

Supershipが推進する

統合型データサイエンスサービス

-戦略パートナーDatabricksとの取り組み-

# 機械学習をより手の届く技術へ

Supershipは、機械学習活用プロセスを包括的にサポート  
データ管理や環境構築にかかる工数/コストを最小限にすることで  
機械学習そのものにフォーカスすることを可能にします

# 機械学習活用における課題

データの管理→モデル構築→活用まで多くのプロセスが存在  
一連のプロセスを生産性高く、整備することが求められる

## データ管理

データの統合、  
権限管理、整形加工  
など管理が煩雑

## モデル構築

分析作業を効率的に実  
現する環境の不足  
コスト高な構造

## 活用

Pjt開始～検証まで  
都度複数課題に直面  
本格活用まで進みにくい

活用に至るまで、時間とコストばかりが嵩んでしまう

# Supershipグループにおける機械学習支援

データ準備・拡張→モデル構築→活用までの一連のフローを支援  
企業が抱える課題に対して、End-to-Endで解決する体制を整備

## データ管理

### データ拡充



貴社データ

Supership  
3rd Party Data

モデルの精度を高めるため  
Supershipのもつ  
3rd Party Dataでデータ拡充

## モデル構築

### 分析環境の提供

分析プラットフォーム



データサイエンティストに  
よるモデル構築支援



## 活用

### データ活用 プロダクト提供

広告ターゲティング活  
用

自社サイト最適化

BI等による可視化

機械学習活用プロセスを包括的にサポート

# 機械学習の活用例

Supershipデータも用いてユーザーの行動予測や興味関心分析をさらに深掘り  
 広告主様やメディア企業様が抱える課題解決アプローチを強化

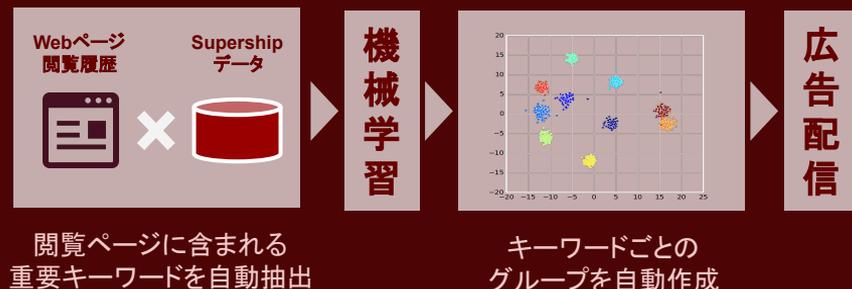
## 行動予測モデルを用いたターゲティング広告

既存顧客と類似する特徴量を保有するユーザー  
 を抽出し、より高度なターゲティングを実現



## キーワードによる興味関心解析の高度化

Webの閲覧ページを解析し、  
 抽出したキーワードをグルーピングすることで  
 ユーザーの嗜好や価値観を高精度に類推



# 機械学習に理想的な分析環境

機械学習を行うにあたって、Supershipが重視している観点は3つ

## Focus on Data Science, not Platform management

分析環境づくりから解放され、  
機械学習への注力が可能

## Knowledge sharing

チームコラボレーションによる  
ノウハウ共有・品質管理

## Small Start, High Scalability

小さく・シンプルに開始  
かつ、ビジネスに合わせた拡大



戦略的パートナーシップを締結し、  
企業のデジタルトランスフォーメーションを推進するための  
統合型データサイエンスサービスを新たに開始しました

# Supership×Databricksの提供価値

高精度なデータと分析技術、統合データ分析プラットフォームが揃うことで、  
開発環境やデータの提供、モデル構築支援までをワンストップで提供



精度の高い  
3rd Party Data



サイエンティストに  
よる  
モデル構築支援

機械学習環境  
3rdPartyデータ  
機械学習ナレッジ



統合データ分析プラットフォーム  
「Unified Data Analytics Platform」

## 事例1: 行動予測モデルを活用した広告配信

## データ管理

データのサイロ化

ガバナンスの複雑化

管理工数の肥大化

## モデル構築

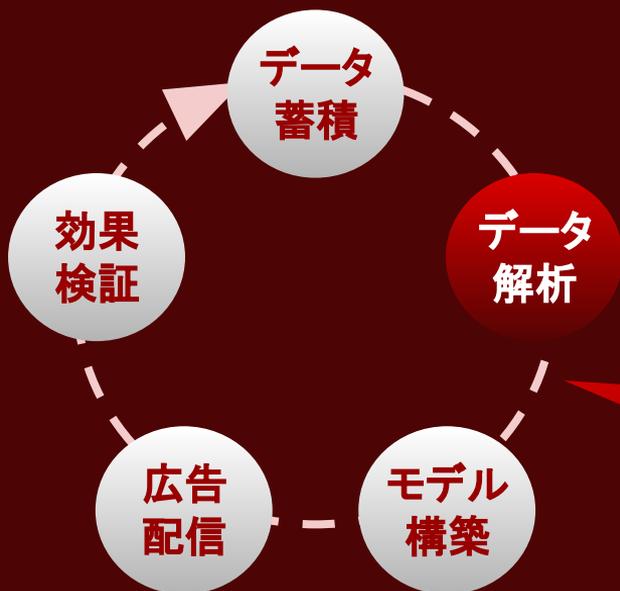
デマンド  
サイド大量データ処理が  
ボトルネックとなり  
精度向上をやりきれない大規模データ集計/  
自動化/コスト等  
複数課題に直面

## 活用

準備から検証に  
時間がかかり、  
本格活用まで至らない

# 事例1: 行動予測モデルを活用した広告配信

広告配信におけるCVR/CTRの最大化は、我々の事業テーマ  
課題は、高度な取り組み実施や効果検証までに時間を必要とする点

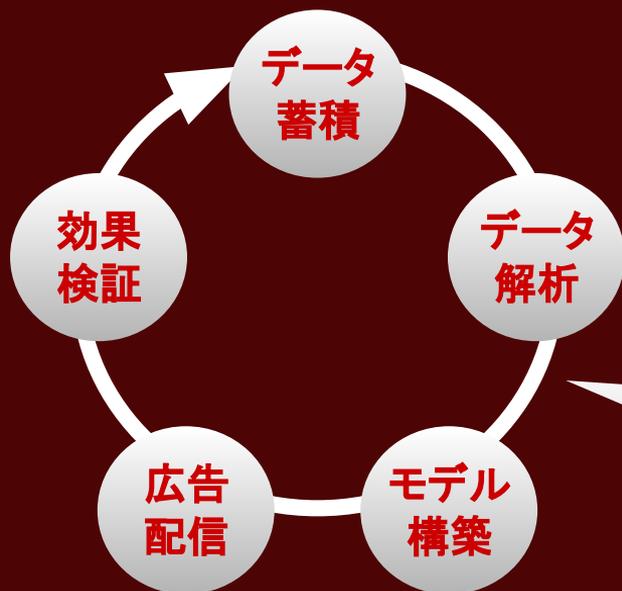


膨大なIDの集計・解析を必要とし  
技術的課題への対処が必要

案件ニーズが生まれてから  
準備～検証に時間がかかり  
本格活用まで至り切らない

# 事例1: 行動予測モデルを活用した広告配信

営業～データサイエンティスト一体化となり必要とする機能要件を具体化  
加えて、必要環境も整備することで、定常的なサービス提供へ進捗



**分析者のみで完結する分析環境**

-弊社の場合は、Databricks+AWS

**顧客ニーズを落とし込んだ  
予測モデル開発プロダクト準備**

# 事例1：行動予測モデルを活用した広告配信

機械学習を用いた予測モデルをクライアントごとに構築することで  
同じ特徴量を持つユーザーを予測し、マーケティング活動に活用



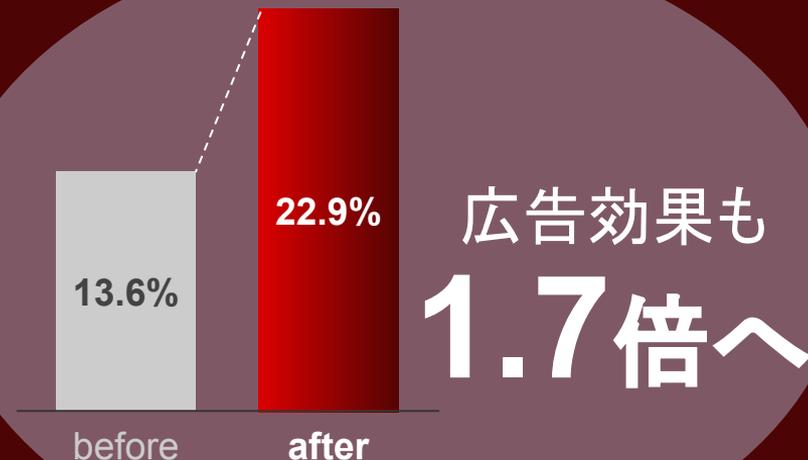
データ取り込みからセグメント作成

**約3日間**

生産効率の大幅改善、パフォーマンスの向上により  
ビジネス活用がよりスムーズに



作成リードタイム



広告パフォーマンス(CTR)

## 事例2 記事コンテンツを通じたユーザー理解

## データ管理

データのサイロ化

ガバナンスの複雑化

管理工数の肥大化

## モデル構築

大量データ処理が  
ボトルネックとなり  
精度向上をやりきれないサプライ  
サイド大規模データ集計/  
自動化/コスト等  
複数課題に直面

## 活用

準備から検証に  
時間がかかり、  
本格活用まで至らない

## 事例2 記事コンテンツを通じたユーザー理解

メディア様各社は自社コンテンツの”お客様”を正しく理解したい  
アプローチとしては、粗くなく、精緻に理解することを目標としたい

ページタイトル/  
記事カテゴリによる限定的な理解

読了内容まで加味した  
深い理解



ありがちな  
ユーザー理解範囲

理想は記事内容等を全て解析  
計算量・技術・コスト視点の  
課題が噴出

## 事例2 研究テーマ

自動化

データの  
保持方法

網羅性

# 事例2 プロセス



ユーザーが閲覧したURLを抽出し、クレンジング



ユーザー



抽出した閲覧URLから不要な情報を削除

`https://www.supership.jp/news/200529/1810290038gg-n1.html?fbclid=IwAR3zwmVLS5jtNR48wcQNOZvjYOKvmMm9mPJrfPmwrCw`



`https://www.supership.jp/news/200529/1810290038gg-n1.html`

## 事例2 プロセス

閲覧URL  
抽出/整理

テキスト抽出

キーワードを  
グルーピング

ユーザーに付与

URL内のテキストを全てクローリングし、記事文章を抽出



ページ内のテキストを識別

本サービスは、Supershipが提供する高精度で豊富なデータと分析技術、Databricks社が開発・運営する統合データ分析プラットフォーム「Unified Data Analytics Platform」を用いて、高速処理が可能なデータ分析基盤の構築から3rd Partyデータの提供、分析基盤での機械学習などを用いたデータ分析まで、ワンストップで提供するものです。

記事を抽出

## 事例2 プロセス

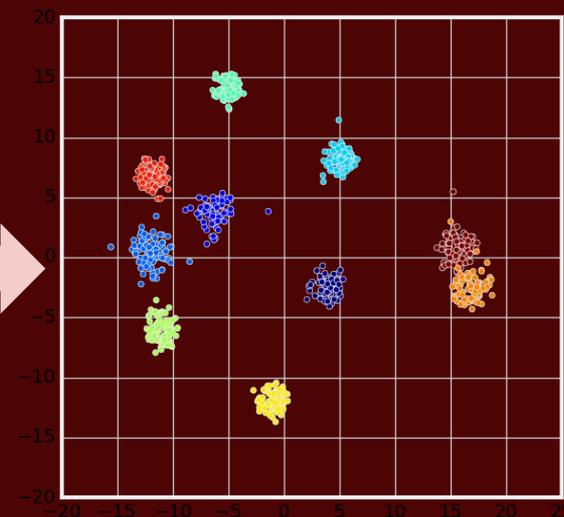


テキストから単語を抽出し、出現頻度から重要度やKW間の距離を推定してクラスター化

### クラスター例:スイーツとアート

"味わい","フレーバー","食感","風味","味","発売","チョコ","チョコレート","コク","スイーツ","期間限定","アイス","トッピング","パニラ","酸味","香り","新フレーバー","クリーム","キャラメル","ミルク","抹茶","税込","果汁","商品","数量限定","楽しめる","SweetS","ストロベリー","希望小売価格","ロどけ","フラペチーノ","カスタード","チョコミント","フルーツ","香ばしい","生クリーム","果肉","パッケージ","アイスクリーム","レモン","甘酸っぱい","ショコラ","赤城乳業","ハーゲンダッツ","カカオ","マンゴー","ホワイトチョコレート","おいしい","税別","ロングセラー","画家","作品","日本画","油彩","油彩画","版画","画壇","展","本展","風景画","浮世絵","絵画","個展","展覧会","同展","画業","所蔵","美術","学芸員","印象派","美術館","絵師","美人画","日本画家","葛飾北斎","企画展","同館","浮世絵師","展示","作品展","伊藤若冲","モネ","東京美術学校","洋画家","特別展","彫刻","原画","美術史","渋沢","木版画","北斎","素描","写実","若冲","屏風","絵","画風","ルノワール","セザンヌ","歌川国芳"

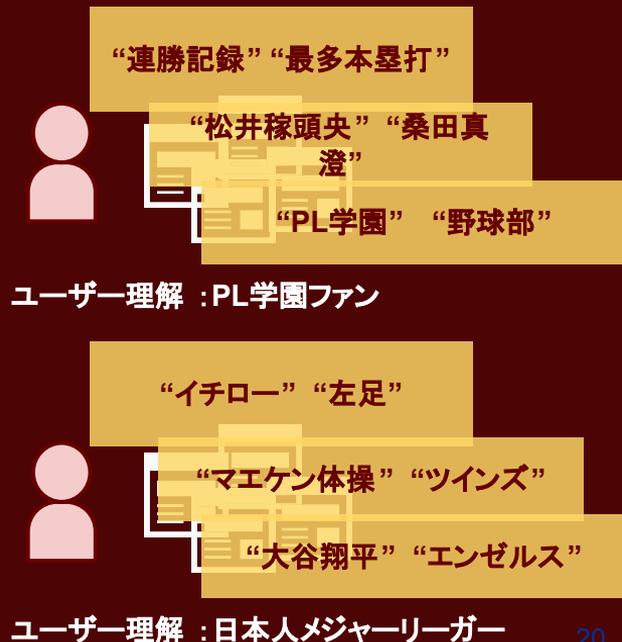
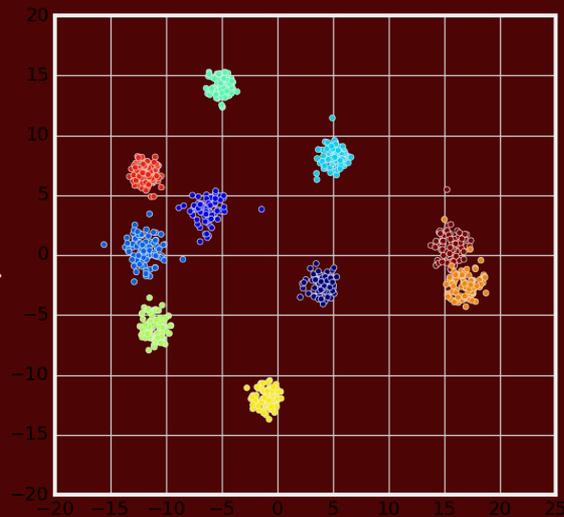
キーワード間の距離を  
機械学習し、  
カテゴリごとに分類



## 事例3 プロセス

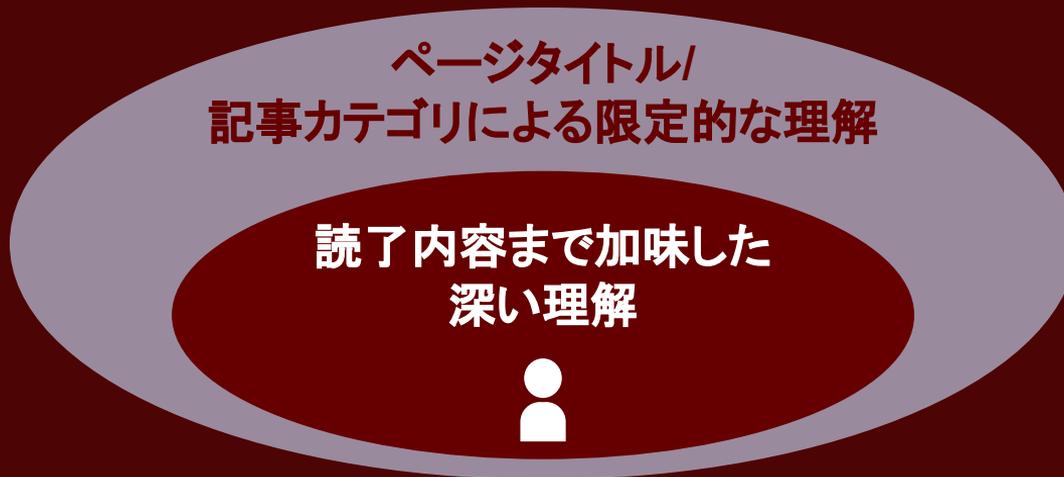


記事に含まれるテキストを束ねたクラスター情報を閲覧ユーザーに対して付与  
対象クラスターに興味を有したユーザーとして認識



## 事例2 記事コンテンツを通じたユーザー理解

サイト自体やサイト内記事カテゴリ情報をもとにしたユーザー理解を超え  
より具体的な興味関心の対象や、属性を推定へ活用



### メディア企業様向け活用

誰が、どのコンテンツ・体験を求めているのか？を把握

- 広告商品企画
- コンテンツ企画、プロダクト企画

### EC事業・ECモール様向け活用

誰が、カテゴリ毎の何を求めているのか？を把握

- 広告商品企画
- 商品レコメンド活用
- カテゴリ毎の体験企画

こちらまでお気軽にご相談ください  
data-solutions@supership.jp