴求したい

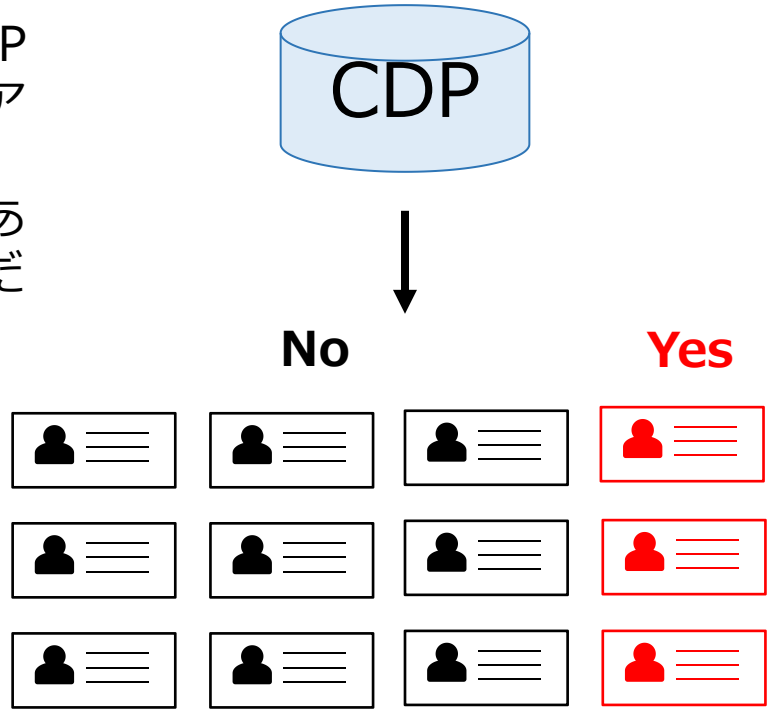
たいくつな毎日を
過ごしてしまっている日が多い



これを機械学習のターゲット（教師データ）にする

アンケートから教師データを作成

- アンケート回答者の属性情報をCDP（顧客データ基盤）より取り出しアンケート結果と結合する
- 属性情報がメディア独自のものであれば、メディアの個性を組み込んだ独自セグメントに成り得る



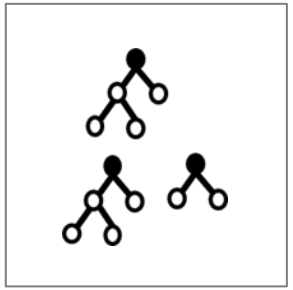
ID	Q1	性別	年齢層	都道府県	属性１
ID001	1	男性	20代	東京都	A
ID002	0	女性	30代	大阪府	B
ID003	0	女性	20代	北海道	A

©GRI Inc.

AutoML（ForecastFlow ）の利用シーン

- アンケート＋属性情報の一枚表から、機械学習を使ってターゲットと特徴量の関係を学習（訓練）する
- 訓練すると、訓練済み予測モデルができあがる

訓練



訓練済み予測モデル

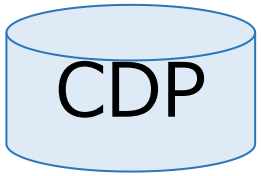
ターゲット　　特徴量

ID	Q1	性別	年齢層	都道府県	属性１
ID001	1	男性	20代	東京都	A
ID002	0	女性	30代	大阪府	B
ID003	0	女性	20代	北海道	A

©GRI Inc.

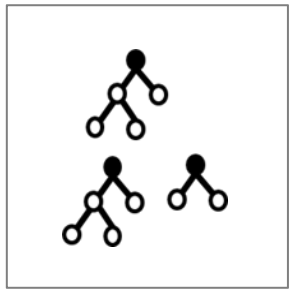
AutoML（ForecastFlow ）の利用シーン

- アンケート＋属性情報の一枚表から、機械学習を使ってターゲットと特徴量の関係を学習（訓練）する
- 訓練すると、訓練済み予測モデルができあがる



ID	Q1	性別	年齢層	都道府県	属性 1
ID991		男性	20代	東京都	A
ID992		女性	30代	大阪府	B
ID993		女性	20代	北海道	A

ID	予測スコア
ID991	0.45
ID992	0.98
ID993	0.32

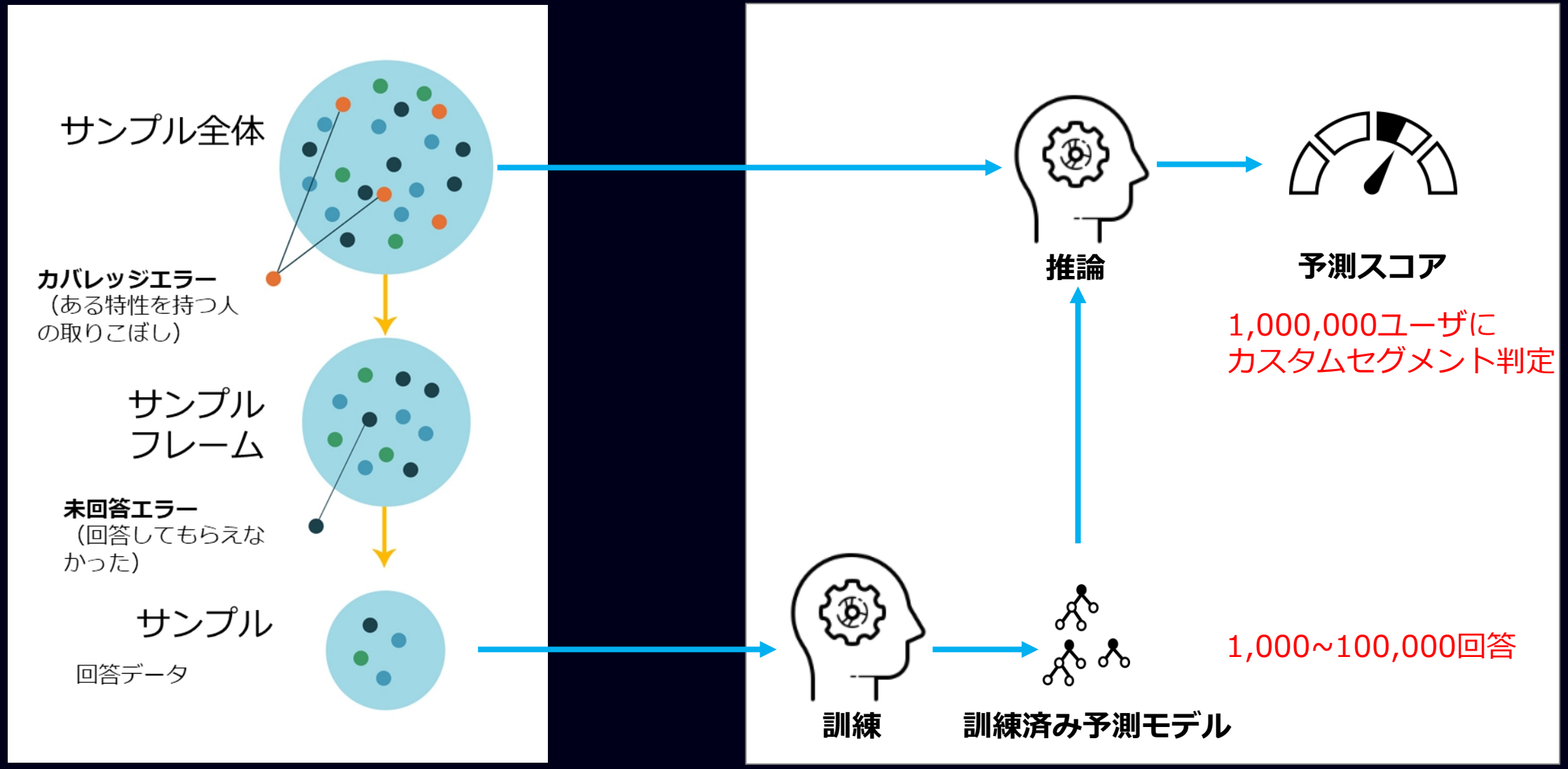


訓練済み予測モデル

例）カスタムセグメントを
スコア0.6以上にする

©GRI Inc.

AutoML（ForecastFlow ）の利用シーン



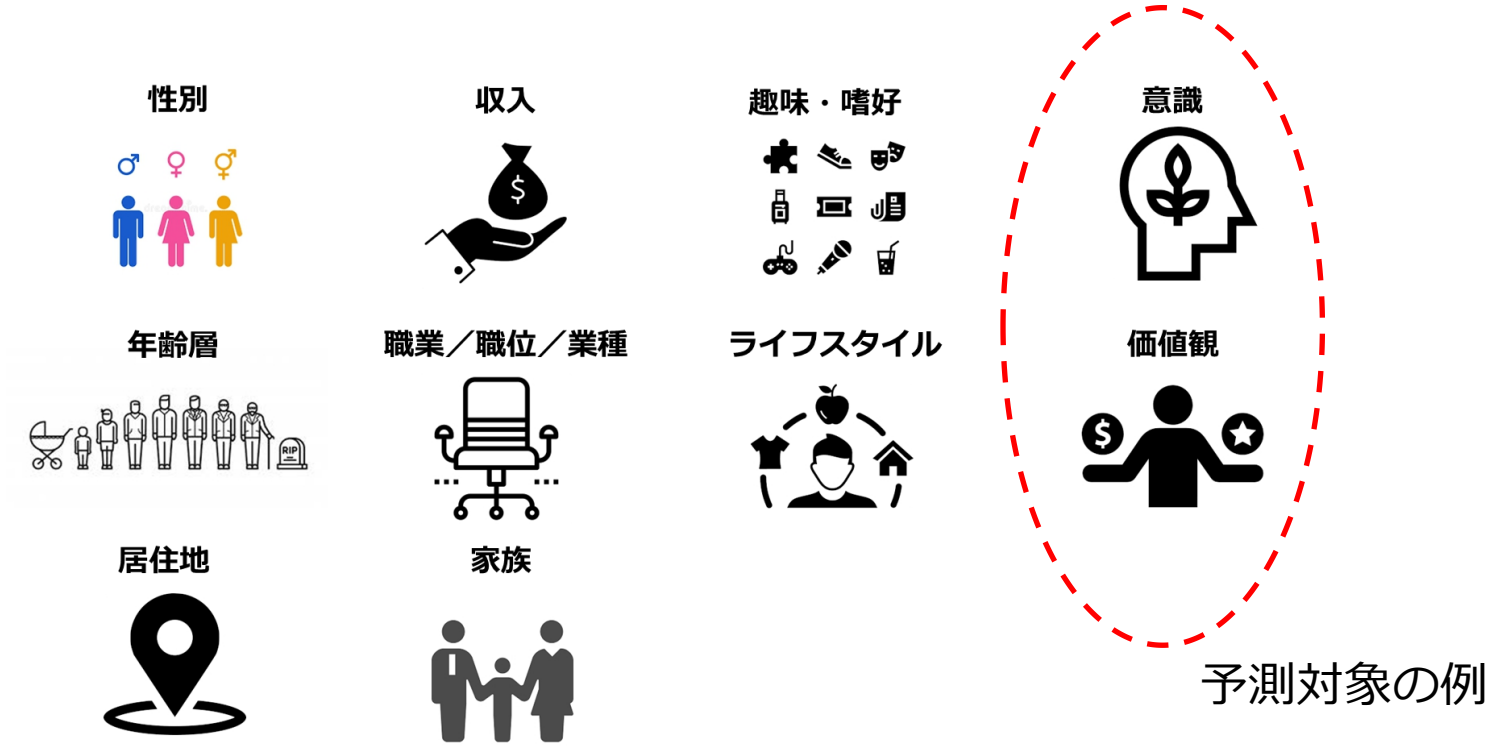
広告セグメント配信のカスタム・セグメント作りのポイント

- 広告出稿主の訴求対象者を切り分ける良い設問を思いつく
  - 広告代理店や調査会社のマーケター的な人
- データ基盤、アンケート配信、自動機械学習の基盤、及びセグメント配信の基盤とのシームレスな接続
- 自社メディアの閲読などの独自の行動情報が広告出稿主へ説得力を持つ
- 上手くセグメントできていたかの判別が難しく、注意深く効果検証を行う

©GRI Inc.

その他の予測パターン

■セグメントの切り口で情報としてないものは、アンケートと組み合わせて設定ができる



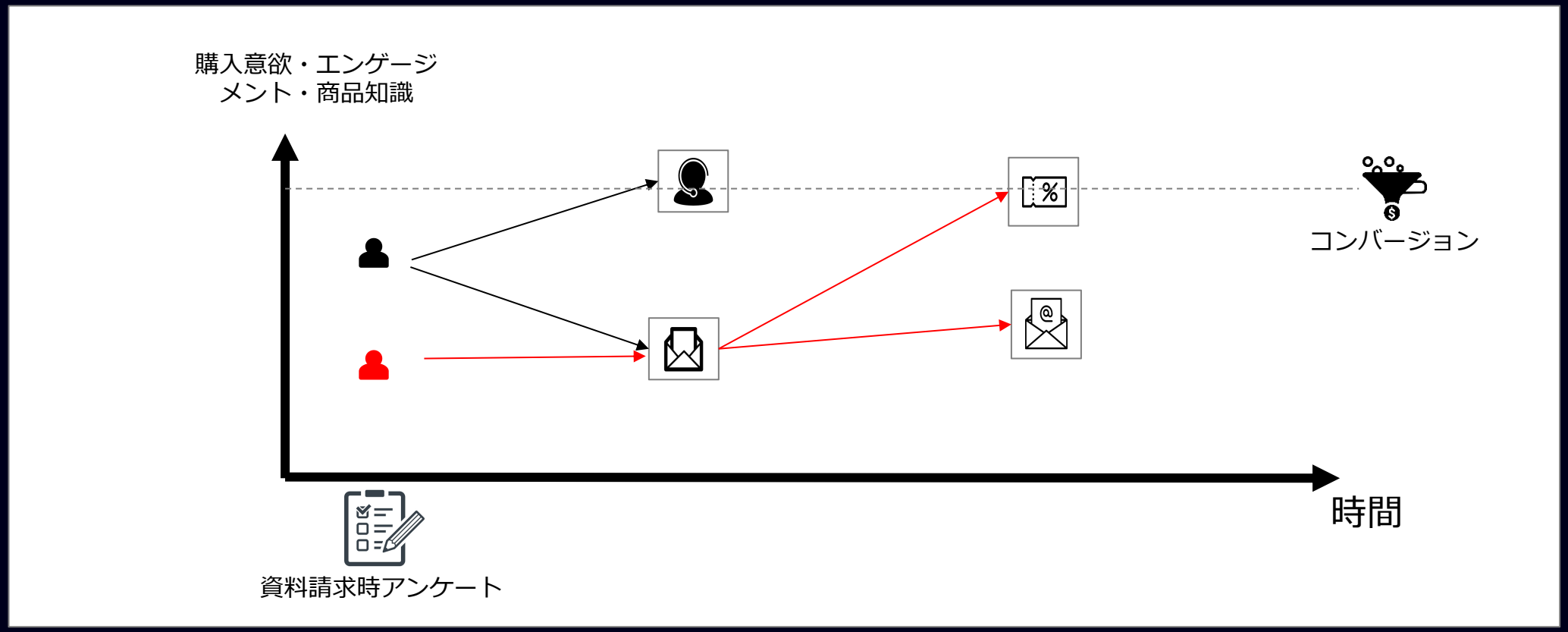
事例２：　顧客のテーストを知る

■シナリオ：　リードナーチャリング

- 最初の接点でアンケートを取得
- コンバージョンに至るまでに、どんな施策をすべきだろうか？

顧客のテーストを知る（シナリオ：　リードナーチャリング）

- 最初の接点でアンケートを取得
- コンバージョンに至るまでに、どんな施策をすべきだろうか？



顧客のテーストを知る（問題：　コンバージョンするかを２つのモデルで説明）

特徴量

- 資料請求時アンケート
- 各種施策の内容
- 各種施策の担当者
- 施策実施までの時間
- 施策実施順序

事前的予測モデル

事後的予測モデル

ターゲット

- コンバージョンしたか？

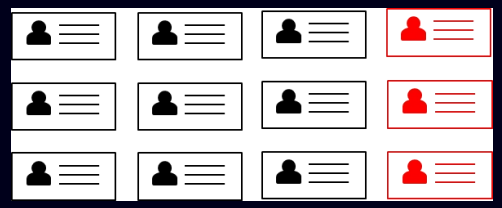
顧客のテーストを知る（アンケートと機械学習でできること）

資料請求時
アンケート

＋

事前的予測モデル

→



施策実施対象ユーザの選定

資料請求時
アンケート

＋

事後的予測モデル

→

施策全体の振り返り
What-if分析

事例３：　アンケートを解約予測モデルに加える

- サブスクの解約予測モデルの特徴量にアンケート結果を加える例
  - サブスクの解約予測モデルに関しては前回のすいすい会の動画を参照
  - https://www.youtube.com/watch?v=ppkgYD69NDU

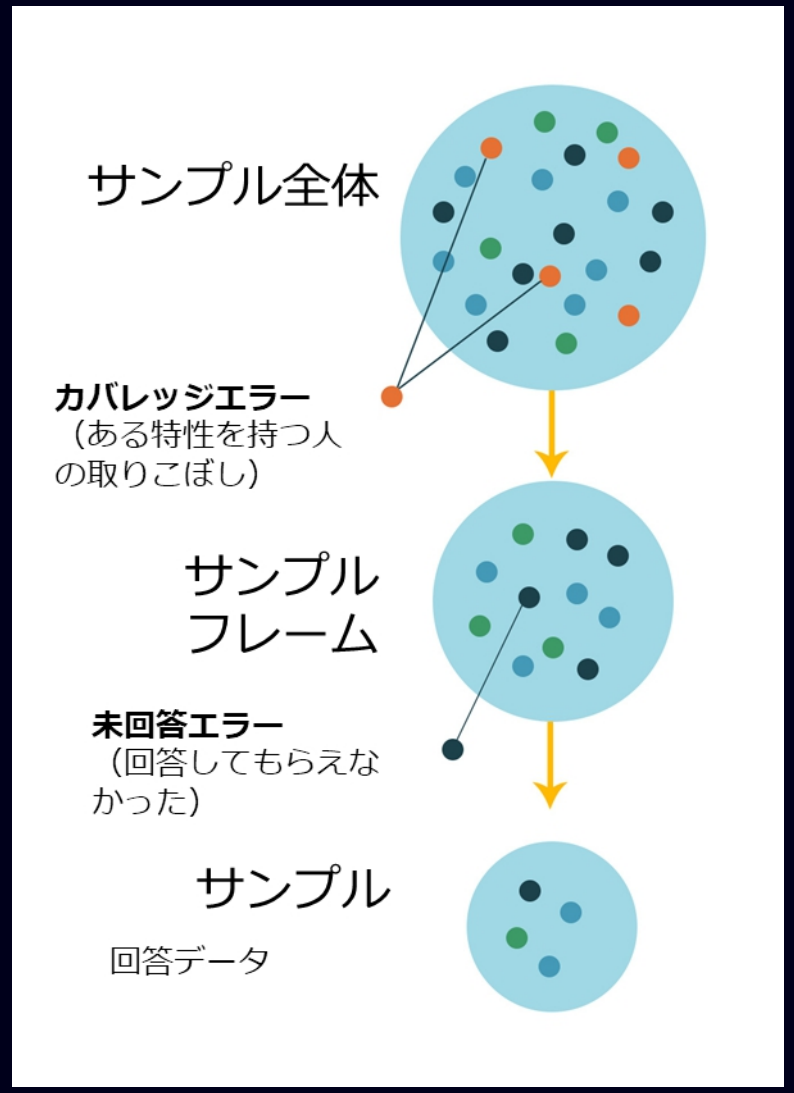
- 仮説
  - アンケートで意識や価値観に迫れるので、サブスクの解約行動の説明に良い示唆が得られる

©GRI Inc.

アンケートを予測モデルに加える（難易度が高い）

- アンケートの回答者は少ないので、解約予測モデルのデータ数が不足する可能性がある

- アンケートに回答している人がロイヤルカスタマーである場合もあり、回答している時点でバイアスを持っている場合もある



まとめ

- アンケートと機械学習の組み合わせの事例を紹介

©GRI Inc.

残っている質問（今後、説明していきたい内容）

- その他のビジネスの実践例を知りたい
- どのくらい正解データの数があるべきか
- 不均衡データの扱い
- 様々な自動機械学習ツールの比較

©GRI Inc.