

電子書籍ストアにおけるデータウェアハウスの利活用と Amazon Redshift MLを使ったビジネス施策への取り組み

2022/03/10

株式会社ブックリスタ

デジタルマーケティング部 データアナリスト 滝澤 俊哉

デジタルマーケティング部 データアナリスト 小田 友莉乃



会社概要

社名：株式会社ブックリスタ

設立：2010年7月1日

代表取締役社長：村田 茂



事業内容

- (1) 出版・新聞コンテンツの制作、販売および仲介事業ならびに出版業
- (2) 出版・新聞コンテンツの管理・プロモーションに必要なシステムの企画、開発、構築、提供
- (3) 情報システムの企画、設計ならびに管理運営に関する業務の受託
- (4) 前各号に付帯関連する一切の業務

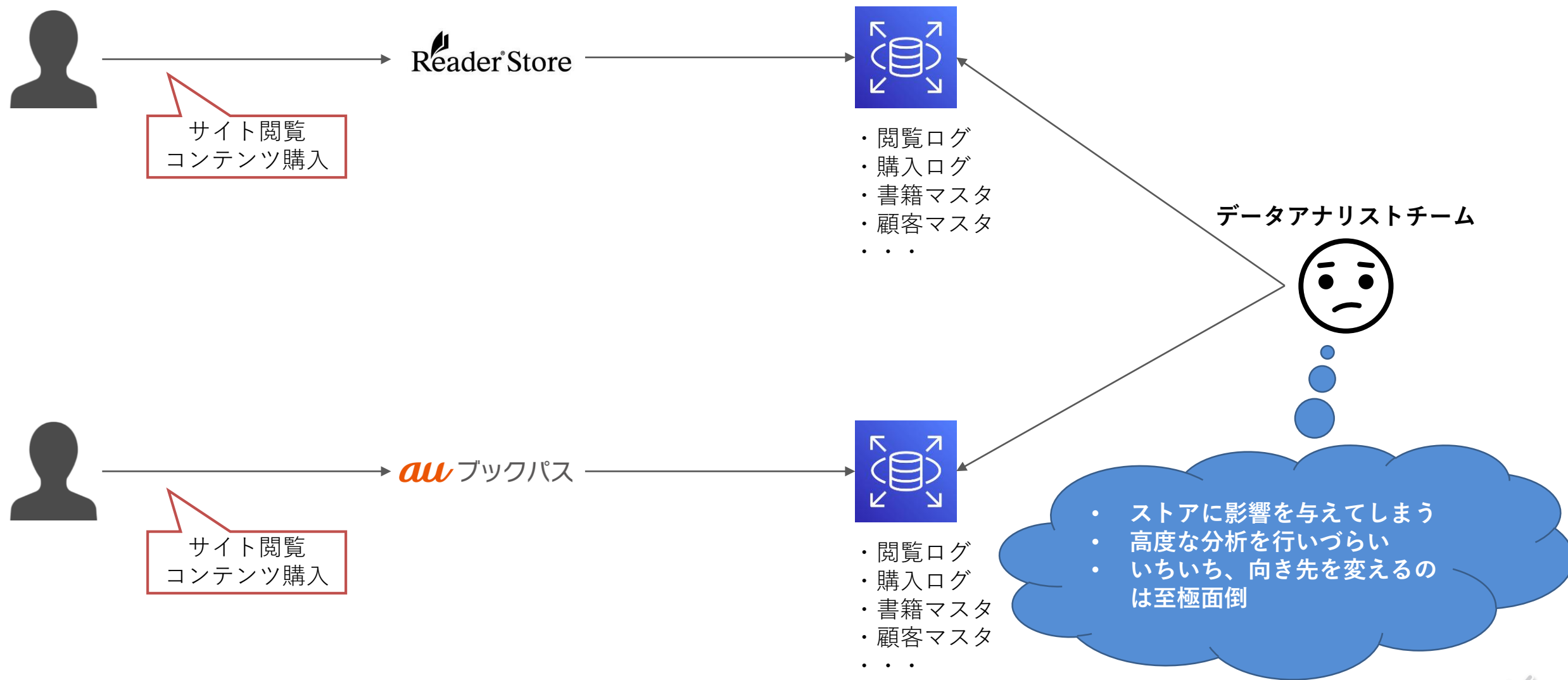


ブックリスタは、電子書籍に関する各種事業を展開し、電子書籍ストアの総合的な運営のサポートをしています。同時に「エンタメ×テック（知的好奇心×感動体験）」のvisionの元、既存事業の枠にとらわれない新規事業の創造に取り組んでいます。

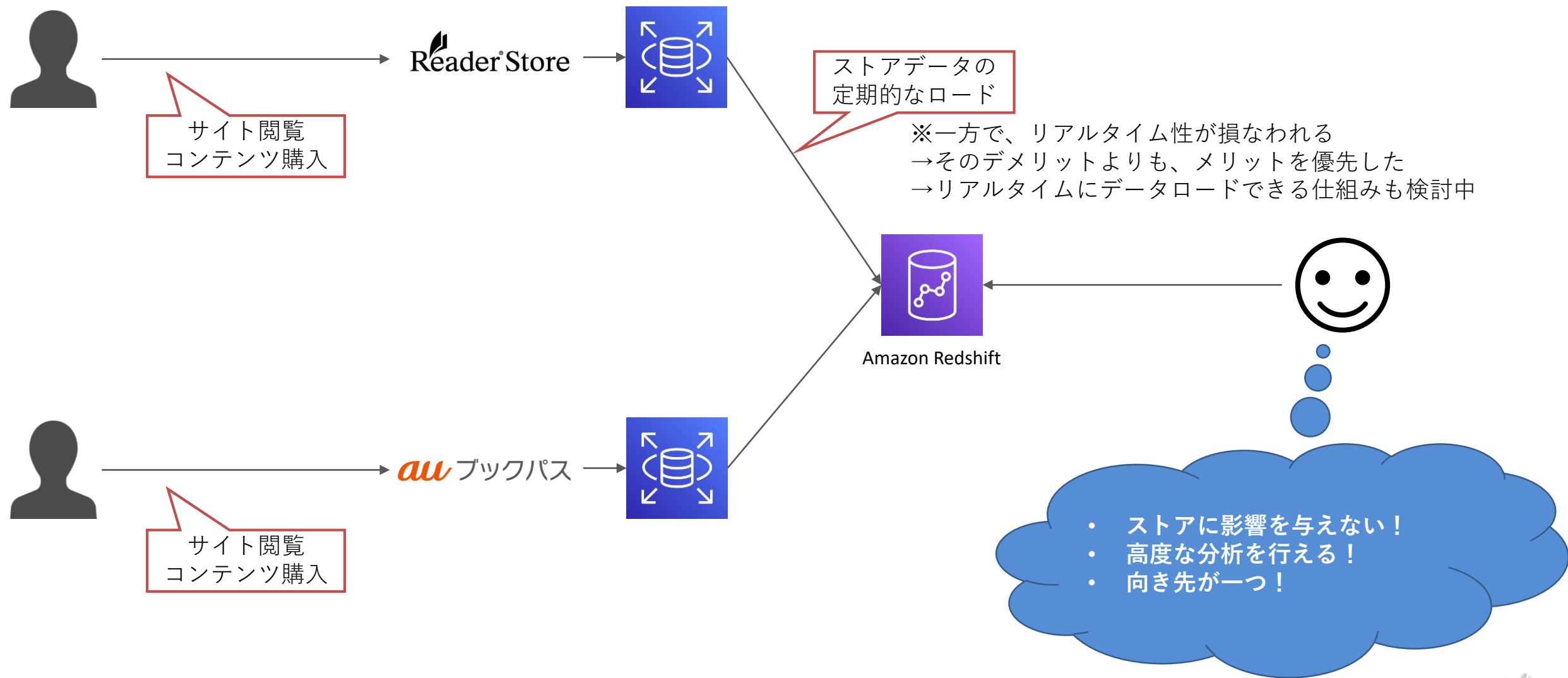




データウェアハウスの構築



データウェアハウスの構築



データウェアハウスの利活用



全社員

- レポートが欲しい
- 行った施策の効果が知りたい



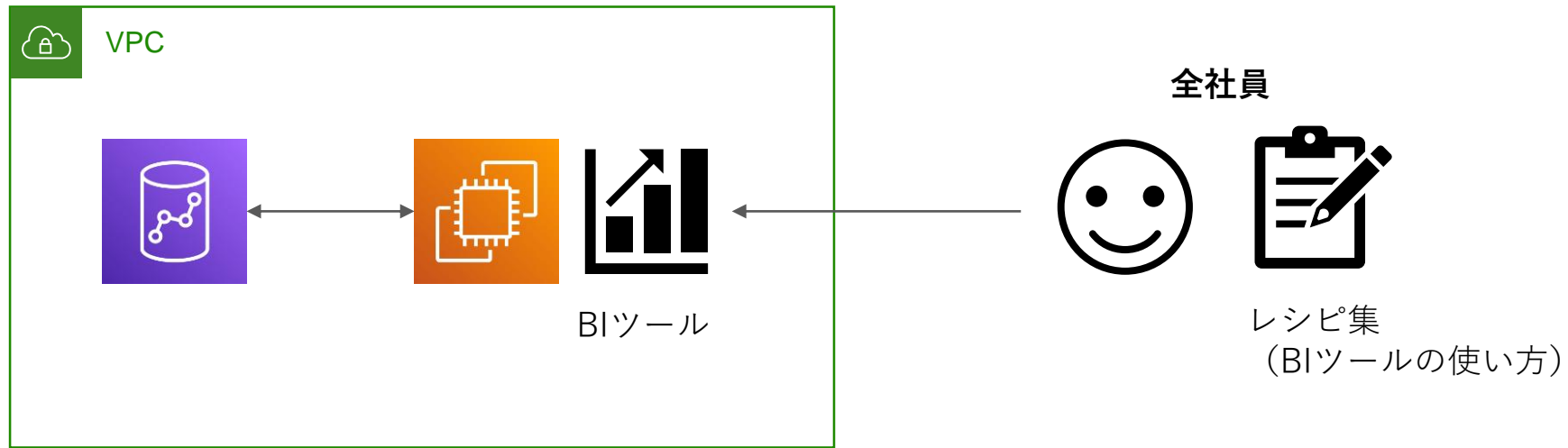
データアナリストチーム

- 予実管理などのストアレポートの作成
- 施策対象者の抽出/実施した施策の効果検証

もっとデータを使いたい

仕事を与えられるだけではなく、
価値を生み出していきたい





データ民主化レクチャー会

The video player displays a Google Sheet titled "データ民主化レシピ集" (Data Democratization Recipe Book). The sheet contains a table with the following content:

対象	共通	レシピ名	データを集めてみよう	概要
表示されているデータを集計して、表示してみます 例として基本操作No.3で作成したデータを用いて、日次の売上金額を集計してみます				

Overlaid on the video are two large, semi-transparent text boxes: "レシピ集" (Recipe Book) on the left and "BIツール" (BI Tool) on the right. The video player interface at the bottom shows a play button, a progress bar at 16:23 / 50:21, and a volume icon. A small video feed of the speaker, 竜澤俊哉 (Ryūzawa Shūya), is visible in the bottom right corner of the video player.



これからは！



全社員

- ・ 自分でレポートが作れる！
- ・ 行った施策の効果を自分で知れる！



データアナリストチーム

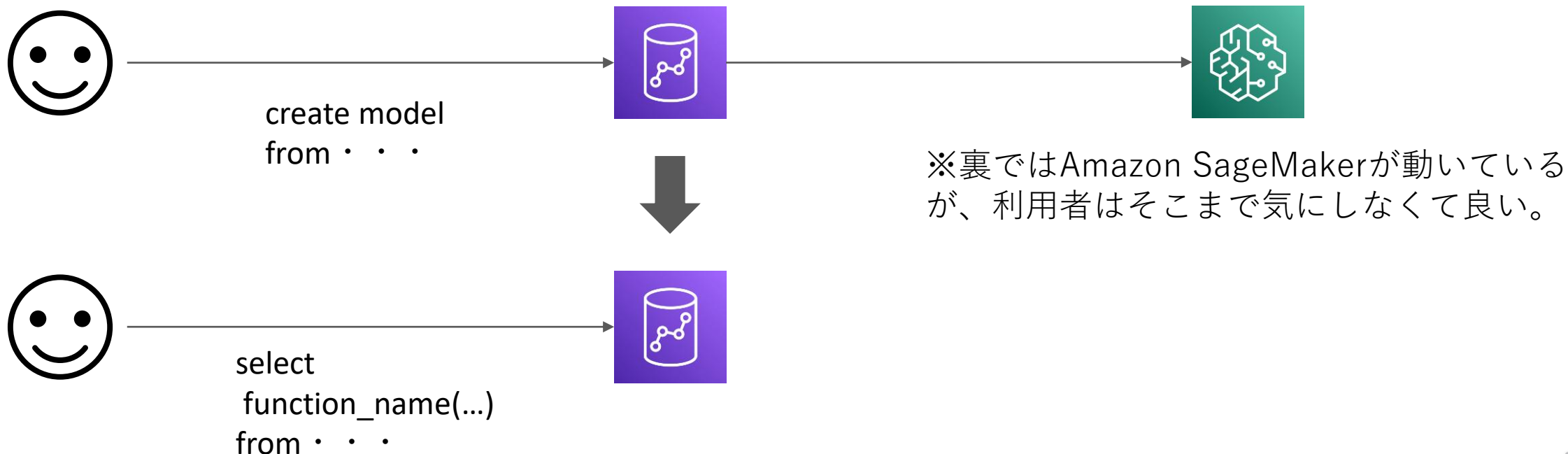
- ・ 全社員へのサポート
- ・ アナリストチームからの施策提案

データ民主化の実施

Amazon Redshift MLを用いた施策提案



- Amazon Redshiftを用いて直接、機械学習モデルの作成、トレーニング、デプロイを実施できる
- SQLをメインで使っているデータアナリストでも学習コストがそこまで高くない



従来の施策概要

- 期間限定ポイントのプレゼント(付与)
- 目的：購買促進、稼働人数増加
- 付与条件：付与時点までに同月1冊も購入していないユーザー

問題

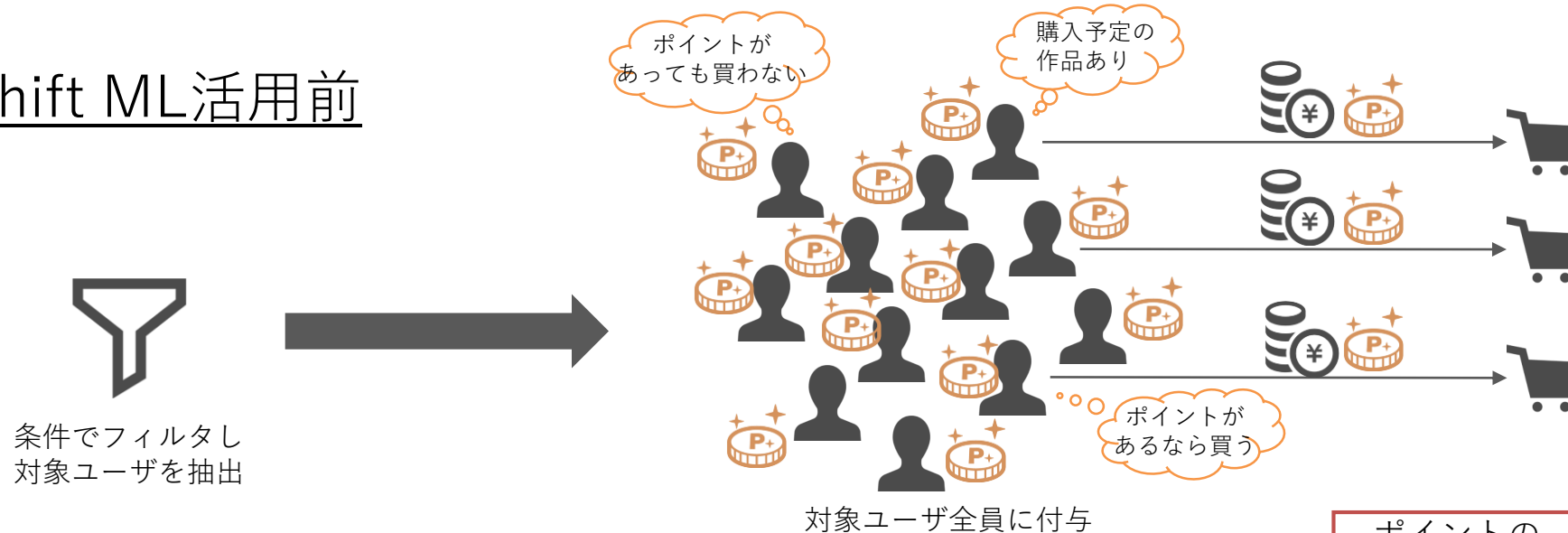
- 付与しなくても購入する人も付与対象(無駄な経費)
- 使用率に幅があるため予算の管理が困難
 - 例えば… 5万人に100p付与
 - 50%が使えば250万円、1%が使えば5万円



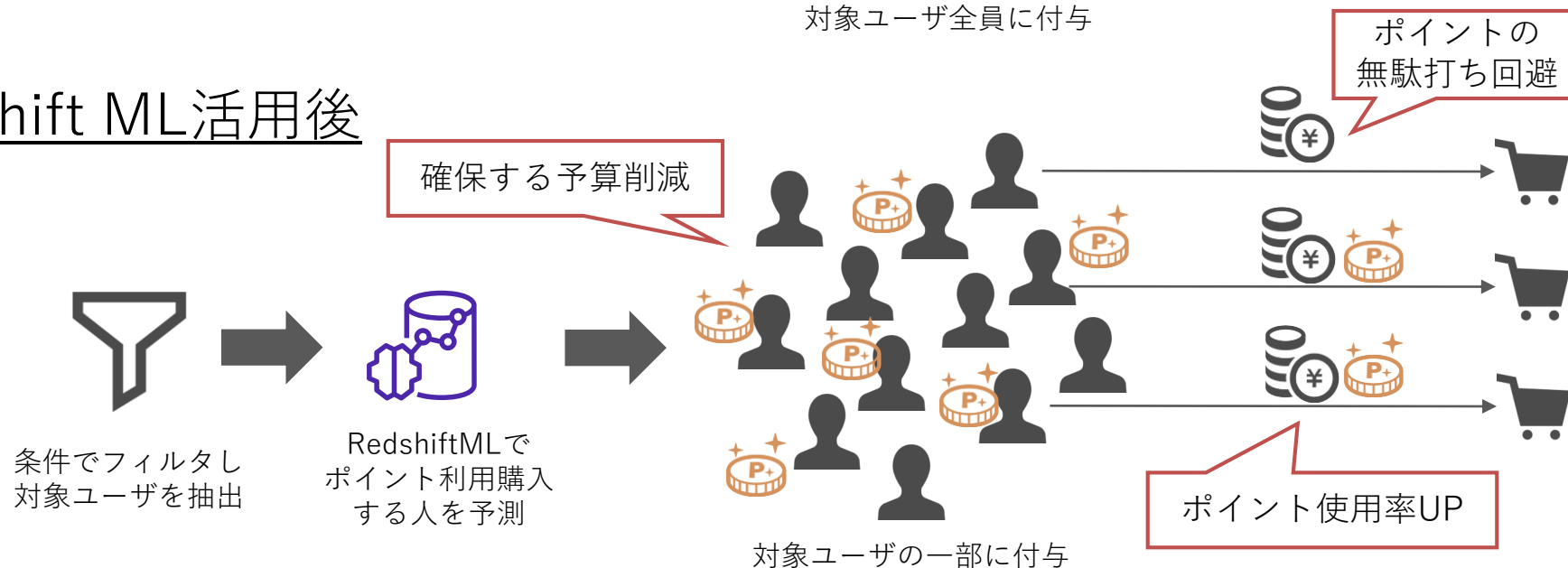
⇒ **ポイントを使って購入してくれるユーザーを予測できれば解決するのは!?**

Redshift ML活用で見込まれる効果

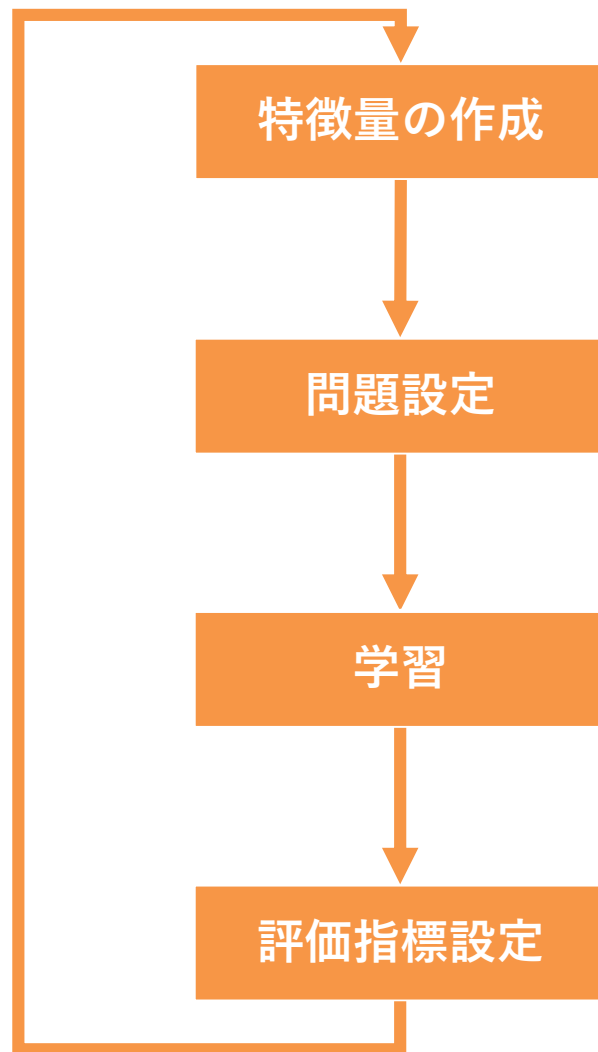
Redshift ML活用前



Redshift ML活用後



学習の外側を人間が決定

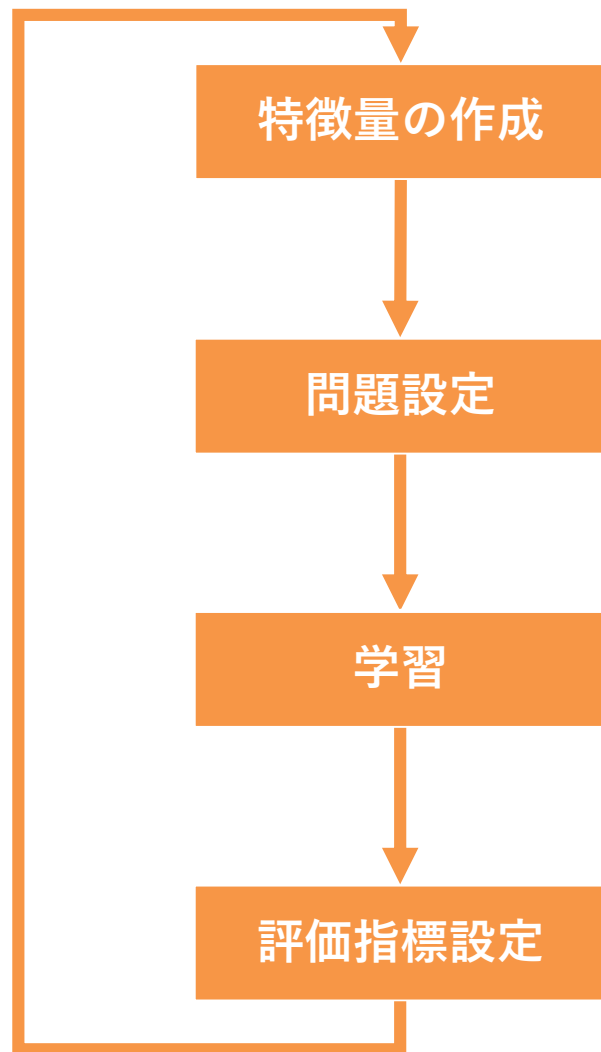


「ポイントを使う人」とはどんな人？

どんな値で表現できるのかを考え、Redshift MLに入力する情報を決定
Ex. 累計購入金額、ポイント使用額、購入冊数

← **Redshift MLにお任せ！**

学習の外側を人間が決定



「ポイントを使う人」とはどんな人？

どんな値で表現できるのかを考え、Redshift MLに入力する情報を決定
Ex. 累計購入金額、ポイント使用額、購入冊数

分類問題 (0/1) ? 回帰問題(実数) ?

データの偏りが激しく分類問題は機能しない…
回帰問題に設定し、予測結果を閾値処理する方針に決定

Ex. 1/1000がポイントを使用するデータ

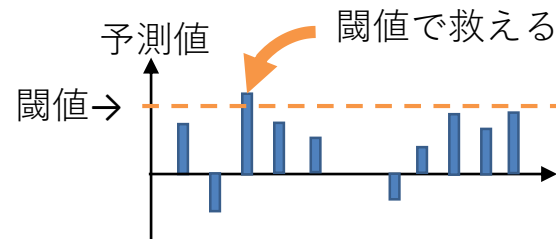
分類問題の場合

正解 0 0 1 0 0 ... 0 0 0 0 0
予測 0 0 0 0 0 ... 0 0 0 0 0

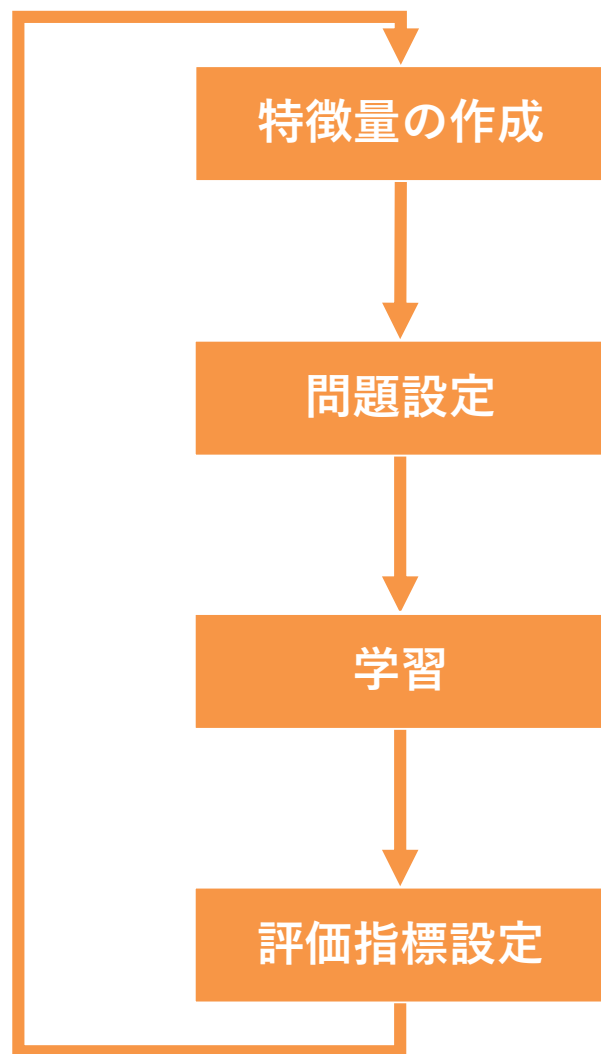
全てゼロと予測 ⇒ 正答率 99.9%

回帰問題の場合

正解 0 0 1 0 0 ... 0 0 0 0 0



学習の外側を人間が決定



「ポイントを使う人」とはどんな人？

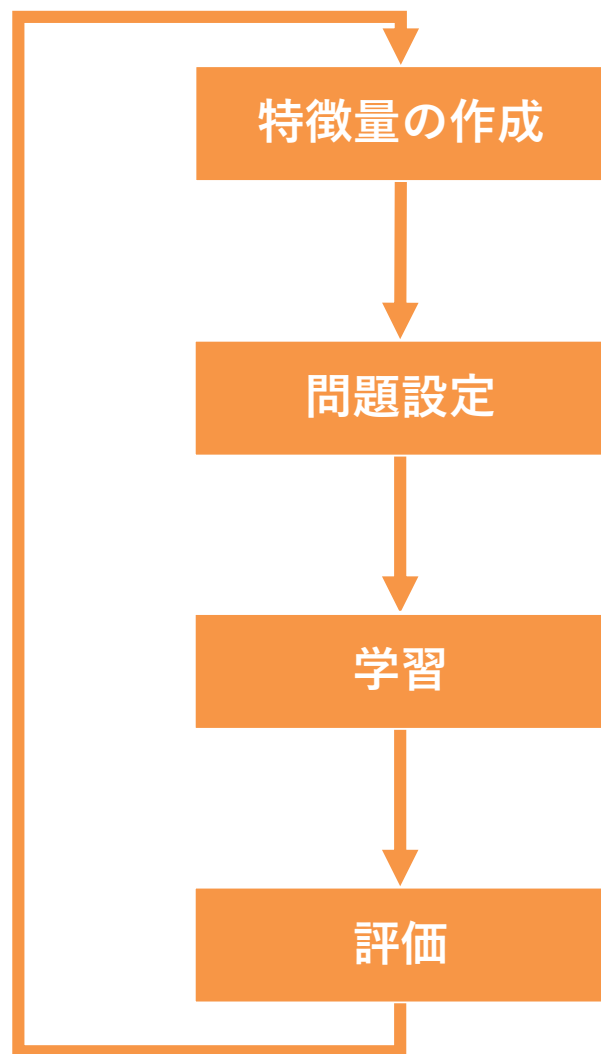
どんな値で表現できるのかを考え、Redshift MLに入力する情報を決定
Ex. 累計購入金額、ポイント使用額、購入冊数

分類問題 (0/1) ? 回帰問題(実数) ?

データの偏りが激しく分類問題は機能しない…
回帰問題に設定し、予測結果を閾値処理する方針に決定

← **Redshift MLにお任せ！**

学習の外側を人間が決定



「ポイントを使う人」とはどんな人？

どんな値で表現できるのかを考え、Redshift MLに入力する情報を決定
Ex. 累計購入金額、ポイント使用額、購入冊数

分類問題 (0/1) ? 回帰問題(実数) ?

データの偏りが激しく分類問題は機能しない…
回帰問題に設定し、予測結果を閾値処理する方針に決定

← **Redshift MLにお任せ！**

目的に合った評価値は？

稼働してくれる人を落とさずに、無駄な付与数を極力減らしたい
⇒ 再現率を重視し、その中で適合率を上げる

テスト内容

- A群：従来通りのポイント付与
- B群：Redshift MLによる購買予測を用いたポイント付与

※稼働率：各郡におけるポイントの有効期間内に購入した人の割合。ポイント使用は問わない

結果

- ポイント使用率が想定より低い
- 想定使用率と結果の乖離具合がユーザ層で変化
- 稼働したユーザの内、ポイント不使用での購入者の割合は従来以上

累計購入金額や購入頻度が低い層

群	ポイント使用率	稼働率	ポイント不使用の 購入者率
A	3.94%	5.32%	1.38%
B	4.44%	2.39%	1.84%

想定使用率
8.14%

累計購入金額や購入頻度が中程度の層

群	使用率	稼働率	ポイント不使用の 購入者率
A	11.70%	17.12%	5.42%
B	15.28%	13.24%	6.18%

想定使用率
20.66%

テスト内容

- A群：従来通りのポイント付与
- B群：Redshift MLによる購買予測を用いたポイント付与

※稼働率：各郡におけるポイントの有効期間内に購入した人の割合。ポイント使用は問わない

結果

- ポイント使用率が想定より低い
- 想定使用率と結果の乖離具合がユーザ層で変化
- 稼働したユーザの内、ポイント不使用での購入者の割合は従来以上

累計購入金額や購入頻度が低い層

群	ポイント使用率	稼働率	ポイント不使用の 購入者率
A	3.94%	5.32%	1.38%
B	4.44%	2.39%	1.84%

想定使用率
8.14%

累計購入金額や購入頻度が中程度の層

群	使用率	稼働率	ポイント不使用の 購入者率
A	11.70%	17.12%	5.42%
B	15.28%	13.24%	6.18%

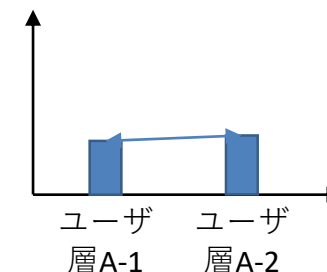
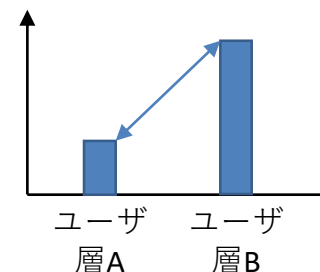
想定使用率
20.66%

⇒ 対象のユーザ層を絞り、モデルを改善する

改善点①：絞り込んだユーザ層のみのデータを使用して学習

改善点②：特徴量の精査

- ・ ユーザ層が絞られたことで不要になる値の削除
- ・ 同一ユーザ層を更に区別するための値の追加修正



改善点③：閾値設定の自動化

- ・ 「稼働してくれる人を落とさずに、無駄な付与数を極力減らしたい」
- ・ 最大化問題に変換
 - ・ $\operatorname{argmax} f(x) = \alpha TP(x) + \beta FFP(x) + \gamma TN(x) + \delta FN(x)$
 - ・ x ：閾値、 $\alpha \sim \delta$ ：係数
 - ・ 購入金額などをもとに係数を決定

予測→ 正解↓	使用	不使用
使用	真陽性 TP	偽陰性 FN
不使用	偽陽性 FP	真陰性 TN

※稼働率：各郡におけるポイントの有効期間内に購入した人の割合。ポイント使用は問わない

テスト内容

- ・ ユーザ層：累計購入金額や購入頻度が中程度の層
- ・ A群：従来通りのポイント付与
- ・ B群：Redshift MLによる購買予測を用いたポイント付与

結果

- ・ ポイント使用率が想定に接近
- ・ 稼働率は従来をわずかに下回る程度
- ・ 稼働したユーザの内、ポイント不使用での購入者の割合は従来以上

累計購入金額や購入頻度が中程度の層

群	ポイント使用率	稼働率	ポイント不使用の 購入者率
A	12.14%	20.35%	8.20%
B	20.84%	18.07%	9.83%

想定使用率
19.83%

⇒ 実運用に乗せられる！！

短期的な展望

- Redshift MLによる予測を通常運用フローに組み込む
- システム詳細が**分かる人**がミスなく運用できる仕組み作り
 - パラメータ設定の簡略化
 - スクリプトの自動実行

長期的な展望

- システムの詳細が**分からない人**でも運用できる仕組み作り
- 予測を適用可能なユーザ層の拡大

課題

- 付与条件の変更に対応する方法(累計購入金額、訪問頻度、購入した書籍ジャンル etc.)
- アラートのあげ方や、その対応方法
- モデル更新タイミングの指標

booklista

創業10周年を経て
次のvisionは、

エンタメ×テック=
知的好奇心×感動体験。

採用情報はこちらから→<https://www.booklista.co.jp/corp/recruit/>