

『Meet up Chubu』 vol.18 インフォマティクス

MIを材料開発に！コスト削減と開発の加速を支援

株式会社 日立ハイテク
SCLレジリエンス推進本部
マテリアルソリューション部
林 貴之

株式会社 日立ハイテク (日立製作所から出向中)
SCレジリエンス推進本部 マテリアルソリューション部

林 貴之 (ハヤシ タカユキ)

担当業務	主要顧客	民間企業のR&D部門
	役割	データサイエンティスト
	業務内容	お客様の材料データをお預かりし 機械学習などを活用してデータ分析を代行し 分析結果をレポートします。
データ 分析関連	資格	• 応用情報技術者 • 統計検定2級 • 統計調査士 • JDLA Deep Learning For GENERAL (2023#1) • AWS Certified Machine Learning – Specialty (MLS-C01)

アジェンダ

- ①材料開発における課題
- ②MIを用いた解決事例
- ③サービス概要

① 材料開発における課題

①材料開発における課題

高品質な材料探索に
時間とコストがかかりすぎる

原材料と配合比率の
パターンが多すぎる...

A	X B	X C	X D
80%	30%	10%	10%
70%	20%	8%	8%
60%	10%	6%	6%
...	...	...	...



画像から品質を判定するが
属人的になってしまう

電子顕微鏡画像の品質を
熟練者が経験・勘で判断している



これでは**工数**がかかるし
属人的で**再現性**がない...



大量の文献から実験データを
抽出するのは難しい

新規特許から
実験データを作成したい



でも...大量の文献から
実験データを抽出するのは
負担が多い



高品質な材料探索に
時間とコストがかかりすぎる

原材料と配合比率の
パターンが多すぎる...

A	B	C	D
80%	30%	10%	5%
70%	20%	8%	3%
60%	10%	6%	2%
...	...	...	...



画像から品質を判定するが
属人的になってしまう

電子顕微鏡画像の品質を
熟練者が経験・勘で判断している



これでは工数がかかるし
属人的で再現性がない...



大量の文献から実験データを
抽出するのは難しい

新規特許から
実験データを作成したい



でも...大量の文献から
実験データを抽出するのは
負担が多い



MIで解決

①材料開発における課題：MIとは

説明変数X

- 実験データ
 - ・ 化合物
 - ・ 原料配合比
 - ・ 実験プロセス
 - ・ 画像データ
- シミュレーションデータ
- 特許データ

順解析

機械学習モデル

$$Y=f(X)$$



目的変数Y

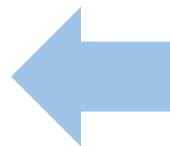
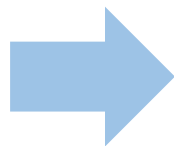
- 実験データ
 - ・ 物性/活性/毒性
 - ・ 材料品質
 - ・ 画像の異常面積

逆解析

実験データを活用し、機械学習やAIを駆使した
多次元分析を実施することで材料特性の仮想実験を行う

説明変数X

- 実験データ
 - ・ 化合物
 - ・ 原料配合比
 - ・ 実験プロセス
 - ・ 画像データ
- シミュレーションデータ
- 特許データ



目的変数Y

- 実験データ
 - ・ 物性/活性/毒性
 - ・ 材料品質
 - ・ 画像の異常面積

【機械学習モデル】
説明変数Xと目的変数Yを用いて関数 f を構築する

説明変数X

- 実験データ
 - ・ 化合物
 - ・ 原料配合比
 - ・ 実験プロセス
 - ・ 画像データ
- シミュレーションデータ
- 特許データ

順解析

機械学習モデル

$$Y=f(X)$$



目的変数Y

- 実験データ
 - ・ 物性/活性/毒性
 - ・ 材料品質
 - ・ 画像の異常面積

【順解析】

変数Xに値を入力 → 予測値Yを得る

説明変数X

- 実験データ
 - ・ 化合物
 - ・ 原料配合比
 - ・ 実験プロセス
 - ・ 画像データ
- シミュレーションデータ
- 特許データ

機械学習モデル

$$Y=f(X)$$



目的変数Y

- 実験データ
 - ・ 物性/活性/毒性
 - ・ 材料品質
 - ・ 画像の異常面積

逆解析

【逆解析】

物性値Yに値を入力 → 変数Xの値を得る

③ MIを用いた解決事例

②解決事例：順解析・逆解析（概要）



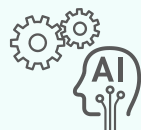
お悩み①：高品質となる条件の探索に時間・コストがかかる

ご準備頂くデータ

実験番号	主成分 A	主成分 A 比率	主成分 B	主成分 B 比率	添加剤 A	添加剤 A 比率	添加剤 B	添加剤 B 比率	..	物性値
1	a	20	b	60	c	30	d	10	..	88
2	x	35	y	20	z	15	w	45	..	60
:	:	:	:	:	:	:	:	:	..	:

主成分の候補数×添加剤の
候補数×比率パターンという
広大な材料探索空間

MI



物性予測モデル

既存実験データの
パターンを学習



最適化アルゴリズム

$f(\text{原料, 比率}) = \text{物性}$



物性値が高く(低く)なる原料・比率を
仮想空間で探索

アウトプット

MIが算出した実験候補

実験候補番号	主成分 A	主成分 A 比率	主成分 B	主成分 B 比率	添加剤 A	添加剤 A 比率	添加剤 B	添加剤 B 比率	..	物性推定値
1	y	25	b	30	z	10	d	20	..	99
2	a	30	x	30	z	20	w	10	..	98
:	:	:	:	:	:	:	:	:	..	:



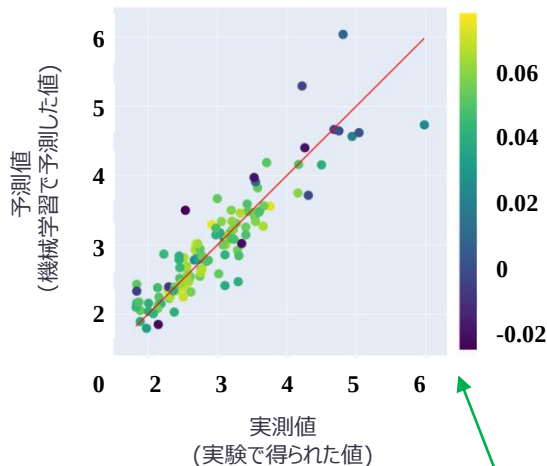
- ✓ 実験回数・コストの削減
- ✓ 品質の向上

材料特性の予測精度と解釈性を両立して、材料設計指針に資する情報を提供

■ 予測性能

*イメージ図

<予測値-実測値グラフ>



#	アルゴリズム(例)	相関係数 (R)	平均誤差 (MAE)
1	eXtreme Gradient Boosting(XGB)	0.83	0.21
2	Gaussian Process(GP)	0.81	0.23
3	Support Vector Machine(SVM)	0.83	0.22
4	Partial Linear Square Regression(PLS)	0.81	0.23
:	:	:	:

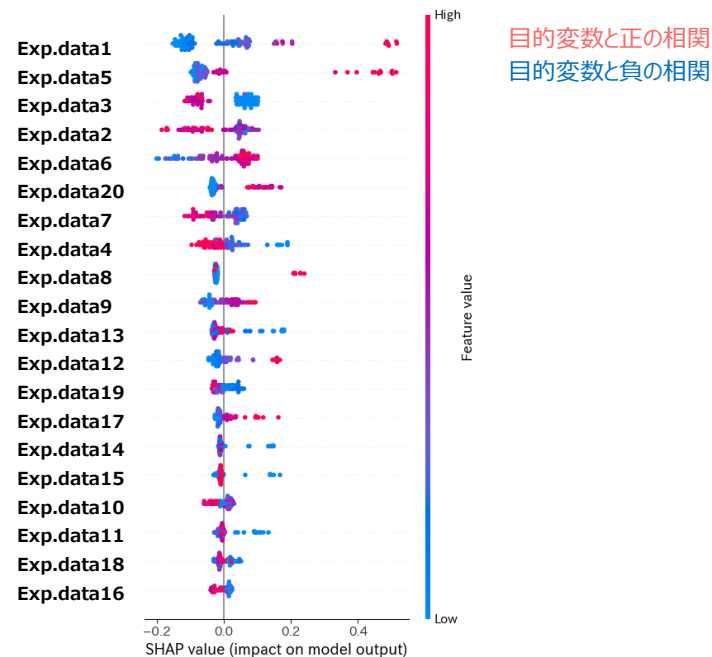
AD (Applicability Domain) スコア

モデルが十分な性能を発揮できるデータ領域。
KNN(K-Nearest Neighbor)やOCSVM(One-Class SVM)などが使用される。
ADスコアが高いほど信頼性のある予測結果となる。

■ 説明変数の重要度

(XAIを用いた予測のメカニズムの可視化)

*イメージ図

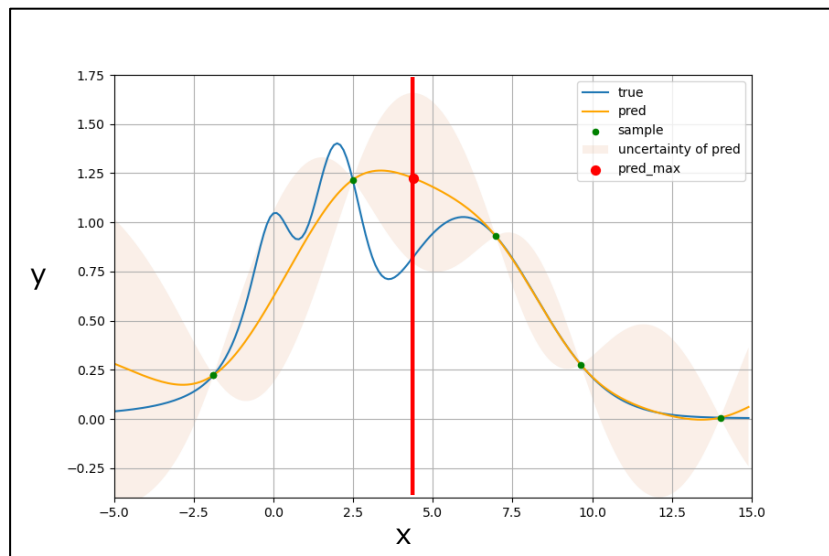


目標とする材料特性を達成し得る実験候補のリストを算出（多目的対応、制約条件付与）

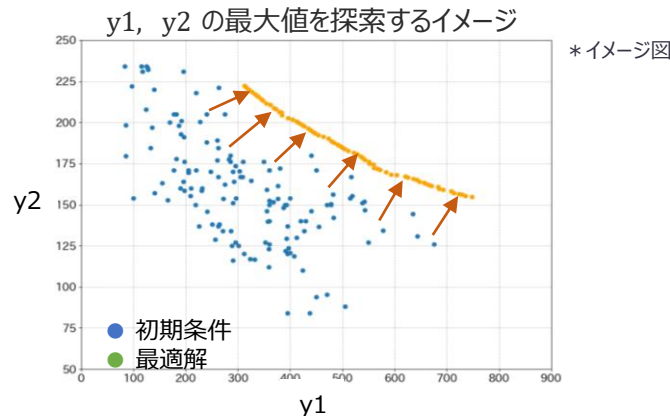
- A) 勾配降下法を用いて応答曲線の停留点を探索
- B) データが未観測の領域を重点的に狙うベイズ最適化も可能

-ベイズ最適化-

*イメージ図



- C) 複数の目的変数を対象とした多目的最適化にも対応



- D) 説明変数の制約条件や目的変数の目標範囲指定など
複雑な条件下での逆解析も可能

例) 制約条件	目標値
$10 \leq \text{原料A} \leq 20$	物性A = 100
原料A + 原料B + 原料C = 100	物性B ≤ 30

化合物情報をAIが読み取れる情報（記述子）に変換

#	記述子	特徴量の数	概要	使用ライブラリ
1	RDKit	208	ケモインフォマティクスの分野で最も使用されている記述子です。 分子量や電荷の情報といった特性記述子と、環構造の数などの構造記述子とから構成されます。	RDKit ^{*1}
2	Avalon FP	512	FP(fingerprint)は、化学構造をある規則でもって 部分構造に分解し、その部分構造を持つ/持たないのビットベクトルで表現したものです。 規則(ベクトルの次元数, 部分構造の定義方法)はこれまでかなりの量が考案されています。AvalonやMACCSKeyという名前が規則を表します。	
3	MACCSKey FP	166		
4	Morgan FP	1024		
5	Mordred	1613 (2D) 1826 (3D)	RDKitと比較して記述子の種類が豊富であるという特徴を持ちます。 2次元での化学情報に加え、化合物の立体構造を加味した 3次元の化学情報を算出することも可能です。	Mordred ^{*2}

^{*1} <https://www.rdkit.org/>

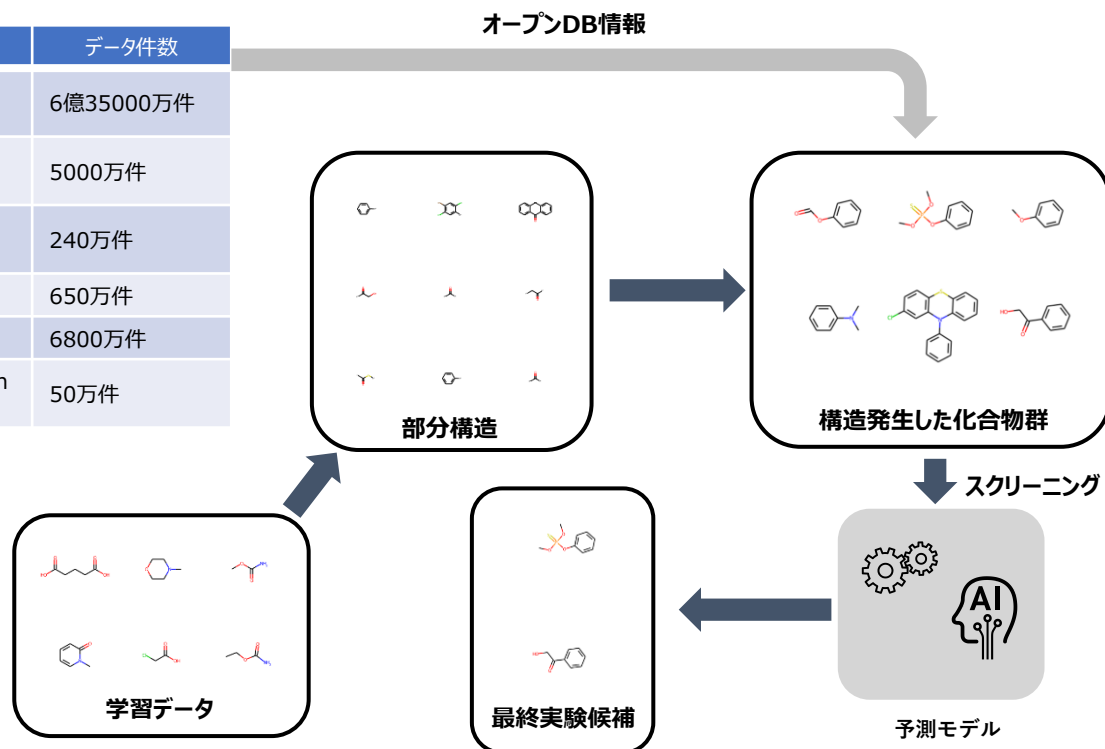
^{*2} https://www.jstage.jst.go.jp/article/ciqs/2016/0/2016_Y4/_pdf/-char/ja

②MI分析のコア技術【バーチャルスクリーニング】

構築したAIモデルで新規生成した化合物 & オープンDBから仮想実験が可能

*イメージ図

#	DB名	出典	データ件数
1	GDB13	University of Bern Reymond Research Group	6億35000万件
2	GDB17	University of Bern Reymond Research Group	5000万件
3	ChEMBL	European Molecular Biology Laboratory	240万件
4	Namiki	Namiki Shoji Co., Ltd.	650万件
5	ChemSpider	Royal Society of Chemistry	6800万件
6	Pubchem	National Center for Biotechnology Information U.S. National Library of Medicine	50万件



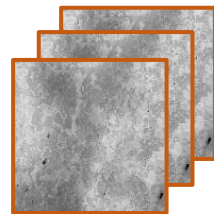
②解決事例：画像分析



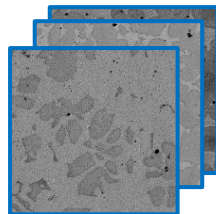
お悩み②：画像から品質を判定するスキルが属人化している

ご準備頂くデータ

電子顕微鏡画像



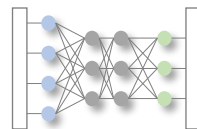
品質Good



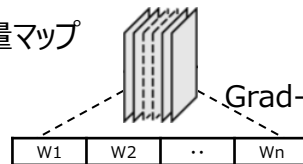
品質Bad

MI

深層学習で
【品質⇔画像の質感】を学習
(CNN : Convolutional Neural Network)

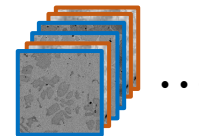


特徴量マップ

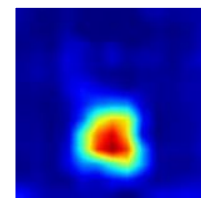


アウトプット

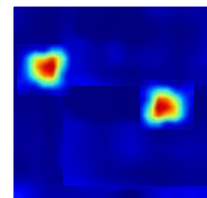
品質の自動判別結果



Good



Bad



判別根拠の可視化例



自動判別による開発効率化
スキルの定量化による品質安定化

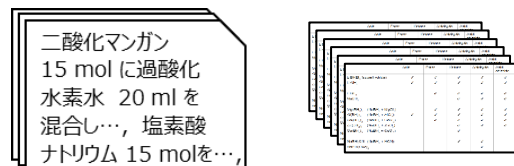
②解決事例：テキストマイニング



お悩み③：大量の文献から実験データを抽出するのは難しい

ご準備頂くデータ

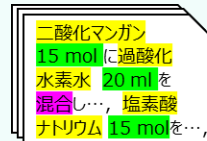
特許と実験データ



物質名・量・操作・温度など
抽出したい
複数カテゴリの情報

MI

自動ラベリング



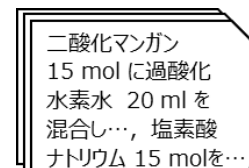
学習

抽出したい単語の周辺に現れる
特徴的な表現を学習

データから文意を分析してマッピング

アウトプット

新規特許から 実験データを抽出



抽出元文章	物質	量	温度	操作	...
塩酸 20 ml	塩酸	20 mol	30℃	混合	...



少量の学習データから自動抽出
システム構築による業務効率化

③ サービス概要

③ サービス概要：各サービスの特徴

分析 環境



MI分析ができるクラウド環境をご提供

使いやすさ

- 操作性の高いGUI（プログラミング不要）
- 結果のグラフ表示による可視化
- csv出力による情報活用

コンサルサポート

- 月1回データサイエンティストによるディスカッション

機密性の担保

- 流出事故なし

分析 支援



弊社データサイエンティストがMI分析を代行

テーマに合わせた分析が可能

- 順解析
- 逆解析
- 画像解析
- テキストマイニング

データサイエンティストが多数在籍

様々なバックグラウンドがあり、ドメイン知識が豊富なデータサイエンティストがテーマに合わせて対応

開発 DX



実験データを一元管理する環境をご提供

使いやすさ

- ユーザービリティの高いGUI
- 高い検索性

高い自由度

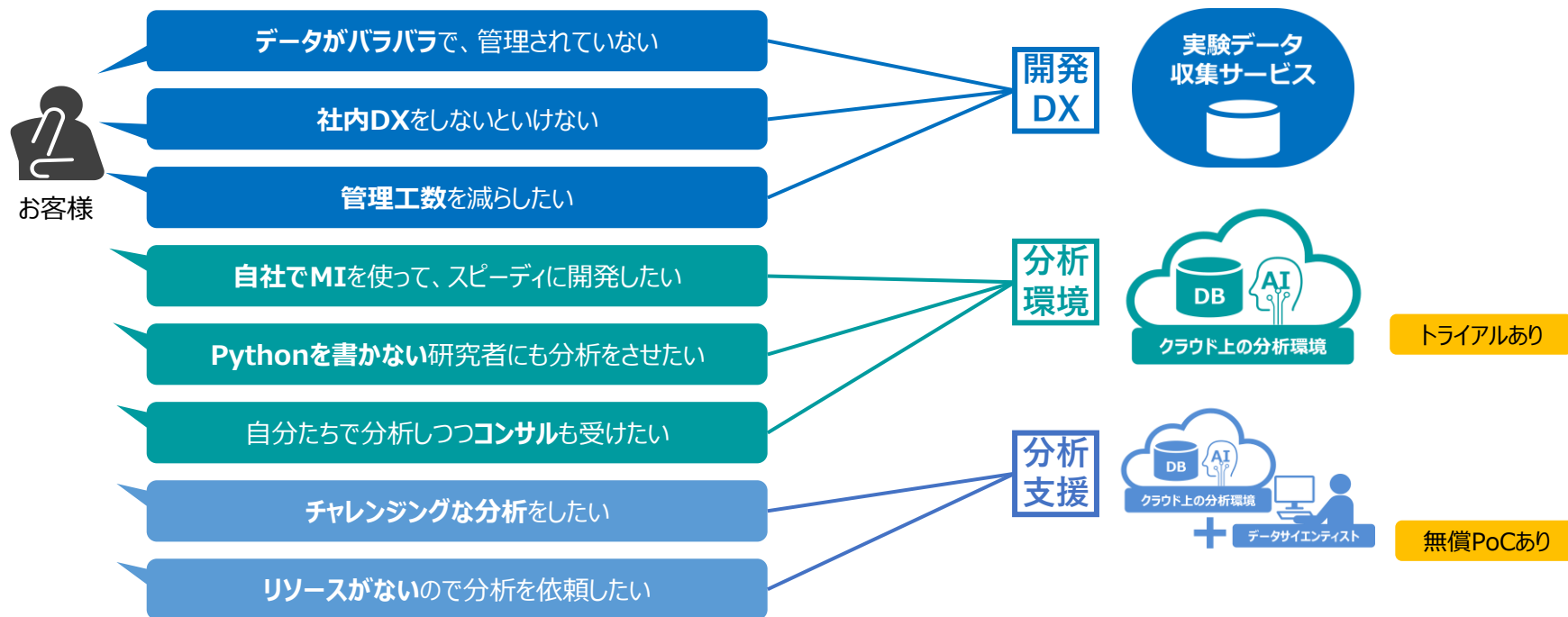
- お客様の環境に合わせた拡張

安全・安心

- 柔軟なアクセス管理
- 手厚い保守サポート

手厚いサポートと豊富な実績、安全性の高さに定評があります

③ サービス概要：頻出課題と日立ハイテクが提供するソリューション案内



お客様の課題に合わせて選択可能・組み合わせ自由

③ サービス概要：材料データ 分析支援 サービス

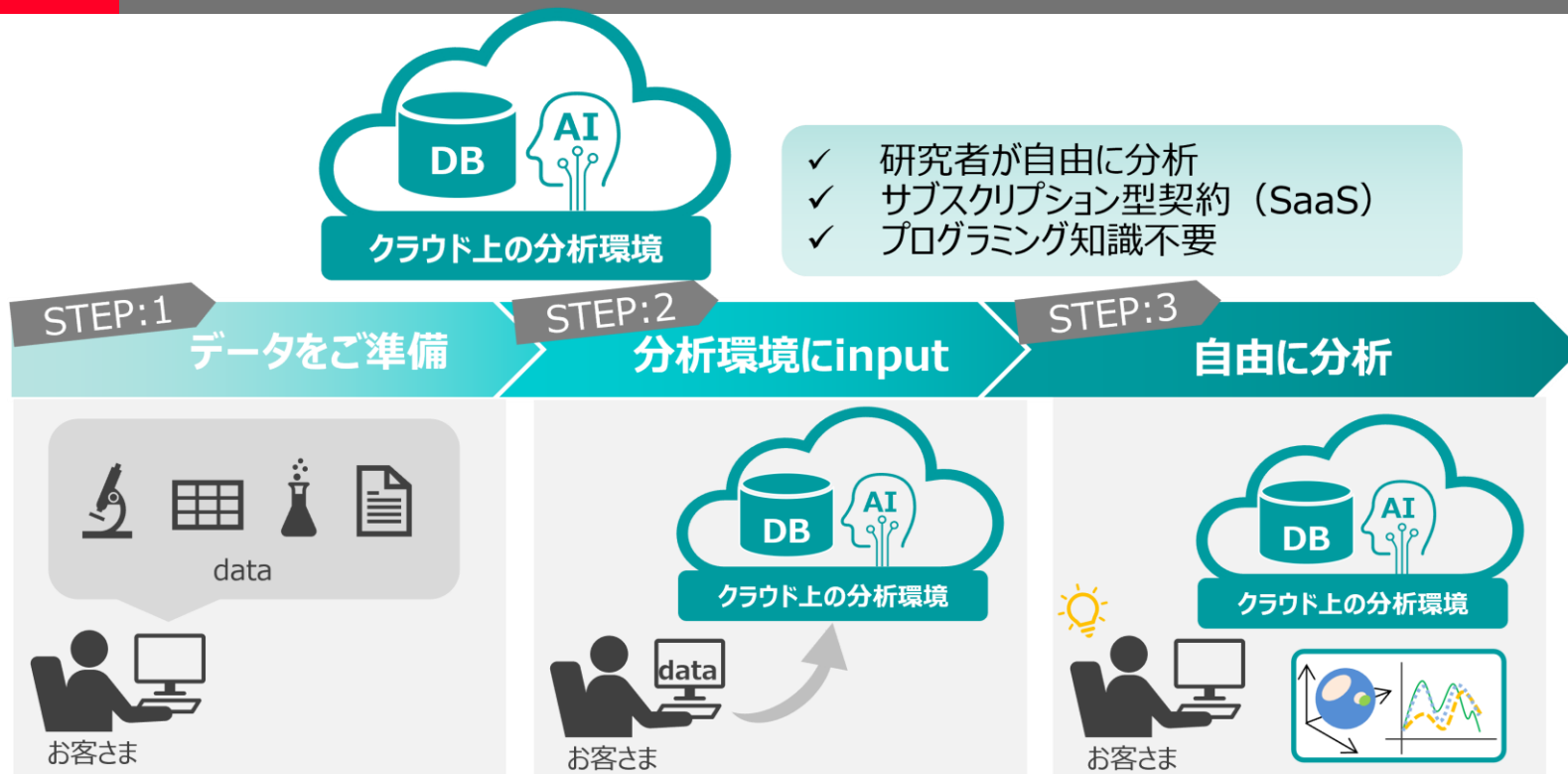


テーマに合わせて日立のデータサイエンティストが分析を行います

③ サービス概要：材料データ 分析環境 提供サービス

分析
環境

HITACHI
Inspire the Next



お客様がご自身で分析を行うサブスク型のMIツールです

③ サービス概要：体験可能な2つのサービス

分析
支援

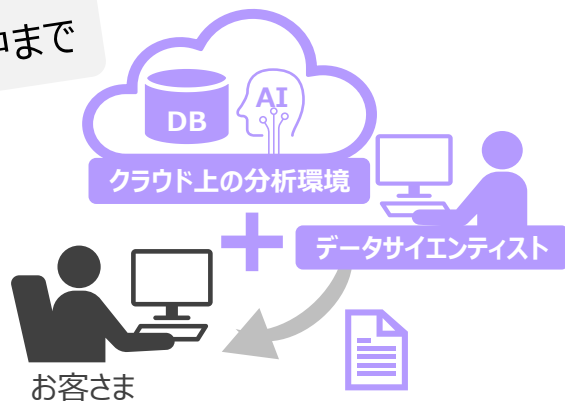
分析
環境

HITACHI
Inspire the Next

初回のみ

プレ分析(無償PoC)

分析途中まで



- ✓ 分析支援の前段階
- ✓ データサイエンティストの分析
- ✓ 予測モデルの構築まで
- ✓ 無償

トライアル分析環境

操作レクチャー

お試し



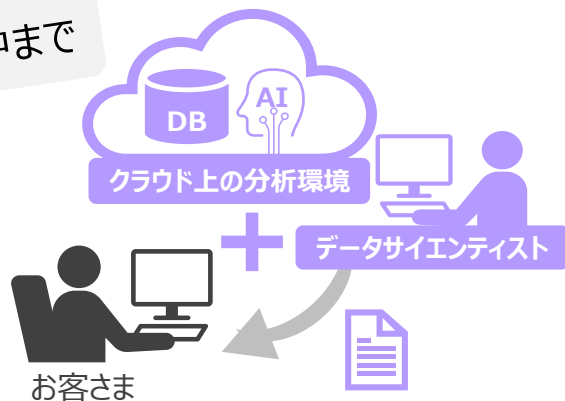
- ✓ 分析環境の前段階
- ✓ お客様による分析
- ✓ 操作レクチャー
- ✓ 有償

③サービス概要：体験サービス プレ分析（無償PoC）

初回のみ

プレ分析(無償PoC)

分析途中まで



- ✓ 分析支援の前段階
- ✓ データサイエンティストの分析
- ✓ 予測モデルの構築まで
- ✓ 無償

操作レクチャー

お試し



- ✓ 分析環境の前段階
- ✓ お客様による分析
- ✓ 操作レクチャー
- ✓ 有償

③ サービス概要：体験サービス プレ分析（無償PoC）

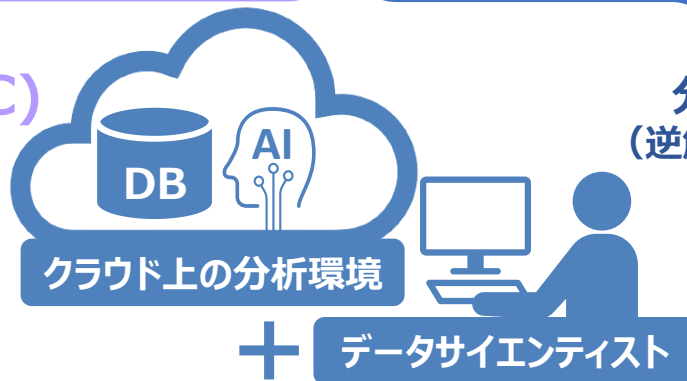
初回のみ

課題のヒアリング

データ整形

学習モデルの
構築学習モデルの
精度向上逆解析による
実験候補の策定

プレ分析(無償PoC)

分析支援の範囲
(逆解析がテーマの場合)

プレ分析【無償】を体験可能

③ サービス概要：体験サービス トライアル分析環境

分析途中まで



- ✓ 分析支援の前段階
- ✓ データサイエンティストの分析
- ✓ 予測モデルの構築まで
- ✓ 無償

初回のみ

トライアル分析環境

操作レクチャー



お試し



サポート



- ✓ 分析環境の前段階
- ✓ お客様による分析
- ✓ 操作レクチャー
- ✓ 有償

③ サービス概要：体験サービス トライアル分析環境

初回のみ

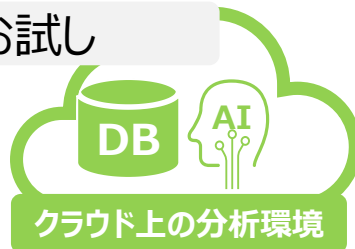
トライアル分析環境

操作レクチャー



データサイエンティスト

お試し



クラウド上の分析環境



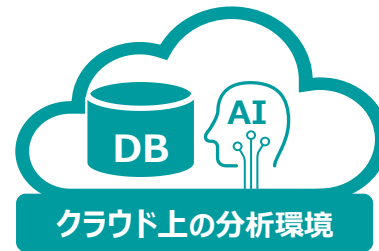
お客さま

サポート



データサイエンティスト

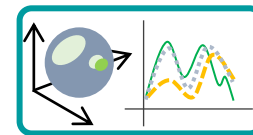
分析環境



クラウド上の分析環境



お客さま



特別価格で体験可能



Hitachi Social Innovation is
POWERING GOOD