

# IoT/BigData/AI活用を担う 人材育成の状況・課題・対応策について

2016.12.9 第一回人材育成推進会議  
株式会社オプトホールディング  
最高解析責任者CAO  
データサイエンスラボ代表  
齊藤 秀



Copyright © 2016 OPT HOLDING, Inc. All Rights Reserved.



## 齊藤 秀 (さいとう しげる)

- 博士 (システム生命科学)
- 株式会社オプトホールディング  
最高解析責任者 CAO 兼 データサイエンスラボ代表
- 東証 1 部上場企業で初のCAO (Chief Analytics Officer)
- グループ18社のビッグデータ利活用・AI実装
- ライフサイエンス & コンピュータサイエンス & マーケティング  
学部 (生物学)、大学院 (機械学習)、仕事 (ネット広告)
- ビジネス & アカデミア

国立がん研究センター客員研究員、統計数理研究所客員准教授

# 日本最大のデータサイエンティストコミュニティ

3

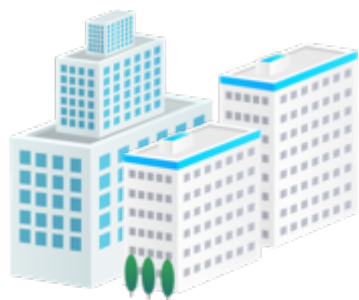


データサイエンスの  
シェアリング・エコノミー  
<https://deepanalytics.jp/>

【企業・行政】  
データ活用に悩み



【データ  
サイエンティスト】  
活躍の場が欲しい



オンラインデータ分析  
コンテスト



IoT推進ラボ  
第1回・2回コンテスト  
の運営  
多数の民間コンテスト



ソースコード  
各種権利  
アイデア  
ノウハウ

様々な領域の  
データサイエンティスト  
学生・研究者・大企業  
**約2700人**



課題の設定・データ設計・コンテスト開催・検収  
弊社のデータサイエンティストがサポート

## 2014.10.18：ビッグデータ

**66%**

“インダストリアル・インターネット戦略を推進するためにビッグデータを活用しなかった場合、今後1年から3年以内に現在の市場ポジションを失うことになる”

**93%**

“ビッグデータを活用することで自社を差別化している企業がすでに市場に新規参入している”

**88%**

“ビッグデータ活用が自社にとっての最優先事項である”

**29%**

“「予測分析」や自社ビジネス最適化のために全社横断的にビッグデータを活用している”

※世界7カ国250名の経営幹部, 売上げ規模1億5千万ドル以上, 半数が10億ドル以上へのヒアリング  
【引用】GEアクセンチュア, “Industrial Internet Insights for 2015”

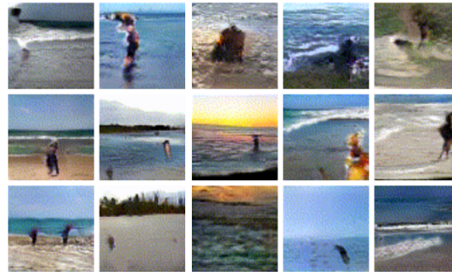
## 2016.12.9 : AI

- 画像認識・言語翻訳・生成モデル・ゲーム・自動運転・ロボティクス…
- 単純にUXとして“おおっ”というレベルの事例が出てきている

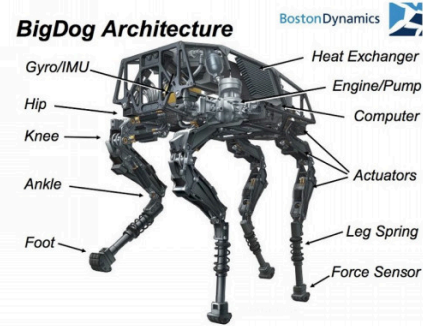
### ゲーム



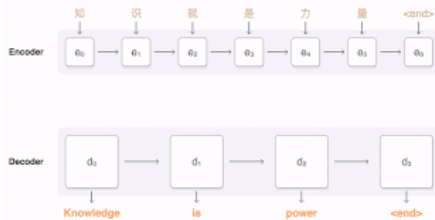
### 画像認識・動画生成



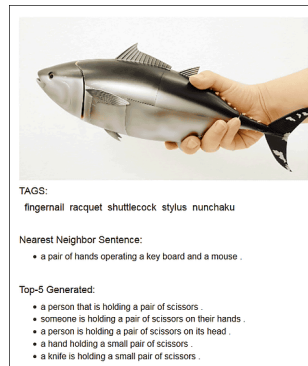
### ロボット・自動運転



### Chatbot・多言語翻訳



### 文章生成・作曲・絵画



IoT

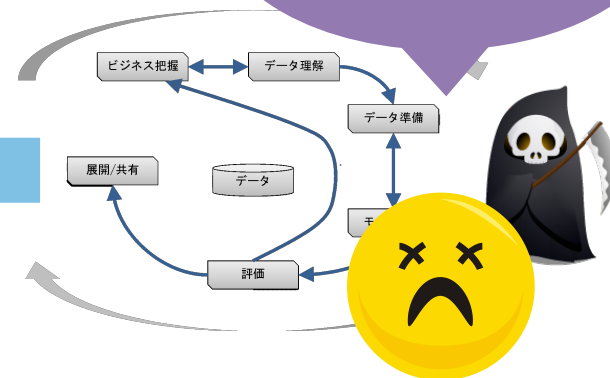
Big  
Data

属人的  
労働集約

## 【必要条件】

1. 良質かつ大規模なデータの整備
2. スケーラブルなデータ分析体制

AI



【問題点】  
どちらも  
人海戦術

## Exponential

私たちが直面している技術革新に伴う社会変化は、時間に対して指数的に進んでいます。

- ・ 産業革命の間隔
- ・ 世界のデータ生成量
- ・ 半導体の集積密度（ムーアの法則）

...

その先に訪れる技術的特異点：Singularity？

- ✓ 技術トレンド・社会変化のスピードが速すぎる
- ✓ IT/データサイエンス人材の定義が安定しない
- ✓ 専門性が曖昧（何でも屋）になるほどキャリアが安定する
- ✓ イノベーションは既知技術の組み合わせ  
→ 水平的な専門性・組織体制が重要

## 【一般論】

- データサイエンティストは今後も不足し続ける
  - 米国：14～19万人（McKinsey Global Institute, 2011）
  - 日本：25万人（Gartner, 2011）

## 【実際】 大手人材仲介企業

- 転職市場でデータサイエンティストの支援に苦戦している
  - 求人は出ているものの成約率が低い
  - 求人側の要件が曖昧
  - 求職者の評価が困難
  - 求職者も貢献の具体性に欠ける
    - なんでもいいからデータ分析の仕事をしてみたい
    - 最新技術を使うことが目的化されている



## 課題

産業界は問題解決能力を持つデータサイエンティストを欲しているが、現状では**統計教育が数理統計に偏っており、実務における問題解決能力と必ずしも結びついていない。**

## 棟梁（500人規模）

社会人に対し最先端の手法を**PBL**で実地経験させ、棟梁レベルの人的ネットワークを形成。

## 独り立ち（5000人規模）

大学院において**PBL**に基づく専門育成プログラムを推進。

## 出口戦略

**育成された人材は、民間・アカデミア問わず広く活用されなければならない。**

## 提言

社会全体のリテラシーを向上させるために(中略)**コンテストの開催を行う。**

データサイエンティスト協会が定めたスキルレベル（2014年12月）

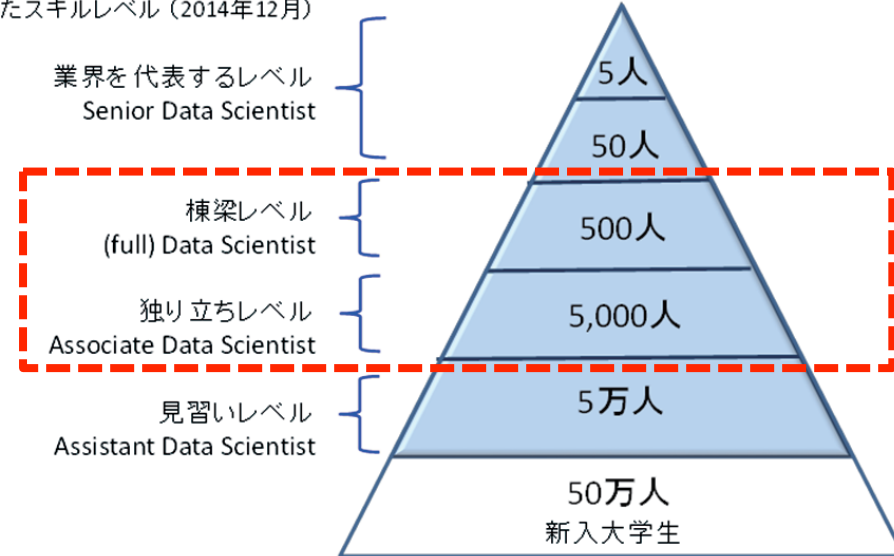


図2. 育成レベルと、毎年の育成目標人数

【引用】ビッグデータの利活用のための専門人材育成について  
H27.7.30 大学共同利用機関法人 情報・システム研究機構  
ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会

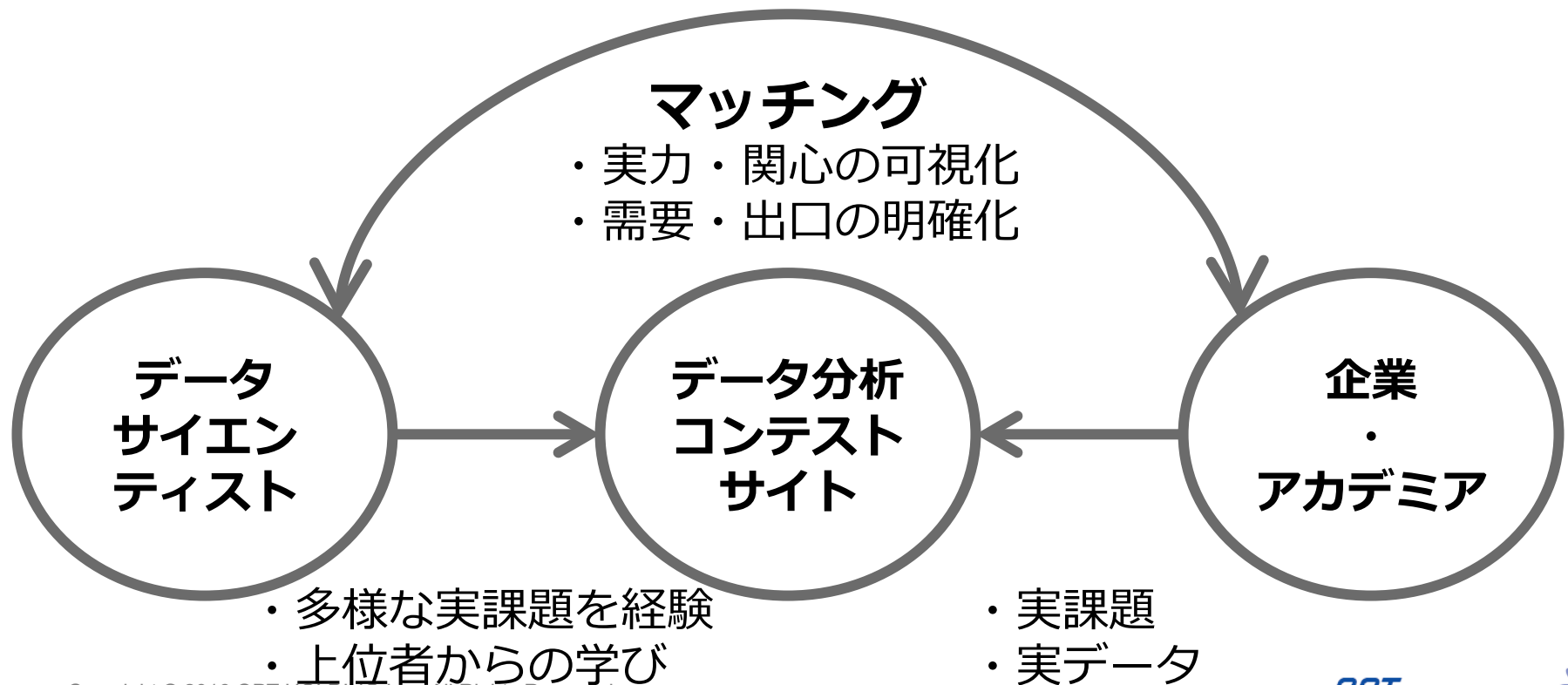
- **建築家：データを材料に価値ある出力をデザイン**
  - 市場と科学がわかる人：全体の設計（サーベイ・実験計画・データ収集・分析・仮説創出・前向き実験）
  - 意思決定者に近い人：社会的・組織的・ビジネス的制約の中での説得・遂行
  - **ステークホルダーを説き伏せ社会実験を実装できる人**
- **材料屋：データを生む人**
  - データ整備従事者：人海戦術・クラウドソーシング
  - 職人：教師になれる人、その人の思考・行動が教師データ
- **道具屋・職人：データから創る人**
  - AI研究者：原理解明・応用技術創出・ライブラリ
  - データサイエンティスト：産業での利活用・Fine tuning
- **管理人：創ったものを保守・運用する人**
  - AIセキュリティ：悪用や攻撃を防ぐ
  - AIメンテナンス：品質の劣化を防ぐ、定期検診

- 見習い→独り立ち→棟梁あたりの教育
- 実践を通じた教育・評価・出口

【教育】 実際の課題・データに向き合うことによる総合的な教育

【評価】 精度で測る、分析の過程も可視化

【出口】 可視化された成果、問題に取り組む情熱



# 教育活動

# 米国におけるコンテストサイトの教育利用実態



Stanford



UCB



UCSD



UCI



CalTech



Columbia



Cornell



NYU



Michigan



GA Tech



UMass



Purdue



CB



CD



WUS



Virginia Tech



Rutgers



North eastern



ISU



SCU



Texas A&M



Rice



Trinity



UIUC



OHSU



Wake Forest



SLU



UMCP



Delaware



Wayne State



UNCC



ASU



GMU

Stanford, UC Berkeley, Caltech, Columbia, Cornell, New York 等  
全米の33の大学がPBL/実習で利用



## 滋賀大モデル

**データアナリシス**  
大規模データを分析・解析  
するための知識とスキル  
(統計学)

**データエンジニアリング**  
大規模データを加工・処理する  
ための知識とスキル  
(コンピュータ科学)

新たな知見

価値創造

ビジネスや政策などの領域で課題を読み取り、データエンジニアリングとデータアナリシスにより得られた知見を現場の意思決定に生かす  
(領域分野での成功体験 = PBLの繰り返し)

- 価値創造を重視
- 職場で、一人でもデータ分析を担当できる人材
- コミュニケーション力を重視

アメリカでは  
統計 + 計算機  
だけで就職できる

滋賀大学データサイエンス学部 (H29/4~)

データサイエンスに焦点を合わせた日本初の本格的な学部の新設  
データサイエンティスト養成のための特色ある文理融合教育プログラムの展開

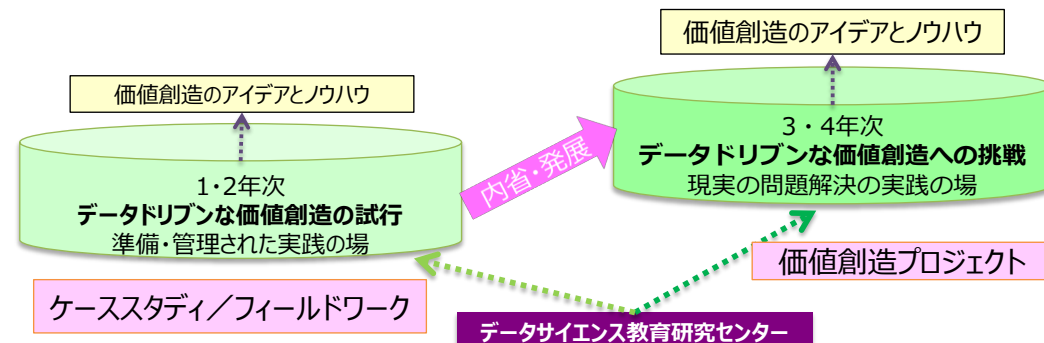
PBL演習(Project-Based Learning, 課題解決型学習)

1年次から4年次まで通して、実際のデータを扱った価値創造を体験するためのPBL型の演習を行います。

この演習の中では、様々なデータサイエンスの実践例の紹介から始まり、データサイエンスが使われている様々な現場の見学、データ分析の試行錯誤を重ねながら、価値創造の経験の積み重ね、最終的には実データを使った価値創造(現場の意思決定に生かす実際の提案)を行うプロジェクトを行います。

## データ駆動型価値創造PBL演習

- 演習では、実際のビジネスの現場で得られるデータを扱うことにより、**実際のデータ**から価値創造を経験
- さまざまな**企業**や地方公共団体との**連携**を依頼中



【引用】データサイエンティストの育成のための大学教育 - 滋賀大学データサイエンス学部のモデル



<http://www.opt.ne.jp/holding/news/gourp/detail/id=3755>

**2016/9/30 日本初のデータサイエンス学部を設置する  
滋賀大学と協定締結。コンテストサイトをPBLに活用予定。  
東京大学・筑波大学・慶應大学等、複数の大学と協定調整中。**



**DataScientist  
Society**



[http://www.datascientist.or.jp/dssjournal/2016/05/26/lecturereport\\_n01/](http://www.datascientist.or.jp/dssjournal/2016/05/26/lecturereport_n01/)

## データサイエンティスト養成講座

～データを分析してディスカッションしよう～

カリキュラム  
の特徴

座学では学習しづらい、分析データのクリーニングや、データハンドリング、様々な分析手法の選択などを、実際に課題を解決する中で学習して身につけることができます。

開催日時

【開校式】

2016年5月18日(水) 19時00分～19時30分

【講座（計8回）】

2016年5月18日(水)・5月25日(水)・6月8日(水)・6月22日(水)  
7月6日(水)・7月20日(水)・8月3日(水)・8月17日(水)  
各回 19時00分～21時00分（毎回2時間）

【修了式】

2016年8月17日(水) 21時00分～21時30分

会 場

東京都渋谷区または、東京都心部

受 講 費

法人会員：40,000円（税別）

定 員

30名

**12社30名**

**データサイエンティスト協会にコンテンツ機能を提供  
共同でデータ分析の実践的教育を実施！**



- 入賞者のモデリング方法から学ぶ
- 分析対象・データタイプ・業種等で、うまくいく特徴設計・アルゴリズムの傾向がわかる
- Deep Learningがうまくいくケース、行かないケース。うまくいく場合のアーキテクチャ・学習法
- 分析視点・解釈、外部データの利用
- Overfitting・Imbalanced data等、ミスしがちなケースを経験

...

実例から学ぶことは多い。

※過去コンテストの入賞者ノウハウを公開（右図）、データを教材として活用

順位	概要	こだわり・工夫
 1位 : toshi_k 	・データの特性を生かした変数選択 (工事の除外/夜中2時での切替) ・予測値を時系列に並べ平滑化(補正)	変数 モデル 後処理
 2位 : neko 	・変数はとにかく投入 ・4モデルを統合しパラメータで調整	変数 モデル 後処理
 3位 : marugari 	・変数選択を熟慮したモデル	変数 モデル 後処理

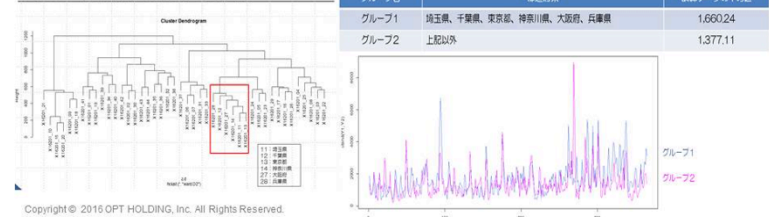
Copyright © 2014 OPT, Inc. All Rights Reserved.



$$Y_t \sim N(\mu_t + week_t + H \cdot holiday_t + L \cdot level_t, \sigma_t^2)$$

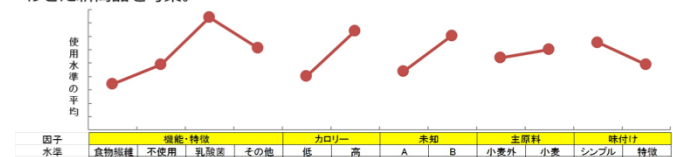
①トレンド項    ②曜日項    ③連休項    ④噴火警戒レベル項

クラスター分析結果 富山 (From 国内)



Copyright © 2016 OPT HOLDING, Inc. All Rights Reserved.

コンジョイント分析により、乳酸菌、カロリー高、小麦と味付けがシンプルといった要素や、商品をシンプルに説明するネーミング等が重要と分析。それらを組み合わせた新商品を考案。



# 評価・出口戦略

- **海外では人材獲得の1手法として定着**
  - KDD Cupのバンケットはヘッドハンターの狩場
  - Facebook, Yelp等は自社でホストする分析コンテストの上位者から面接
  - コンテストの実績をCVに記載すると価値が上がる
- **国内でも活用がはじまりつつある**
  - 弊社への問い合わせが多いのは、人材採用の話
  - IoT推進ラボ第1回ビッグデータコンテストの優勝者は就活中の学生、同結果が有利に働き、大企業内定を多数獲得
    - 同じく入賞者である銀行のクウォンツは、社内で注目され、新規データサイエンスプロジェクトの責任者に任命された

**勝者は必ずしも専門分野の人材ではない。業種・専門を超えた人材のマッチングがコンテストを通じて起きている！！**

**人材の流動化、multi majorへのきっかけ！**



採用目的の分析コンテストを開催  
Facebook, Yelp, Walmart, Airbnb, Telstra  
登録データサイエンティストは約60万人  
求人広告で需給マッチング



過去7回の分析コンテストを開催

- ・ 課題/データはアリババGから提供
- ・ 学生向けの教育志向

500万円以上の賞金が出ている  
登録学生は約2万人

## ビッグデータ先進国では人材採用もオープンな形式で実施！

# DeepAnalyticsを活用した採用コンテスト

用意された枠内の項目をラベリングする



元の名刺

サンサングローバル株式会社
取締役副社長
ジョージ ネルソン
〒150-0001
東京都渋谷区神宮前 5-52-2 青山オーバルビル 13F
TEL 03-6758-0033   FAX 03-3409-3133
MOBILE 090-1234-5678
E-mail info@33i.co.jp
SITE http://jp.corp-sansan.com/

ラベリング後の名刺

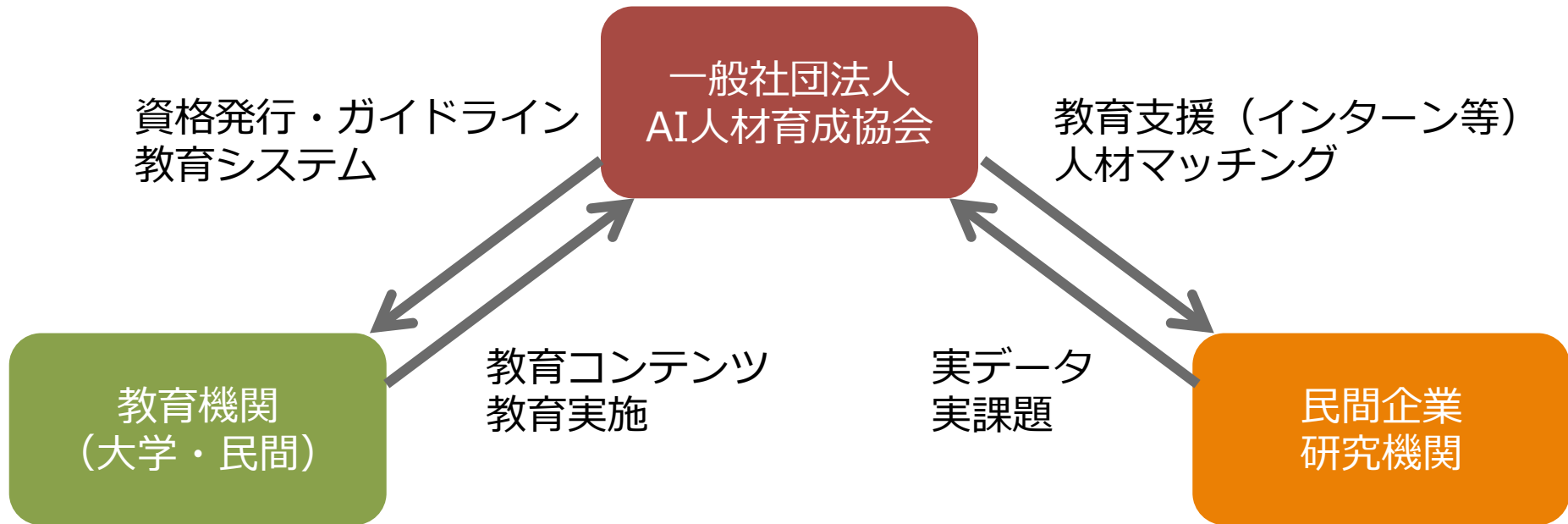
Position	取締役副社長
Name	ジョージ ネルソン
Zipcode	〒150-0001
Address	東京都渋谷区神宮前 5-52-2 青山オーバルビル 13F
Tel	TEL 03-6758-0033
E-mail	MOBILE 090-1234-5678
URL	E-mail info@33i.co.jp
	SITE http://jp.corp-sansan.com/
	Fax FAX 03-3409-3133
	Mobile
	company サンサングローバル株式会社

<https://deepanalytics.jp/compe/26>

日本初のコンテスト形式の採用イベント  
画像処理 X 機械学習 のスキルを持つ技術者が欲しい  
約240名が個人情報を入力し参加  
50人が分析、14人が社内の分析精度を凌駕！

# その他の活動

事業・研究等を通じた社会へのIoT/ビッグデータ/AIの  
戦略的・実際の活用を担う人材の輩出を目的とした協会



- インターンシップ等の民間企業・研究機関におけるAI・データ人材育成の支援
- 教育機関への教育システムの提供・システムを通じた資格の認定、資格に対応するカリキュラムガイドラインの提供

設立準備中

- AI・データ活用は依然として重要
- 関連人材は継続して世界的に不足
- 実践的教育が求められている
- 米国ではデータ分析コンテスト機能を活用した教育が実施されている
- 日本でも、大学のPBL・社会人向けの検証を始めている
- コンテスト活用はリテラシーの向上に加え出口戦略にも有効
- AI人材の育成を支援する社団法人の設立を準備中



# 補足資料

## 観光地の宿泊者数予測

IoTビッグデータの推進の活動として、「観光」をテーマとしたビッグデータ分析コンテストをIoT推進ラボ・経済産業省共催で開催。

全国の観光地の宿泊者数を、過去の実績やSNSデータ、為替データ、気象データ等、観光地に紐づく様々なデータを用いて予測。



期間：1.5ヶ月  
参加者：129人  
分析回数：2,808回

※経済産業大臣による表彰

## 豆腐の需要予測

従来は工場長の判断で計画をたてていた豆腐5種類の需要量を、過去実績や特売情報、気象データを用いて予測。

「食品ロス」という大きな社会問題とも密接に関連する問題解決をデータサイエンスで挑み、最大で15%の精度改善が得られた。



期間：1.5ヶ月  
参加者：114人  
分析回数：1,665回

データサイエンスアワード  
2016 奨励賞受賞！

## 健康菓子の売上予測、新商品開発

IoT推進ラボ・経済産業省共催によるコンテスト第2弾として、POSデータや商品特性データ等を用い、ナチュラルローソン菓子の売上予測、新商品の開発をテーマとした分析コンテストを実施。



期間：約2ヶ月  
参加者：150人  
分析回数：2,250回

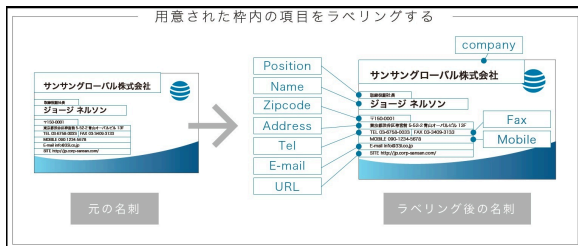


※CEATEC JAPAN 2016での表彰式

## 名刺のラベリング

名刺の画像データと項目の位置情報から、各項目のラベルの内容を予測。日本初のコンテスト形式の採用イベント。

期間：2ヶ月  
参加者：49人  
分析回数：630回



## インターネット広告のクリック予測

インターネット広告の配信精度の改善を目的に、様々な消費者行動データから、消費者の広告への関心を予測。

現行の予測アルゴリズムに対し30%の精度改善が得られた。



期間：2ヶ月  
参加者：77人  
分析回数：849回

## 気象データによる鉄道支障予測

携帯電波基地局で収集している気象データの有効活用の取り組みとして、過去の気象データから鉄道支障予測を実施。

気象要因障害については、80%の予測精度が得られた。

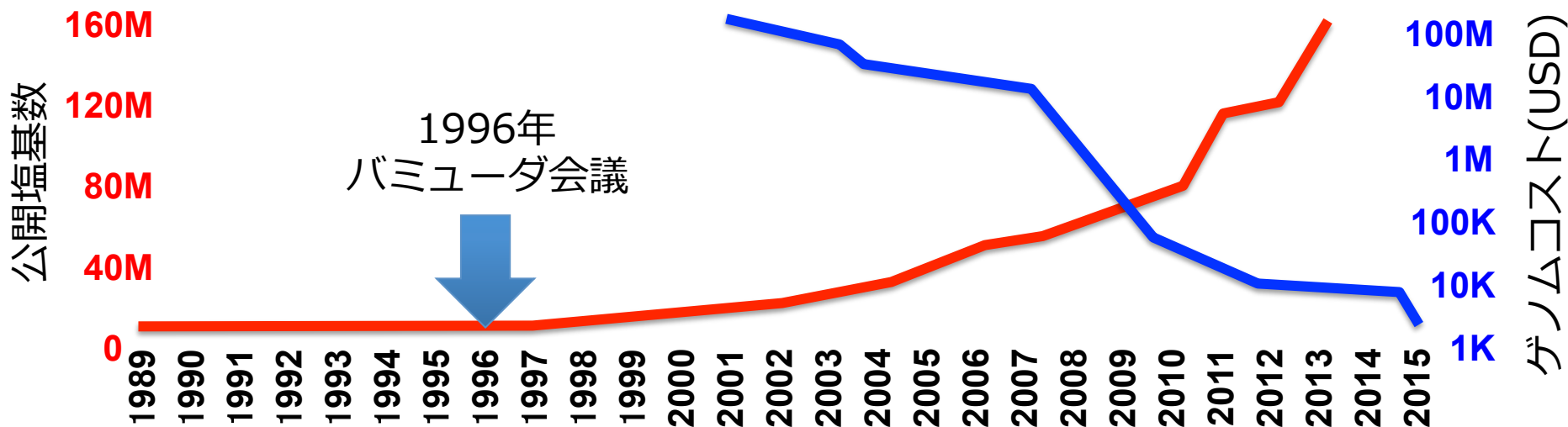


期間：2ヶ月  
参加者：48人  
分析回数：683回



# 例) ライフサイエンス・医薬品領域の状況

27



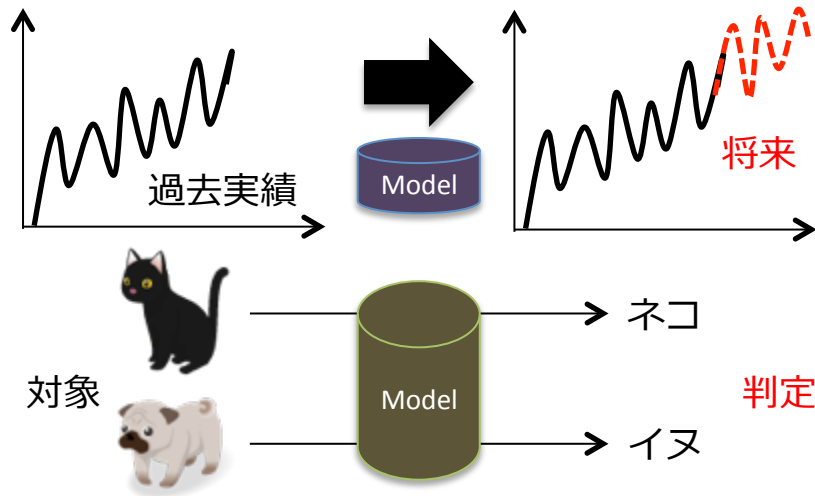
- 1996 : 遺伝子データのオープン化 (バミューダ協定)
  - 2006 : iPS細胞樹立 (データベースの活用)  
【AI創薬スタートアップの台頭】
  - 2007 : Numerate社, 17.42M\$
  - 2012 : Atomwise社, 6.35M\$
  - 2013 : Human Longevity社, Craig Venter氏, 300M\$
  - 医薬品業 : 基本Chemistry, 1980以降=Biotech, 今後=Bigdata/AI
- ※日本 : 本領域のAI研究・産業利用は非常に遅れており、人材も不足している  
→生物系ポスドクのAI技術武装が有効か

# データ分析コンテストの仕組み

## 予測・分類

過去のデータに基づき  
将来を予測する分析  
対象を分類する分析

- 例) ・ 需要予測、リスク予測  
・ 異常検知、画像認識



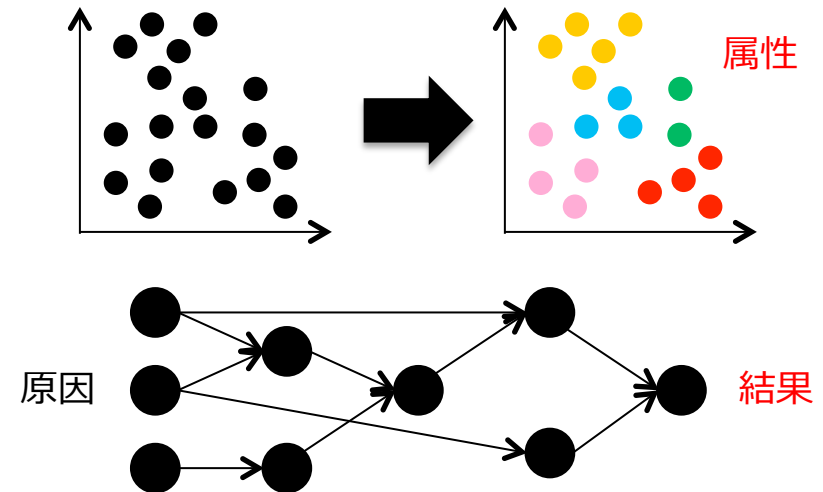
成果物)

- ・ 予測アルゴリズム、ソフトウェア
- ・ 分類アルゴリズム、ソフトウェア

## 理解・発見

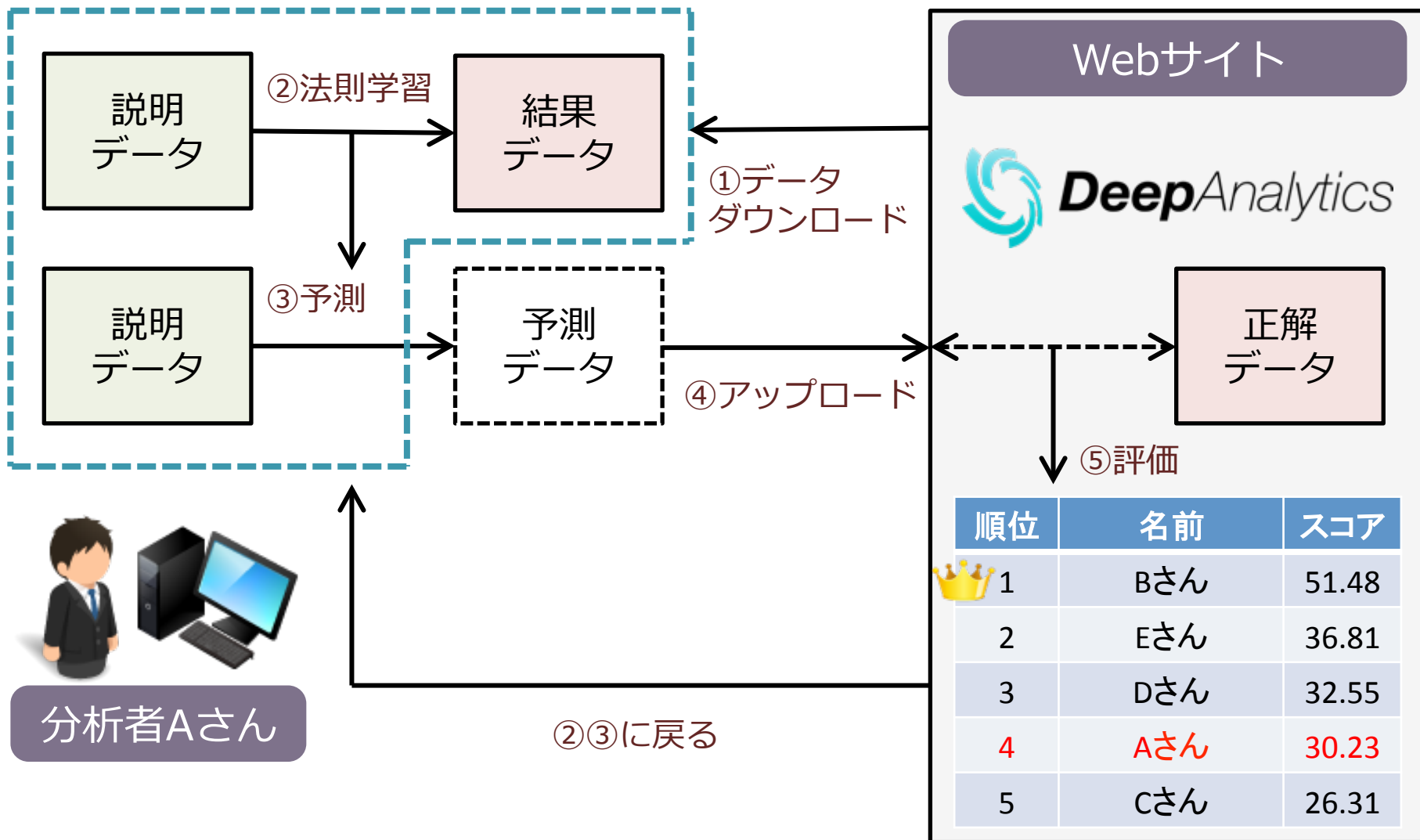
データからルール・パターン・類似性・関係性などを発見する分析  
データマイニング

- 例) ・ ユーザー行動分析  
・ アンケート分析



成果物)

- ・ 現状理解の確認、課題の裏づけ
- ・ 未知の法則発見、課題の発見



ゲーミフィケーションが前人未至の分析精度へと誘う！

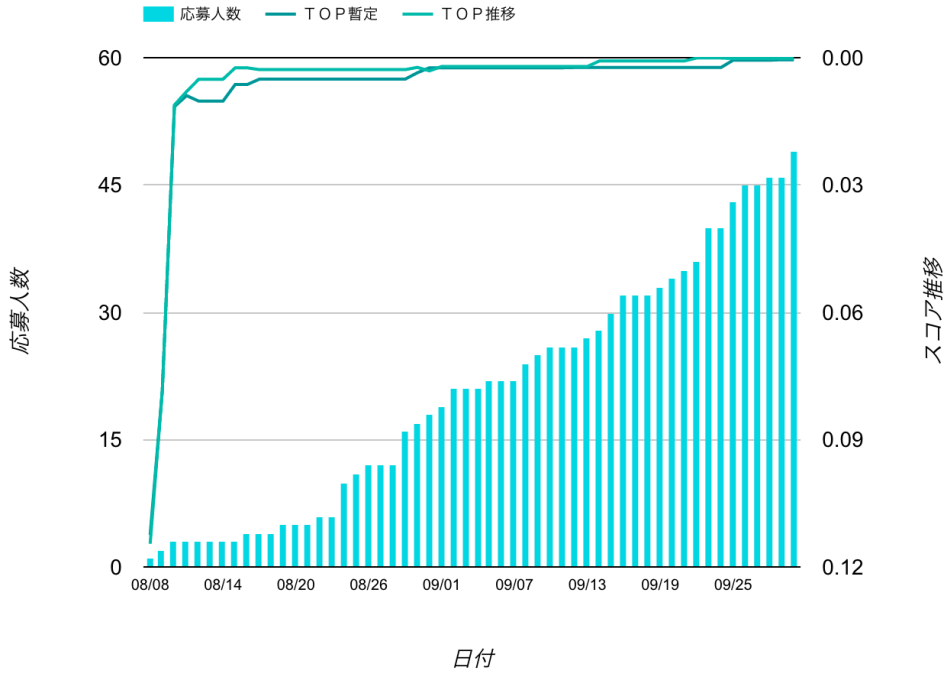
## スコアボード

応募件数：2226件  
 応募人数：138人

※順位のカッコ内は1週間前の順位からのランクのアップダウンを表しています。

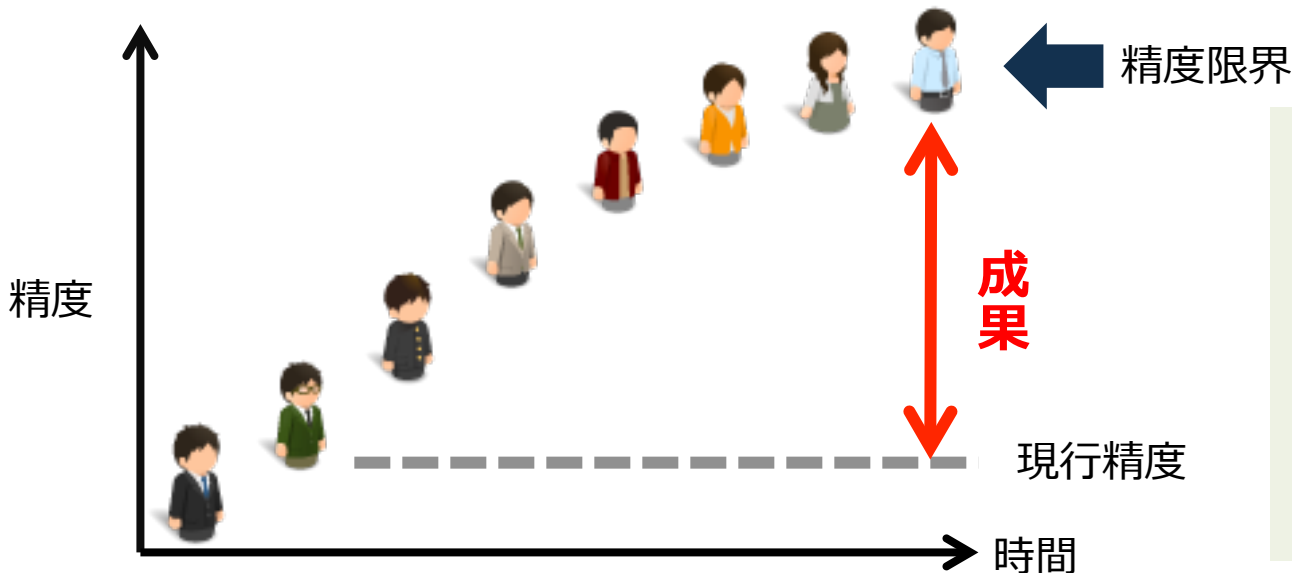
順位	ユーザー名	スコア	応募件数	投稿日時
1	deeshi	0.14806	60	2016/08/30 02:14
2	moemoeと遊び隊	0.14895	64	2016/09/01 23:07
3	t_iwmr	0.14952	127	
4	imenurok	0.15056	87	
5	keizoku	0.15079	48	
6	ZABURO	0.15106	32	
7	tofu	0.15172	16	
8	atg	0.15188	1	

リアルタイムで  
 成績が出るため、  
 悔しくて何回も  
 挑戦！



時間とともに参加者が増え  
 分析精度も上がっていきます！

- 通常では試せない膨大な分析パターンから  
**最高精度の分析**結果が得られます！
- 成果から逆算して投資額を設定する  
**ROIの高い分析**が実現できます！
- 分析精度の向上プロセス・精度限界が可視化され、  
**プロジェクトの継続判断**が可能になります！



【過去実績】  
138人\*平均35時間  
= 4830時間  
= 約600人日  
= 約30人月



- オンデマンドで変化する課題に応じた  
**最適な人材**に分析を依頼できます！
- 過去多くのコンテストで、問題領域の  
**専門外**の分析者が勝利している！
- 人材採用を検討している場合、  
**実際の分析業務ができる人材**にリーチできます！

マーケ

通信

Web

医療

金融

自動車

製造

研究機関

- ビッグデータ・人工知能といった最先端技術に取り組む  
**企業ブランド**を周知できます！
- データ販売企業の場合  
**データの活用事例**を発信できます！



dunnhumby



Genentech  
A Member of the Roche Group



GIGAOM



information security  
amazon



海外では多くの  
大企業・研究機関が  
積極的に活用！  
メディア発信することで  
先端的取組を  
PRしています！

※米Kaggle社から引用

- 投資対効果が出にくいモデル探索フェーズは**オープン**で構築  
調達したモデルからノウハウを抽出、社内R&Dと  
ビッグデータを投入し、**クローズ**で高度化！
- 社内データサイエンティスト・研究者は全体戦略立案・設計  
といった**上流工程**にリソースをシフト
- 事実、海外では、**トップデータサイエンス**企業ほどオープン  
イノベーションを活用している！！

## 米Netflix社

- ✓レコメンデーション技術の運用・改善に**300人**の専門家  
**年間1.5億ドル(約185億円)**もの経費をかけている
- ✓約**1億円**の懸賞金をかけたコンテストを開催  
同エンジンにより**年間90億円**の売上向上

ご清聴ありがとうございました

データを紡いで、未来を編む。

時代をとらえデータを見つめる。

データをつなぎ科学する。

科学を通してビジョンを叶える。

私たちはe-マーケティングで、

社会を動かす新たな価値を生みだしていく。

株式会社オプトホールディング データサイエンスラボ

〒102-0081 東京都千代田区四番町6 東急番町ビル

お問い合わせ: <https://deepanalytics.jp/inquiry>