

そうはいってもどうする、ビッグデータ

2012年4月19日

株式会社 アイズファクトリー



i's FACTORY co., Ltd.
Data mining opens the door to tomorrow.

<http://www.isfactory.co.jp/>

1. ビッグデータ解析実例 お客様の声
2. ビッグデータ概況
3. ビッグデータ解析は目的が重要
4. 目的を見誤らないためのデータマイニング

1. ビッグデータ解析実例 お客様の声

東京大学・東京工業大学で物理学（素粒子・宇宙論）を研究していた研究者が集まって創業。

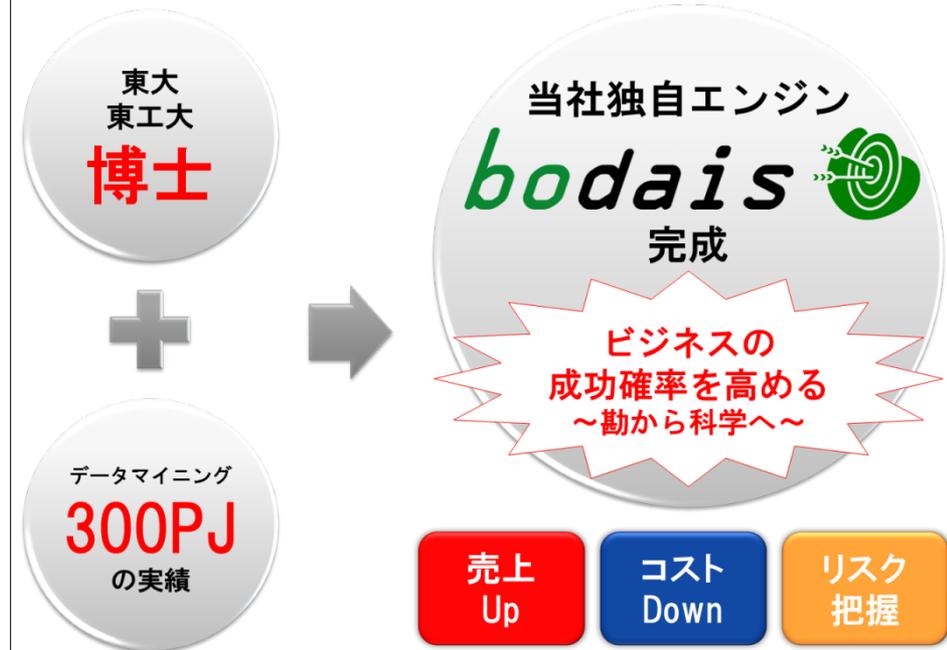
現在もスタッフの半数程度が博士号取得者。

売上UP、コストDown、リスク把握に関するデータマイニングソリューション（bodais）を提供しており、解析手法が確立されていない分野や、複数のデータを組み合わせて解析する場合などに、特に強みを発揮します。



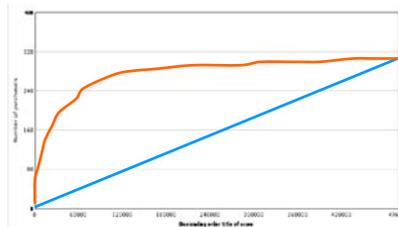
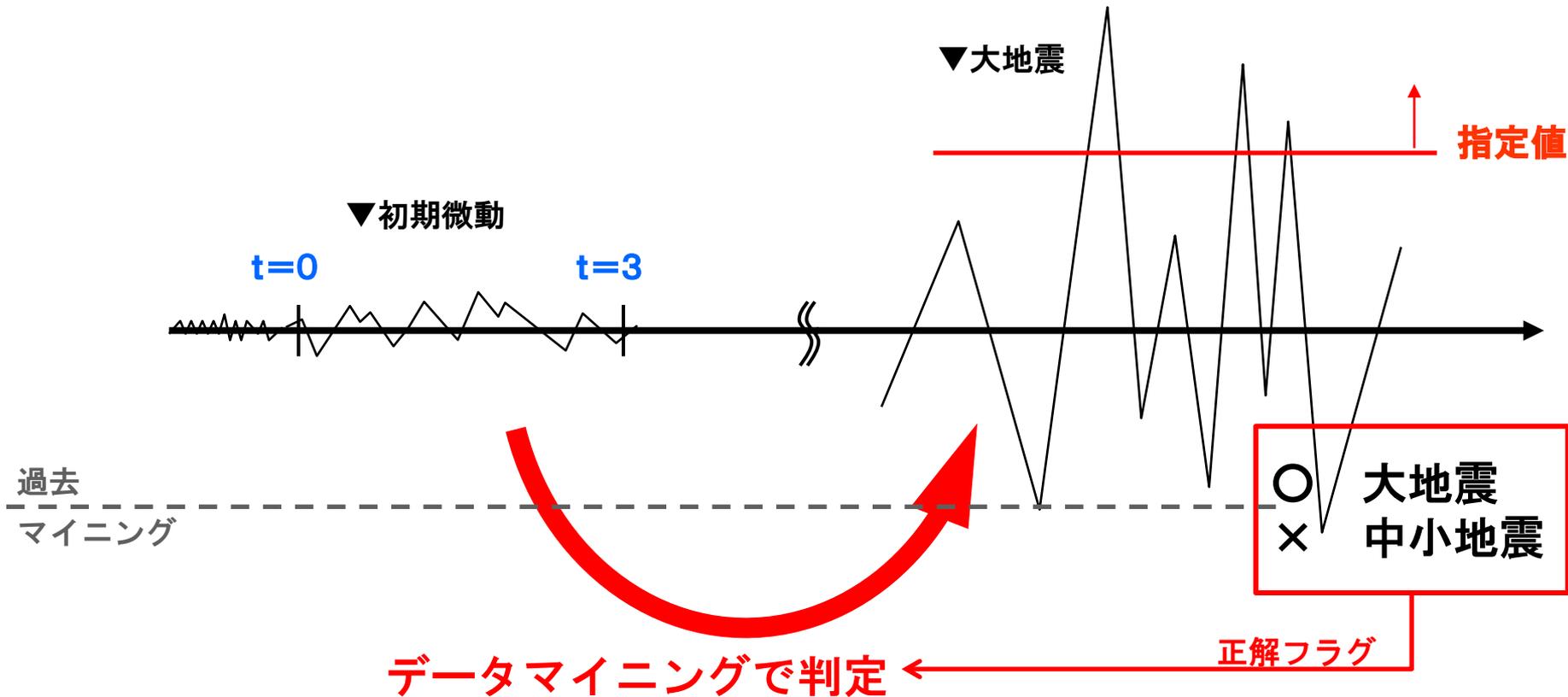
商号	株式会社アイズファクトリー (英表記 i's FACTORY co., Ltd.)
設立	2000年（平成12年）4月25日
資本金	54,906千円
本社	〒101-0054 東京都千代田区神田錦町1-23 03-5259-9004
役員	代表取締役 大場 智康 博士（理学） 取締役 岩崎 哲 博士（環境学） 取締役 佐藤 晴正 博士（理学） 監査役 山口 高志 公認会計士
従業員	35名（2012年4月現在）
事業	データマイニング・テキストマイニング・ 人工知能・数理科学・統計学を用いた データ解析/システム開発 クラウド型BIプラットフォーム「bodais」の提供

ビジネスの成功確率を高める ～勤から科学へ～



⇒各業種向けに、bodaisによるソリューションの提供

■ 大地震を当てる

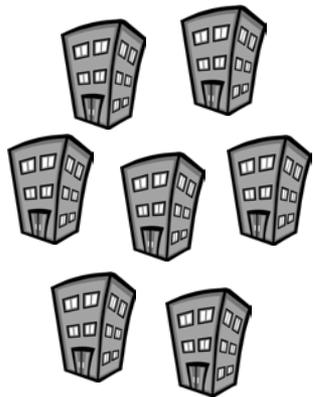


bodais

- ・地震前10秒～終了の約1分間のデータ
- ・XYZの3成分、地表、地下の合計6成分
- ・速度、加速度のテキストデータ

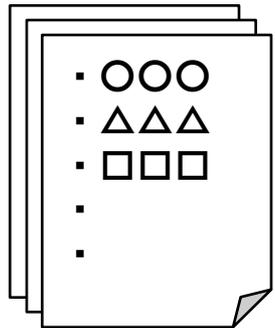
■ 税務申告書の非違を当てる

▼企業



提出

▼税務申告書

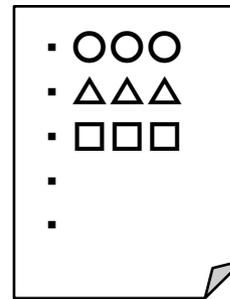


抽出

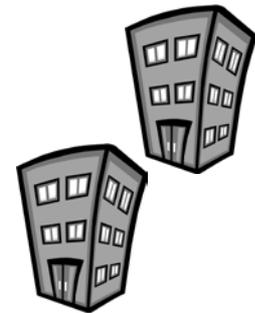
- ・ベテランノウハウ
- ・過去の経験



▼査察候補リスト



査察

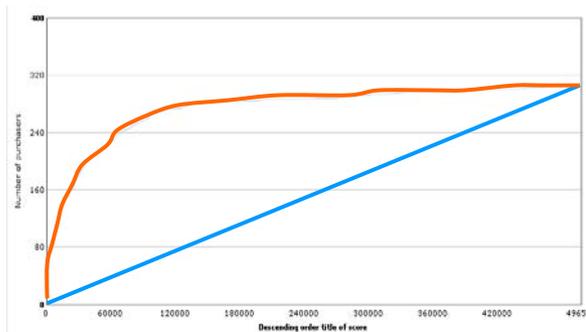


過去
マイニング

○ 課税あり
× 課税なし

正解フラグ

データマイニングで抽出



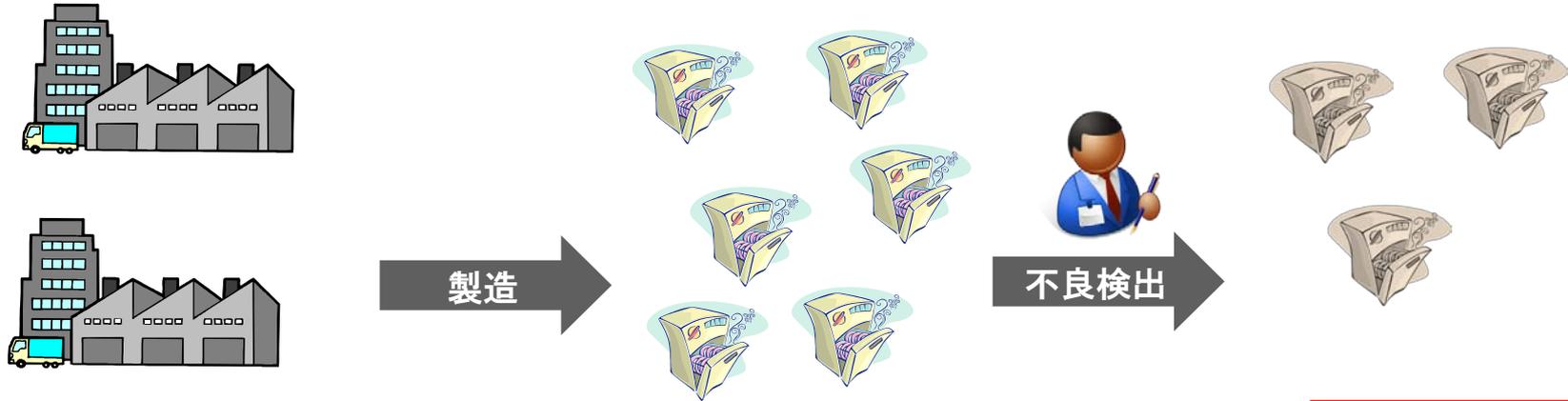
bodais

- ・査察実績
- ・企業属性
- ・申告書数値ほか

■ 工場の不良品率を当てる

▼工場

▼製品

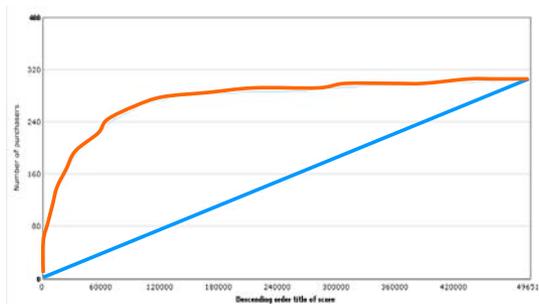


過去
マイニング

○ 不良あり
× 不良なし

正解フラグ

データマイニングで確率計算



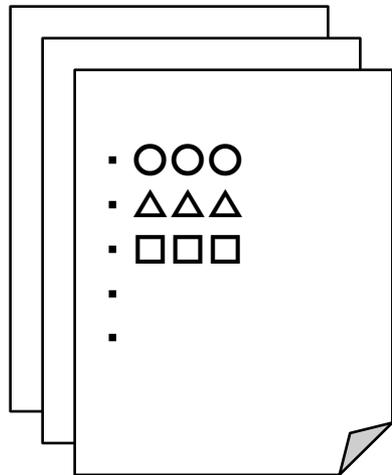
bodais

- ・ 天気
 - ・ スタッフ配置
 - ・ ライン
- 等の過去の製品製造条件

DM送付先を決める

▼送付先候補リスト

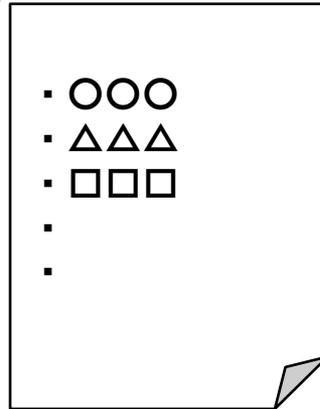
▼送付リスト



- ・ベテランノウハウ
- ・過去の経験



抽出



DM郵送

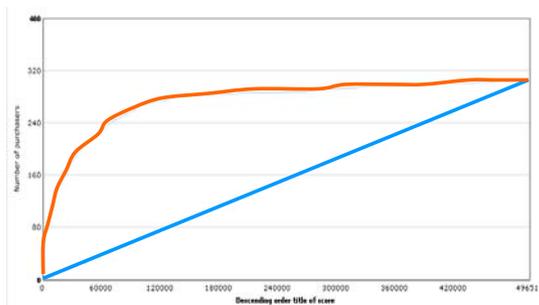


- 反応あり
- × 反応なし

正解フラグ

データマイニングで抽出

過去
マイニング



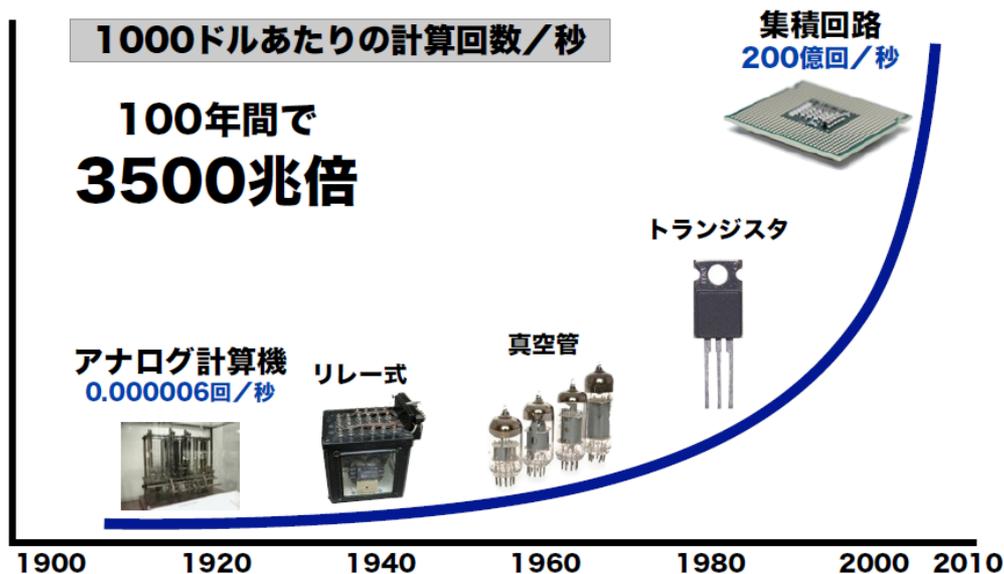
bodais

- ・運用実績
- ・属性（年収、兄弟…）
- ・コンタクト履歴（TEL、DM…）
- ・入退会履歴
- ほか

2. ビッグデータ概況

情報爆発が起きている！

	2010年		2040年
CPU トランジスタ数 	30億個	100万倍	3000兆個 (人間の脳の10万倍)
メモリ 容量 	32GB	100万倍	32PB
通信速度 	1Gbps	300万倍	3Pbps

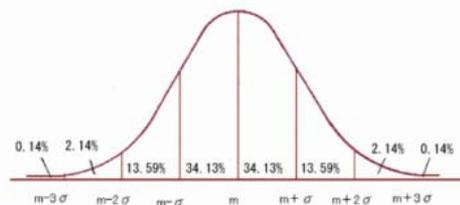


出典: 新30年ビジョン | ソフトバンク株式会社
<http://webcast.softbank.co.jp/ja/press/20100625/>

科学も変わってきている！

1. 物理(確定的世界観)

→ 科学の言語: 数学 【理論科学】

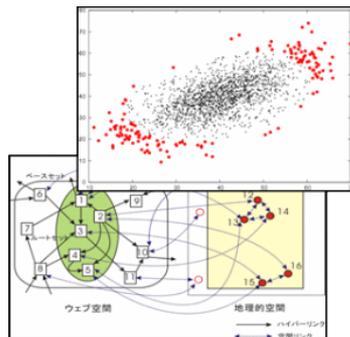
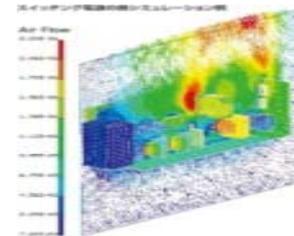


2. 生物・経済(確率的世界観)

→ 科学の文法: 統計学 【実験科学】

3. 非線形・複雑(システムの世界)

→ 計算科学、シミュレーション 【計算科学】



4. 大量・大規模(データの世界) **[現在]**

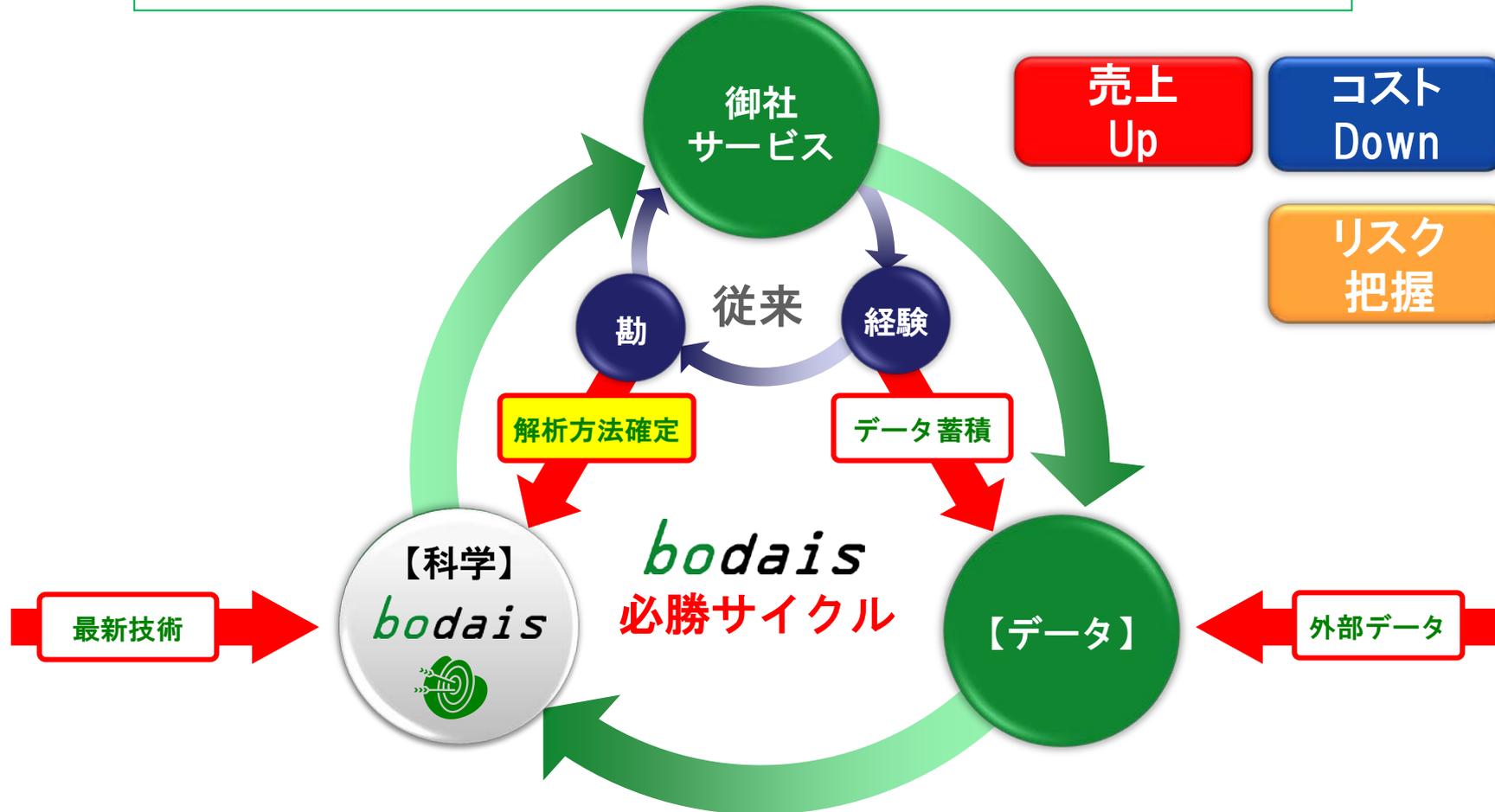
→ データマイニング、発見科学 【データ科学】

出典
全体: 数理システム ユーザーコンファレンス2007 統計的モデリング
-知識社会の基盤技術-北川源四郎 から抽出改変 <http://www.msi.co.jp/userconf/07/>
1. Microsoft Officeクリップアート
2. Weblio <http://www.weblio.jp/content/正規分布>
3. TDK <http://www.tdk.co.jp/techmag/technobox/200611/index.htm>
4. 筑波大学 計算科学研究センター <http://www.rcpp.tsukuba.ac.jp/cos/research/research-j5.html>

■ bodais必勝サイクル

bodais

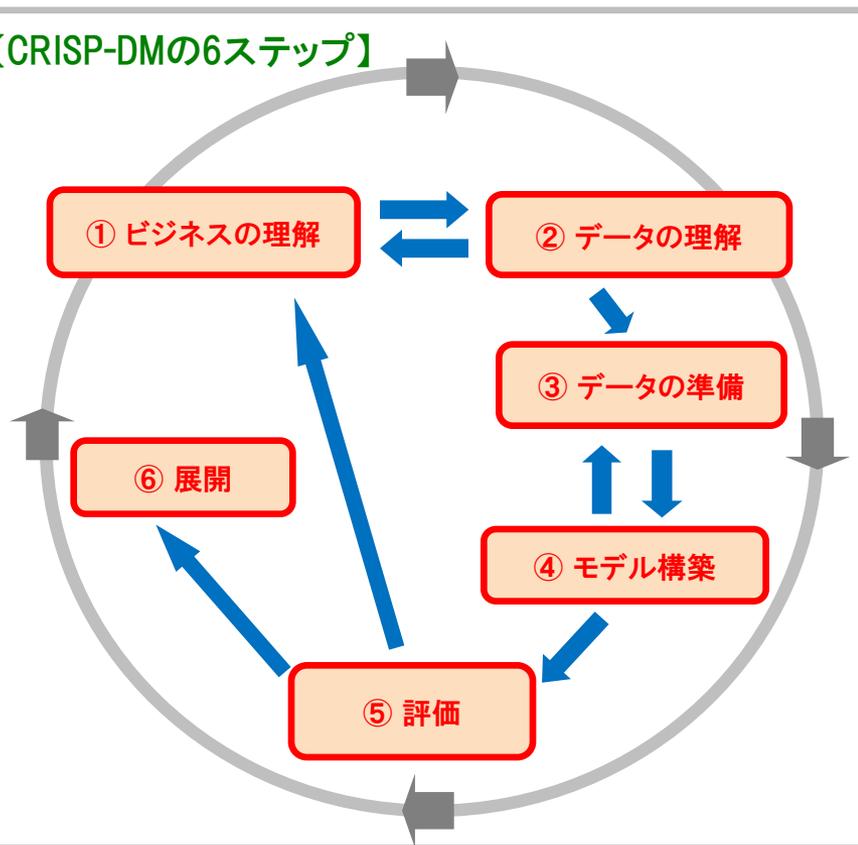
企業内の「経験」や「勘」をデータとして蓄積し、ビッグデータ解析の最新技術で「売上UP」、「コストDOWN」、「リスク管理」へとビジネスの成功へ導きます。



3. ビッグデータ解析は目的が重要

CRISP-DM (Cross-Industry Standard Process for Data Mining) は、SPSS、NCR、ダイムラークライスラー、OHRAがメンバーとなっているコンソーシアムにて開発されたデータマイニングのための方法論を規定したものです。業界横断的に展開可能なデータマイニングのプロセスモデルとなっています。

【CRISP-DMの6ステップ】



① ビジネスの理解

↓ ↑ 目標理解・目標数値設定・スケジュール設定

② データの理解

↓ データは目標を達成するのに十分か
(データ項目・量・品質の調査)
各データの基礎集計、詳細理解

③ データの準備

↓ ↑ マイニングの前処理
(データ洗浄・欠損値処理・サンプリングなど)
マイニング用にデータを加工

④ モデル構築

↓ 学術的な裏づけに立脚したデータ処理

⑤ 評価

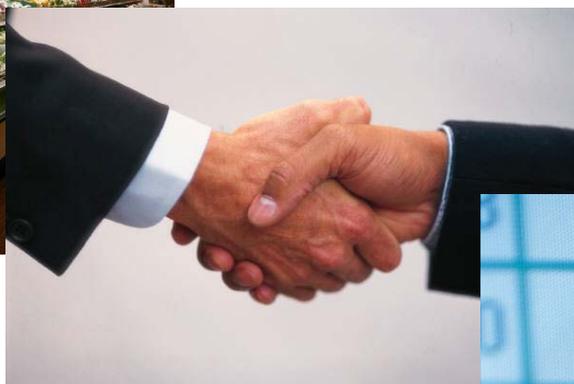
↓ 目標を達成するのに十分な知見が得られたか

⑥ 展開

↓ アクションそして次の目標へ



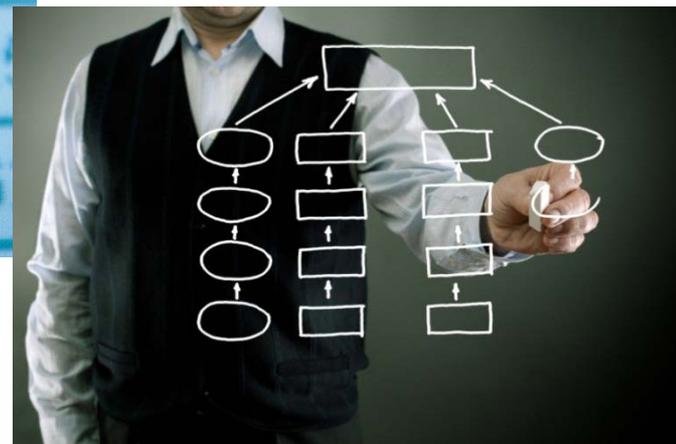
① ビジネスの理解



機密保持契約

② データの理解

この手順の途中で
「提案」する。



実データ解析

ビッグデータの活用をワンストップでご支援いたします。

PJ管理

ビジネス上の課題

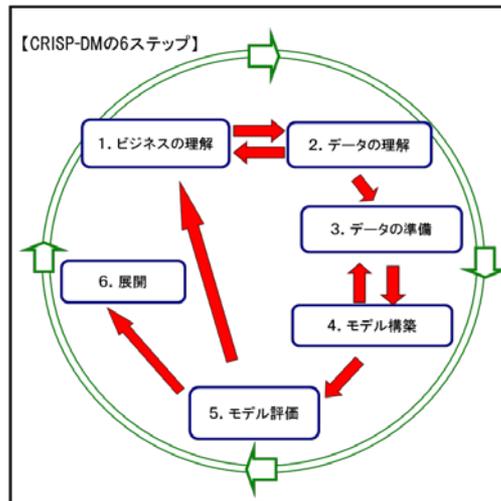


データ活用方法、
解析要件への
落とし込み

- ・ 課題管理
- ・ 施策の検討
- ・ PJ計画策定
- ・ RFP作成支援

データマイニング

CRISP-DMによる
仮説検証プロセス



- ・ 仮説検証
- ・ モデル作成
- ・ KPI設定
- ・ 定量評価

システム開発

システム要件定義

システム開発

運用・改善

- ・ DB/ エンジン設計、開発
- ・ 運用保守
- ・ 定期的なモデル更新

4. 目的を見誤らないためのデータマイニング

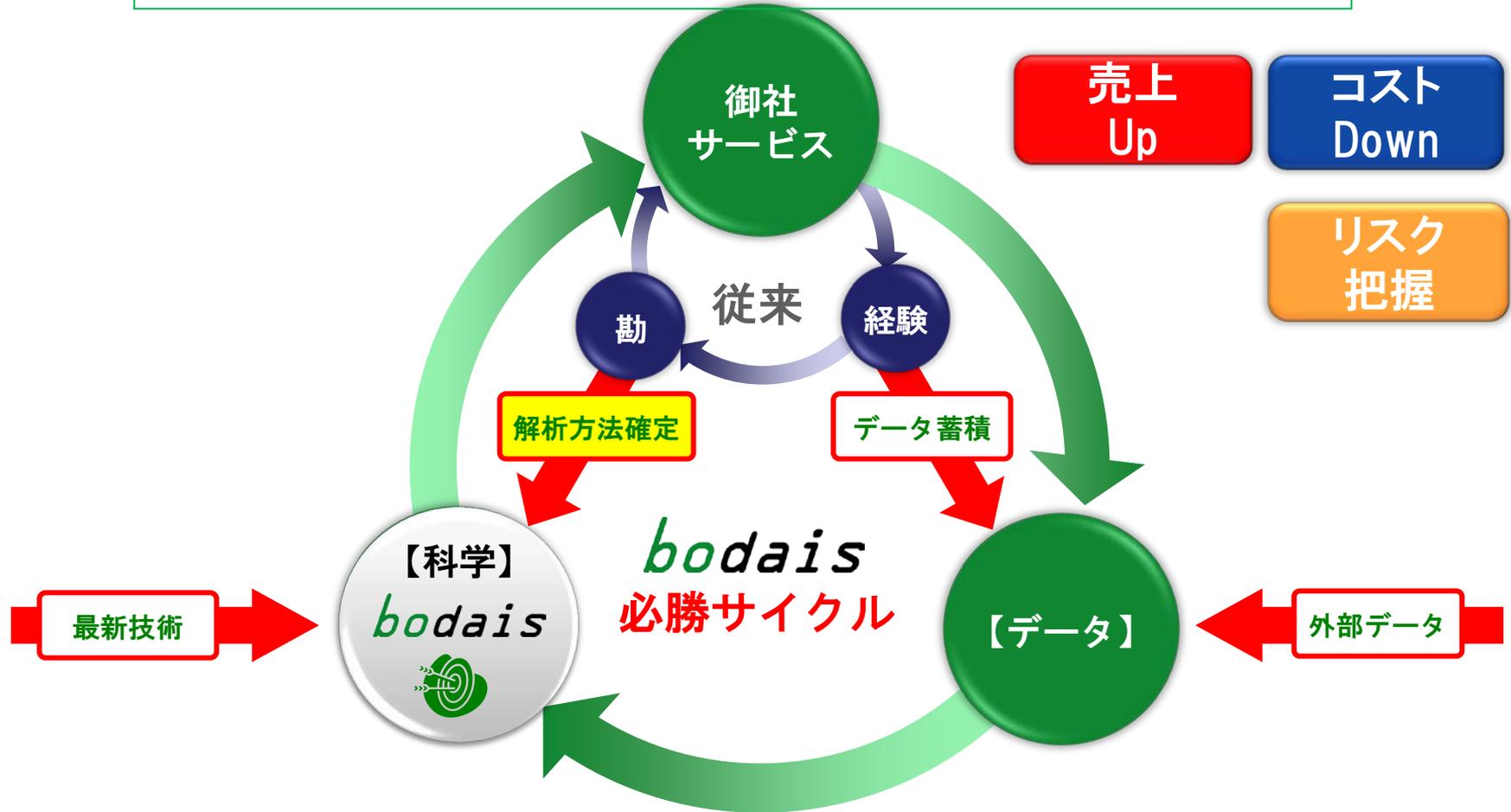
bodais

- 手法はできる限りシンプルに！
解析手法の説明が長くなると、目的を見誤る可能性大
- 効果は明確に！
効果が分からないと投資し辛い
- 利用者にやさしい！
経営者、マーケッター、解析者、それぞれの目的に対応

■ bodais必勝サイクル

bodais

企業内の「経験」や「勘」をデータとして蓄積し、ビッグデータ解析の最新技術で「売上UP」、「コストDOWN」、「リスク管理」へとビジネスの成功へ導きます。



解析の流れ

○ 過去実績ファイル

名前	性別	年齢	住所	...	入会
佐藤	男	29	東京	...	○
鈴木	女	18	大阪	...	×
高橋	男	49	愛知	...	○
田中	男	36	宮城	...	×
佐藤	男	29	東京	...	○
...					

変数が多いほど予測精度が上がります

○ A 変換済みファイル

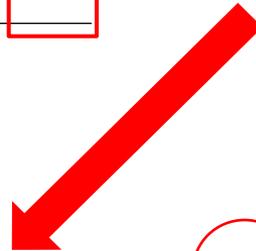
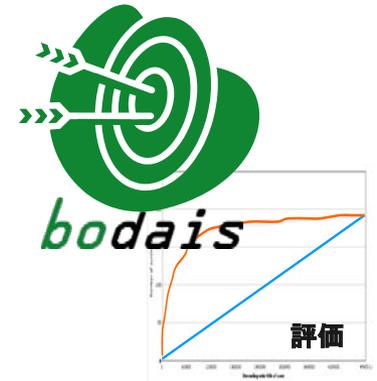
ID	性別	年齢	住所	...	正解
1	0	2	16	...	1
2	1	1	12	...	0
3	0	4	30	...	1
4	0	3	9	...	0
...					

○ F

変換ツール
(フィルタ)

- 変数は個人の特典できないものに変換
- 入会や購買など、目的とする変数を正解フラグとして利用

モデル構築



○ B 新規リスト

ID	性別	年齢	住所	...
11	1	5	2	...
12	1	4	15	...
13	0	1	22	...
14	1	1	39	...
15	1	3	10	...
...				

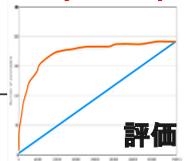
モデル適用



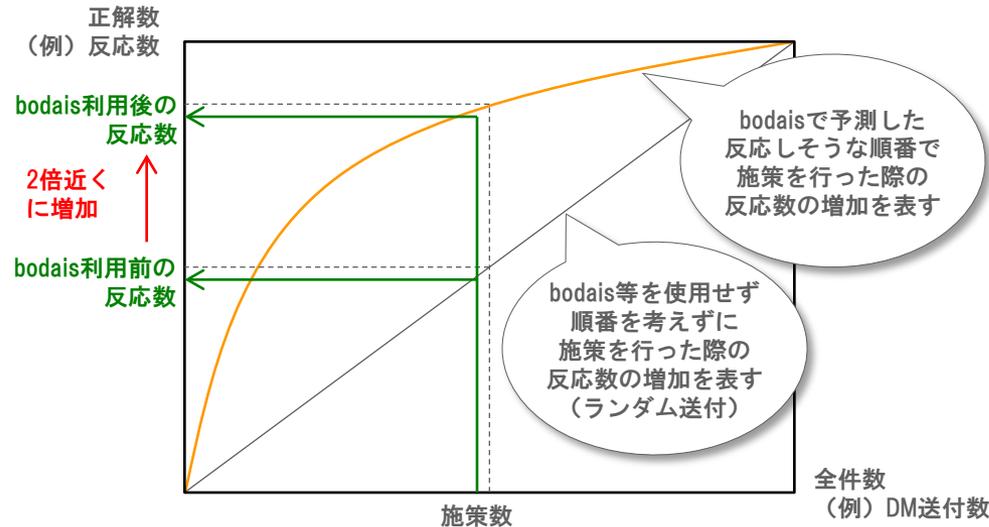
○ C 見込みランキング

ID	性別	年齢	住所	...	確率
11	1	5	2	...	0.674
12	1	4	15	...	0.265
13	0	1	22	...	0.413
14	1	1	39	...	0.064
15	1	3	10	...	
...					

各人に反応確率が付くので
アクションが明確!



【凡例】ゲインチャートの読み方

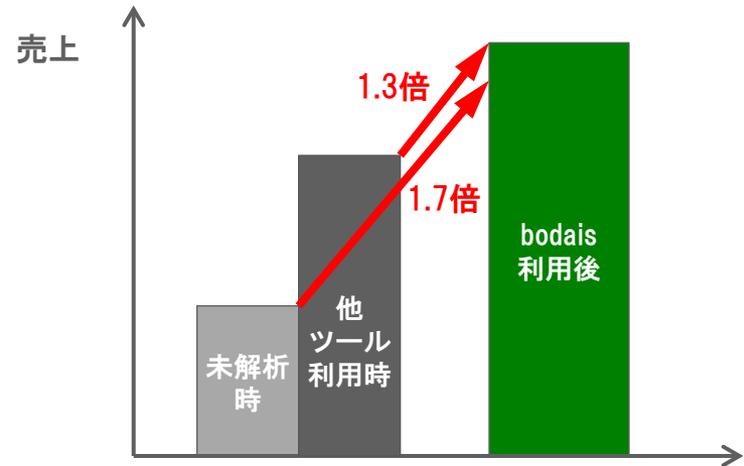
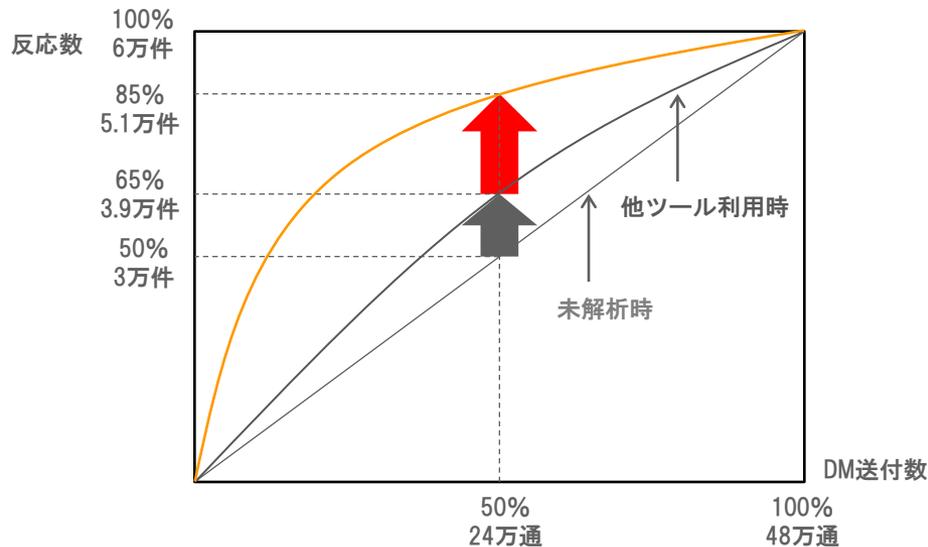


bodais利用前と後でどれくらい効率化できているか、を分かりやすく可視化するグラフがゲインチャートです。

一定数の施策を行った際に、bodais利用前後で、大きく反応数が異なっていることが一目で分かります。

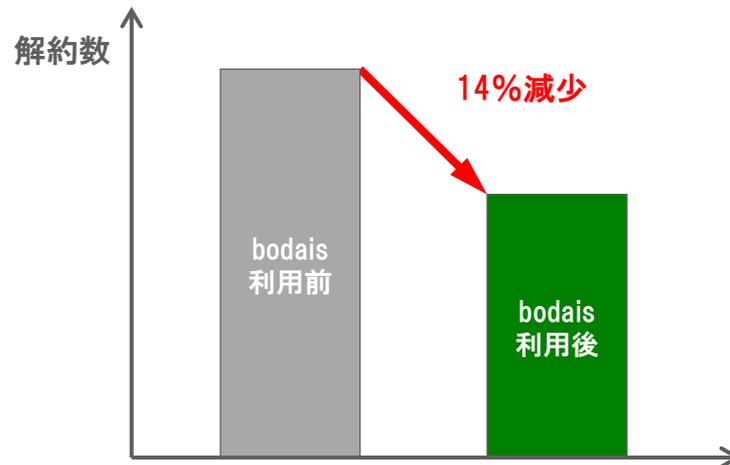
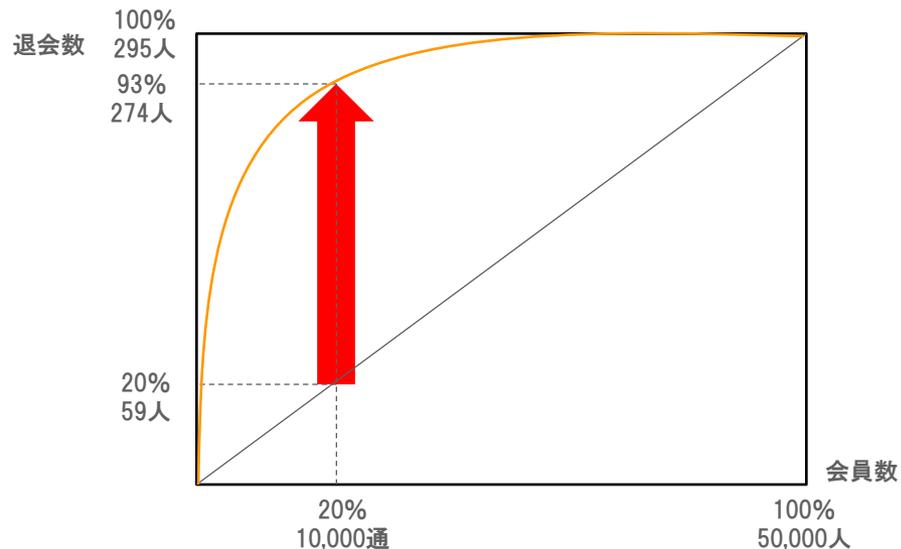
例えば左図の例では、全件の半数程度に対して施策を行った際に、利用前の反応数に比べ、利用後は2倍近くの反応数になっていることがわかります。

【A】某カタログ通販会社 A社 DMIにおける売上が1.7倍に

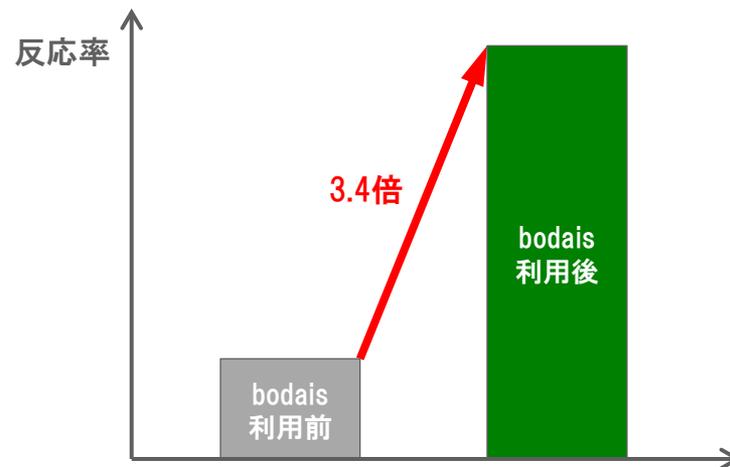
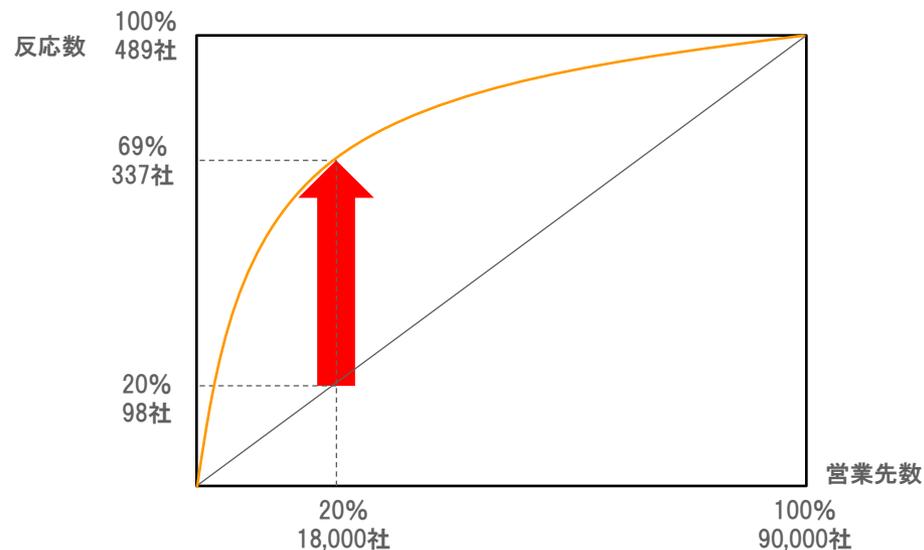


目的を見誤らないためのデータマイニング

【B】 某 会員制有料サイト会社 B社 会員の解約を14%減少



【C】 某 販売会社 C社 営業反応率が3.4倍に



効果指標：

- ・どのカテゴリがモデルの正解に効いているか（影響度）を定量的に表したものの。
モデルがブラックボックスでスコアを出力するのでなく、モデルの中身がわかります。

利用例：

- ・モデルが直感と合うか確認
各カテゴリにおける影響力の大小が、マーケッターの方の感覚と合うかご確認いただくことが可能です。
- ・施策の検討
影響力のあるカテゴリを参考にして、様々な施策を検討いただくことが可能です。

属性名	カテゴリ	効果指標    【初期に戻す】
会員属性	1	10.98
購入回数	40	7.18
年齢	20	5.65
ログイン有無	1	3.88
購入金額	20	3.87
購入金額	15	2.35
年齢	60	-2.57
購入回数	1	-2.98
購入回数	10	-3.28
ログイン有無	0	-3.88
購入金額	0	-7.40
会員属性	3	-10.56