

文部科学省委託事業 データサイエンティスト育成ネットワーク形成 事業報告



丸山宏
情報・システム研究機構 統計数理研究所

事業の概要

- 目的

- ビッグデータを利活用しイノベーションを促進する人材を育成するネットワークを形成する

- 期間

- 平成25年7月～平成28年3月

- 体制

- 受託者：大学共同利用機関法人 情報・システム研究機構 統計数理研究所

- 再委託先：国立大学法人 東京大学 情報理工学系研究科

- 考慮点

- スケーラビリティ

- 出口戦略

- 我が国の実情に合った育成

- 事業の5つの柱

- 1. 認知度向上・啓発

- 2. 人材ローテーションに基づく育成

- 3. ベストプラクティスの調査

- 4. 教材の開発・展開

- 5. ネットワーク形成及び標準化の検討

H25年度の主要な成果

1. データサイエンティスト現状調査

- 統計検定合格者に対するアンケート調査
- 20名の「データサイエンティスト」に対する聞き取り調査
- 米国シリコンバレーにおける現状調査

2. インターンシップ・パイロットプログラムの実施

- 東大学生4名

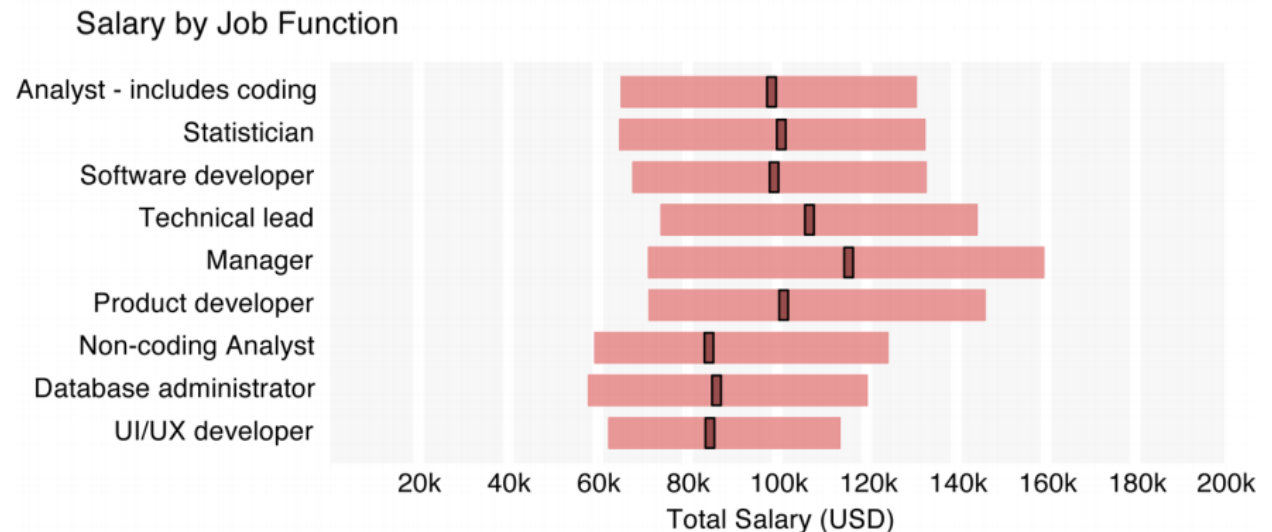
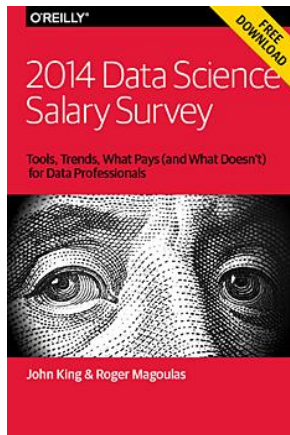
3. オンライン教材の作成

- 「データサイエンティスト・クラッシュコース」

米国での注目



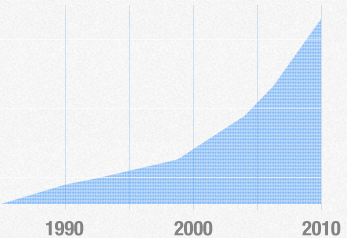
“Data Scientist: The Sexiest Job of the 21st Century”



<http://radar.oreilly.com/2014/12/2014-data-science-salary-survey.html>

米国における育成取り組み例: INSIGHT DATA SCIENCE FELLOWS PROGRAM

ANALYTICS AND DATA SCIENCE JOB GROWTH



1990 2000 2010

7 WEEK FULL TIME TRAINING

FUNDED PHD FELLOWSHIP

FOUNDED IN

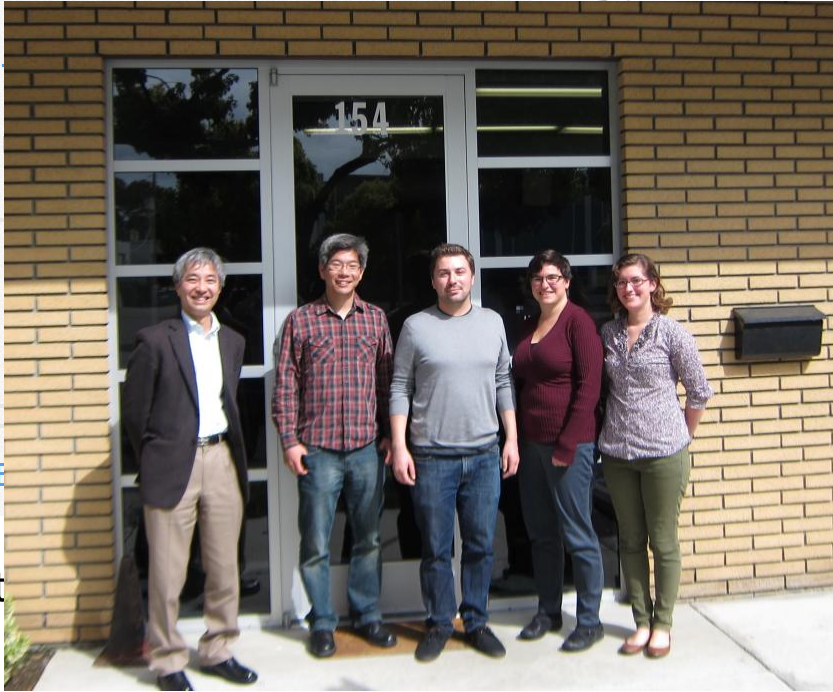
2012

ALUMNI COMMUNITY

200+

LOCATIONS

SILICON VALLEY



Mentor Companies



<http://insightdata.org>

“Fellow”の多くは自然科学の博士号取得者



Julia
Statistics, PostDoc
Insight Fellow 2012
Data Scientist, Adobe



Nick
Physics, PostDoc
Insight Fellow 2012
Quantitative Analyst,
Youtube



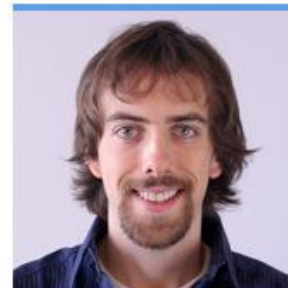
Lauren
Astrophysics, PhD
Insight Fellow 2013
Data Scientist,
Groupon



Krystal
Biology, PhD
Insight Fellow 2013
Data Analyst, Modcloth



Megan
Space Physics,
PostDoc
Insight Fellow 2013
Quantitative Analyst,
One Kings Lane

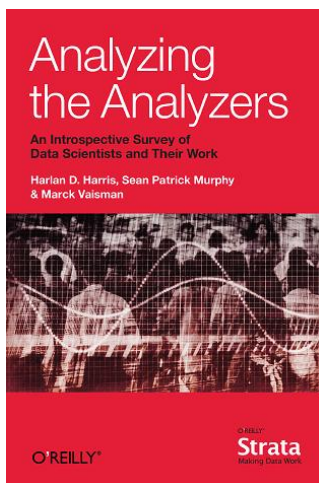


Doug
Physics, PhD
Insight Fellow 2012
Data Scientist, Twitter

“Data Product”の例: CouchTube



データサイエンティストとは、「データプロダクト(動くシステム)」を作る人



米国におけるデータサイエンティストのサーベイ

- Web上でのアンケート(KwikSurveys.comを利用)、5ページ、平均記入時間10分
- 回答者: 250名
- スキル、経験、学歴、自己イメージ、Webプレゼンス(LinkedInなど)の5項目

<http://oreilly.com/data/stratareports/analyzing-the-analyzers.csp>

Business	ML / Big Data	Math / OR	Programming	Statistics
Product Development	Unstructured Data	Optimization	Systems Administration	Visualization
Business	Structured Data	Math	Back End Programming	Temporal Statistics
	Machine Learning	Graphical Models	Front End Programming	Surveys and Marketing
	Big and Distributed Data	Bayesian / Monte Carlo Statistics		Spatial Statistics
		Algorithms		Science
		Simulation		Data Manipulation
				Classical Statistics

Developer	Engineer	
Researcher	Scientist	Statistician
Jack of All Trades	Artist	Hacker
Leader	Businessperson	Entrepreneur

自己イメージの選択項目
(5レベルの選択)

Data Scientist 4つのタイプ

Data Businesspeople

- 経営工学MBA
- コンサルティング会社での経験
- 現在大企業のデータ分析チームマネージャー
- チームと経営層との間の通訳として役割
- データを自分で触りたいが、時間は限られている

Data Researcher

- 分子生物学で博士号を取得
- 元々はアカデミア指向
- 現在は国際的な流通会社でデータサイエンティスト
- 論文は書けるが、マネジメントの経験はない

Data Creatives

- 経済・CS・統計
- 統計コンサルのベンチャーを起業
- 現在大手新聞社に勤務
- 夜はPythonのオープンソース開発
- 自身はハッカーと思っている

Data Developer

- CS修士
- 現在中堅のコンサルファームで機械学習ツールの開発に従事
- 製品レベルのコードが書ける

日本の現状：H25年度に行った調査

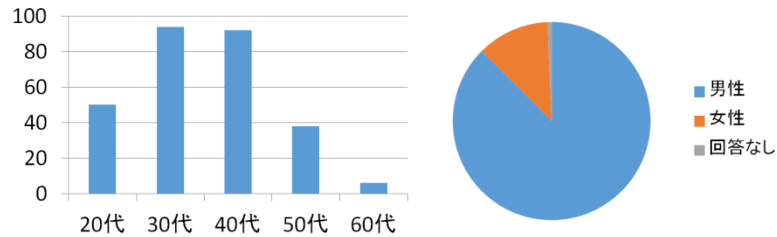
- 定量的調査：統計検定受験者に対するアンケート (319名)
- 定性的調査：“データサイエンティスト”に対する聞き取り調査(20名)
 - 業種：金融、製造、流通、公共、ITベンダー、コンサルティング
 - 規模：個人、ベンチャー、上場企業
 - 役割：部署内分析者、社内コンサルティング、社外コンサルティング、フリーランス



<http://datascientist.ism.ac.jp/pdf/H25DSTN.pdf>

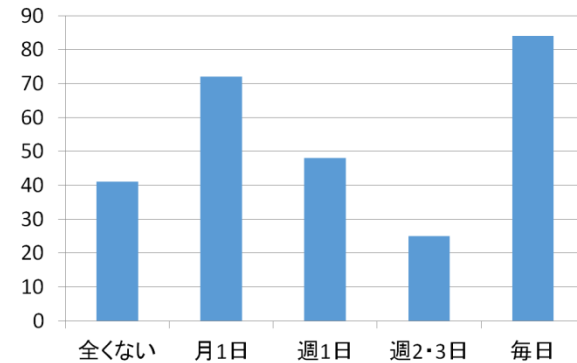
定量的調査

回答者のデモグラフィ



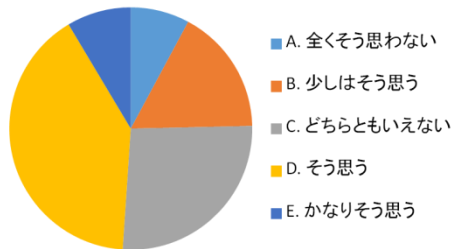
合計319名、女性は11%

Q7. データ分析の頻度

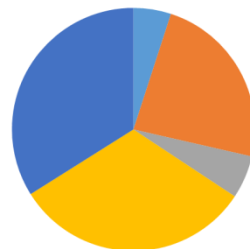


キャリアについて

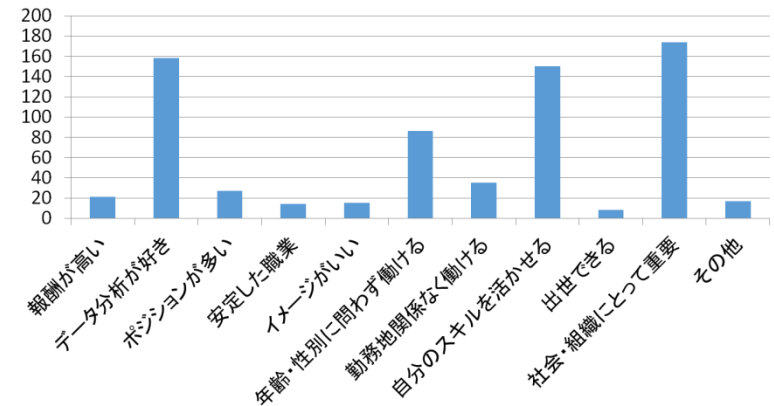
Q18. 現在、自身の能力が十分に発揮されているか



Q19. データ活用の専門家としてキャリアを積みたいか



Q20. データ活用の専門家としてキャリアを積みたい理由



クラスタリングの結果 – キャリアの類型

メーカーの製品開発・企画部門にいる中堅のIT系エンジニア。社内では確実にデータの活用が進んでいる。キャリアパスも見えている。

主に中小のサービス系の企業に勤める女性。比較的自由になる勤務形態を望んでいる。

若手で、まだ実務経験は少ないが、データサイエンティストになりたい夢を持っている。

ITサービス業でデータ分析をプロとして長年実施してきていて、この仕事に誇りを持っている。

定性的調査

定性的調査のFinding(1) : 出身は様々



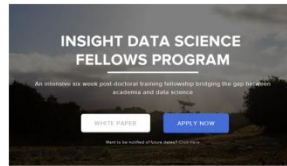
数理工学出身



MBA出身

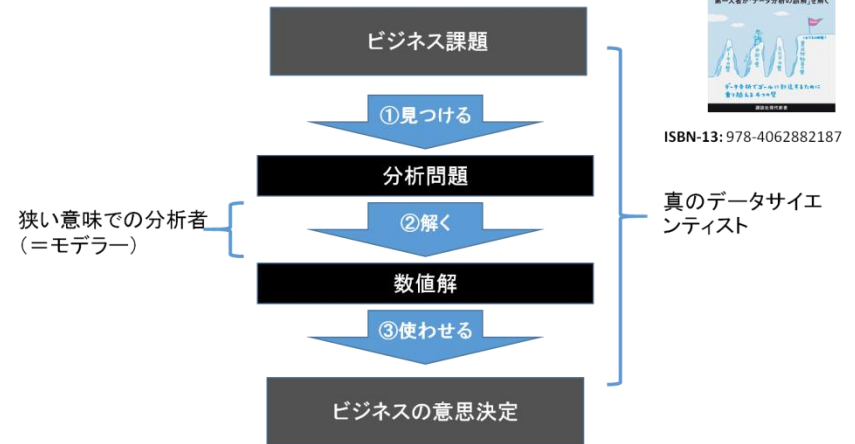


商学部出身



物理・生物など自然科学出身が多い

定性的調査のFinding(2) : データサイエンティストは全人的能力



定性的調査のFinding(3) : データ利活用は組織能力



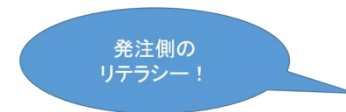
データサイエンティスト

VS



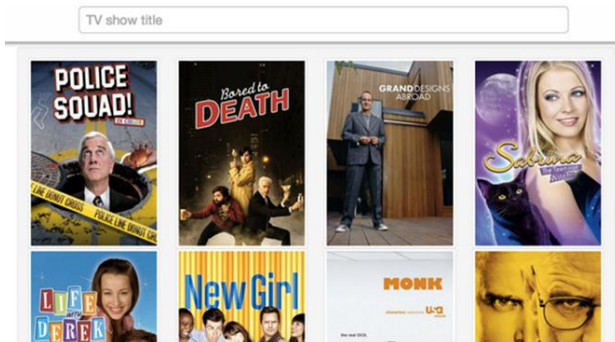
データ分析チーム

定性的調査のFinding(4) : 発注側のリテラシーも重要

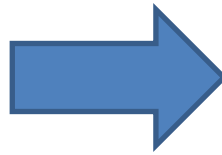


独立行政法人統計センター 戸谷理事長

彼我の違い



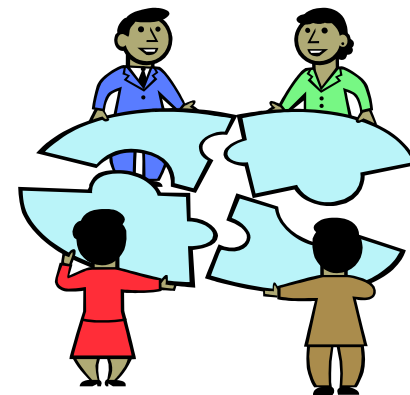
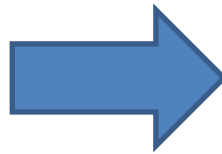
プロダクト



サービス

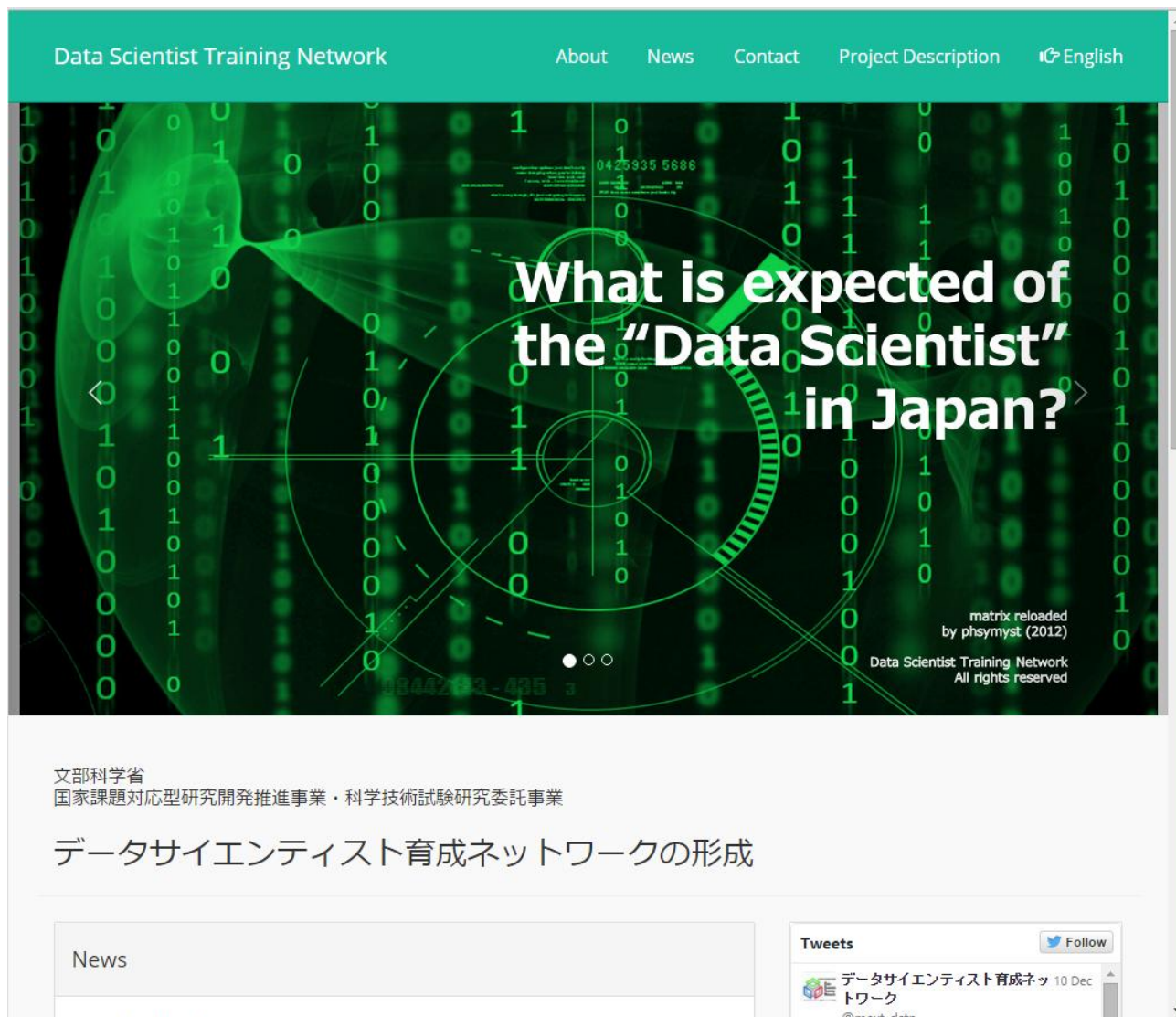


個人の能力



組織の能力

Web・Twitterによる情報発信



<http://datascientist.ism.ac.jp/index.html>



大学共同利用機関法人 情報・システム研究機構

統計数理研究所

	データサイエンティスト育成ネットワーク @mext_dstn · 12月10日 第4回データビジネス創造コンテストの募集が始まっています。今回のテーマは「データと創造力で、子育てに笑顔を」です。 dmc-lab.sfc.keio.ac.jp/digpi/ ← 1 ♡ ...
	データサイエンティスト育成ネットワーク @mext_dstn · 12月10日 「データサイエンティスト育成ネットワークの形成」サイトをリニューアルしました。 ← ♡ ...
	データサイエンティスト協会 @DataScientistPR · 12月9日 分析・データサイエンスの実務に携わる方を対象に開催する「木曜セミナー」。第4回は、12月17日（木）に開催します！↓詳細は、こちらをご覧ください。 eventdots.jp/event/576540 ← 1 ♡ ...
	データサイエンティスト協会 @DataScientistPR · 12月4日 当協会が後援する、日本で唯一のスポーツアナリティクスカンファレンス『SAJ2015-スポーツアナリティクスジャパン2015-』が、12月19日（土）日本科学未来館にて開催されます。 jsaa.org/seminar/saj201... ← ♡ 1 ...
 データサイエンティスト協会さんがリツイート 	MarkeZine（マーケジン）編集部 @markezine_jp · 11月23日 データサイエンティスト協会、データサイエンティストに必要なスキルまとめたチェックリスト公開 dlvr.it/CqJ1fQ ← 3 ♡ 15 ... 概要を表示
 データサイエンティスト協会さんがリツイート 	アナリティクス アソシエーション @a2ijp · 11月21日 データサイエンティストに必要なスキルをまとめたスキルチェックリストを初公開 a2i.jp/topics/others/... ← 3 ♡ 7 ... 概要を表示
 データサイエンティスト協会さんがリツイート 	ITmedia @itmedia · 11月21日 【マーケティング】データサイエンティストに必要な3つのスキル：データサイエンティストに必要なスキルをまとめたスキルチェックリストを初公開

H26年度の主要な成果

1. データサイエンティストの「あるべき姿」に関する提言
 - 日本学術会議 提言「ビッグデータ時代に対応する人材の育成」
-  2. インターンシップ・プログラムの展開
 - インターシップ説明会(6/15/2014)
 - 民間(株式会社アカリク)のノウハウの活用
 - 11社、学生70名
 - インターシップ参加学生からの聞き取り調査
-  3. ベスト・プラクティスの調査
 - 2つの組織(佐賀県・松竹)における追跡調査
 - クラウドソーシング利用に関する調査
4. オンライン教材の公開
 - YouTube上での公開

提言

- 日本学術会議提言「ビッグデータ時代に対応する人材の育成」
 - 日本版Insightプログラム、資格制度など
 - <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-t198-2.pdf>

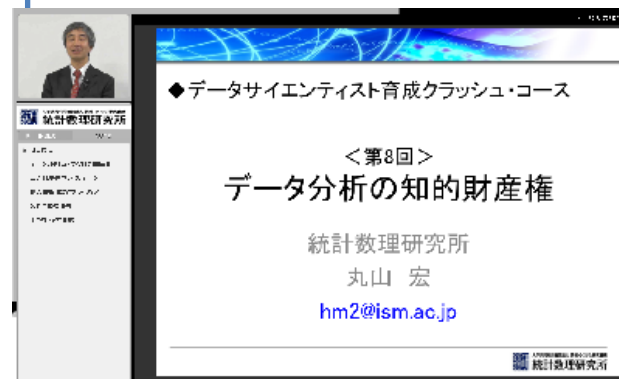


育成教材 「データサイエンティスト・クラッシュコース」

コースの全体像（約20分 × 8）

0. コース概要

1. データサイエンティストとは – ブレインパッド佐藤部長
2. データ解析基礎 – 統数研馬場特命教授
3. データ可視化とツール – 統数研中野教授
4. 統計モデリングと機械学習 – 統数研松井教授
5. 統計的時系列モデリング – 統数研川崎准教授
6. 最適化 – 統数研伊藤教授
7. データ分析と意思決定 – 統数研椿教授
8. データ分析の知的財産 – 統数研丸山教授



内外におけるDS育成プログラム

およそ200件の教育プログラムを網羅。Web上で公開

注) 各講座・教育プログラムについて、本事業が推薦、あるいは支持するものではありません。また、内容の詳細についてまで調査しておりませんので、完全ではない部分、正確ではない部分や、最新ではない部分が含まれている可能性があります。お気づきの点がございましたら、本事業事務局までお知らせください。												
プログラム (講座/教材) 名	提供元	問い合わせ先	URL	概要 (元のURLから引用)	online	実習等	回費条件	期間	料金	データサイエンス型との対応 分析サー ビス提供 範囲	学習認定 範囲	使用言語 備考
1 データサイエンティスト育成トレーニングコース	EMCジャパン株式会社	EMCジャパン株式会社EMCエデュケーション・サービス TEL: 044-520-9830 E-mail: Japan_Education@emc.com 〒151-0053 東京都渋谷区代々木2-1-1 新築 メインタワー	<a href="http://japan.emc.com/micro-
learning/data-science-big-data-
datascientist.htm">http://japan.emc.com/micro- learning/data-science-big-data- datascientist.htm	データ分析の基礎知識及びHadoopエコシステムの理解、各種統計手法について知識を用いたパズル系演習ができる。 1日目: ビッグデータ分析入門/データ分析のライフサイクル (Big Dataの概要/分析技術の現状/データサイエンティストとは/業界別のBig Data分析/データ分析のライフサイクル) 2日目: データ分析の基本〜「R」を使って (R言語の基礎知識/データの読み込みと分析/モデル構築と評価検証) 3日目: 先進的分析の構築と評価 (QやRのクラウド化/リアルタイム分析/ルール/構造化データ/ロジスティック回帰/単純ベイズ分類器 (Naive Bayesian Classifier)/決定木/経路系列分析/テキスト分析) 4日目: 先進的分析の技術とツール (非構造化データの分析 MapReduceとHadoop/Hadoopエコシステム/In-database分析-SQLの要点/In-database分析で活用するSQLとMADlib) 5日目: 分析結果の可視化とプレゼンテーション+課題チャレンジ (分析プロジェクトの企画と運用/最終成果の作りか/ビジュアライズのデモンストラ/課題チャレンジ-データ分析ライフサイクルの適用事例 (ケーススタディ))	×	○	×	5日間	?	○	○	日本語 コースの学習内容は、EMCDSA (EMC Proven™ Professional Data Science Associate) 認定の準備として利用できる。これは、EMC社の認定資格。
2 enPIT-BizApp	enPIT (公益財団「情報社会人育成のための実践型ネットワーク形成事業」分野・地域を越えた実践的情報教育国際ネットワーク)	大阪大学大学院情報科学研究科enPIT事務局 問い合わせフォーム有り	http://www.enpit.jp/fields/bizapp.html	enPIT-BizAppは、進化を続ける先端情報技術や情報インフラを積極的に活用し、潜在的なビジネスニーズや社会ニーズに対する実践的な問題解決ができる人材を育成するプログラム。中堅大学は、筑波大学、公立はこだて未来大学、産業技術大学院大学の3大学。	○	○	☆	約1年間	無料	○	○	日本語 enPITには、クラウドコンピューティング、セキュリティ、組み込みシステム、ビジネスアプリケーションの4分野がある。大学と産業界の全国的なネットワークに基づいた、実践的な情報教育を受けることができる。分野毎に中堅となる大学が推薦選定され、参加大学との単位互換等が行われている。推薦企業へのインターンシップも行われている。ただし、会場等への参加が必須。 ☆対象は、主に博士課程1年生
3 Making Sense of Data	Google	Forum有り。 https://datasense.withgoogle.com/forum	https://datasense.withgoogle.com/course	How to structure, visualize, and manipulate data. Course Content: 0. Welcome to Making Sense of Data 1. Unit 1 - Introducing the Data Process 2. Unit 2 - Answering Basic Questions with Fusion Tables 3. Unit 3 - Finding Patterns and Relationships in Data 4. Hangout on Air #1: Google Sheets 5. Hangout on Air #2: Office Hours 6. Final Project - Part 1 7. Final Project - Part 2 8. Final Assessment	○	○	×	-	Free	○		英語
4 Tackling the Challenges of Big Data	Massachusetts Institute of Technology	MIT Professional Education PHONE: +1 617-324-7693 E-mail: online@mit.edu	http://web.mit.edu/professional-education/courses/tackling-the-challenges-of-big-data.html	Participants can survey state-of-the-art topics in Big Data, looking at the data collection (smartphones, sensors, the Web), data storage and processing (scalable relational databases, Hadoop, Spark, etc.), extracting structured data from unstructured data, systems issues (exploiting multicore, security), analytics (machine learning, data compression, efficient algorithms), visualization, and a range of applications.	○	○	×	約4週間	\$545	○		英語 ディスカッションフォーラムやWebコミュニティでの議論の場に 参加できる。また、これまでに 行われたコースのアーカイブに30日 間アクセスできる。
5 ビッグデータ概説	NECマネジメントパートナー株式会社	テクノロ ジ・情報事業部 TEL: 03-5232-3075 E-mail: dbg@educ.jp.nec.com (本社) 〒211-8601 神奈川県川崎市中原区下田 第1753	https://www.neclearning.jp/courses/detail/courseid/00034/	ビッグデータを「大量の様々な形式のデータから新しい知見を導き出し、ビジネスに活用する」ための先端分野」と位置づけている。ビッグデータの概要、活用するためのデータマイニング技術、および効率的に収集・管理するためのシステム基盤技術などのトレーニングを受けることができる。 ビッグデータについて、概要、活用方法、およびビッグデータのためのシステム基盤などを1日間で分かりやすく紹介してもら える。	×	×	×	1日間	43,200円 (税込)	○	○	○ 日本語

<http://datascientist.ism.ac.jp/pdf/20141201DSMLv1.0.pdf>

インターンシッププログラムの実施

データサイエンティスト インターンシッププログラム 合同説明会

会場 ▶ 慶応義塾大学日吉キャンパス 来往舎2階大会議室

6.15 (日)

大学共同利用機関法人 情報・システム研究機構
統計数理研究所
The Institute of Statistical Mathematics

 IMJ GROUP アイ・エム・ジェイ	 IBM 日本アイ・ビー・エム	 NTT Communications NTTコム オンライン・マーケティング・ソリューション NTTコムオンライン マーケティングソリューション	 構造計画研究所 KOZO KIKAKU ENGINEERING Inc. 構造計画研究所
 opt e-marketing company オプト	 CyberAgent サイバーエージェント	 Deloitte トーマツ 有限責任監査法人トーマツ	 FUJITSU 富士通
 FreakOut フリークアウト	 BrainPad ブレインパッド	 YAHOO! JAPAN ヤフー	



企業への依頼内容

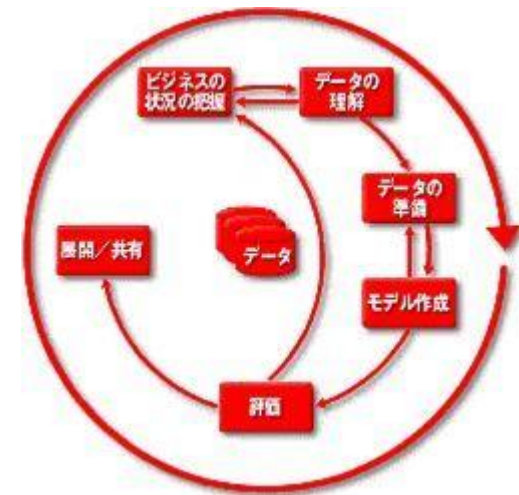
- 日程・期間：
 - 平成26年 7月～8月、2w～6w程度
- 対象：
 - 物理、情報、統計、経営などの分野でデータ分析を学んだ学生（修士、博士、学部）
- 受入機関：
 - データ分析を行う企業・研究機関
- 内容：
 - タイプⅠ：特定のデータの深い分析を行う
 - タイプⅡ：データ分析プロセスをひと通り経験する
 - タイプⅢ：データに基づく経営意思決定の現場を見聞きする

タイプ I : 深い分析

- 目的: 実際のビジネスデータを分析する経験
 - 品質データ、実験データ、マーケティングデータ、Webデータ、...
 - 一つの問題に集中して、現実のデータ分析の難しさ(ノイズ、探索的分析など)を学ばせてください
- 対象: ある程度データ分析の経験のある学生
- その他お願いしたいこと
 - ツールの利用環境 (R, SPSS, SQLなど)
 - データ生成の現場の視察

タイプⅡ：プロセス体験

- 目的：データ分析プロセスをひと通り経験
 - ビジネス分析・データ収集・分析・評価・展開
 - 例：社内でデータ分析サービスを提供する部門
- 対象：データ分析の基礎知識があり、コンサルティング等のサービスビジネスに興味がある学生
- その他お願いしたいこと
 - チームとしての活動
 - 顧客(分析依頼者)との接点



タイプⅢ：意思決定

- 目的：データ分析を意思決定につなげる現場を体験
 - マーケティング部門、事業企画部門など、データ分析の発注側の部門
 - 必ずしも自社内発注でなくとも構わない
- 対象：データに基づく合理的な経営に興味のある、マネジメント系の学生
- その他お願いしたいこと
 - 「カバン持ち」で結構です。できるだけ意思決定の場面に立ち会わせてください

説明会参加学生のデモグラフィー (n = 63)

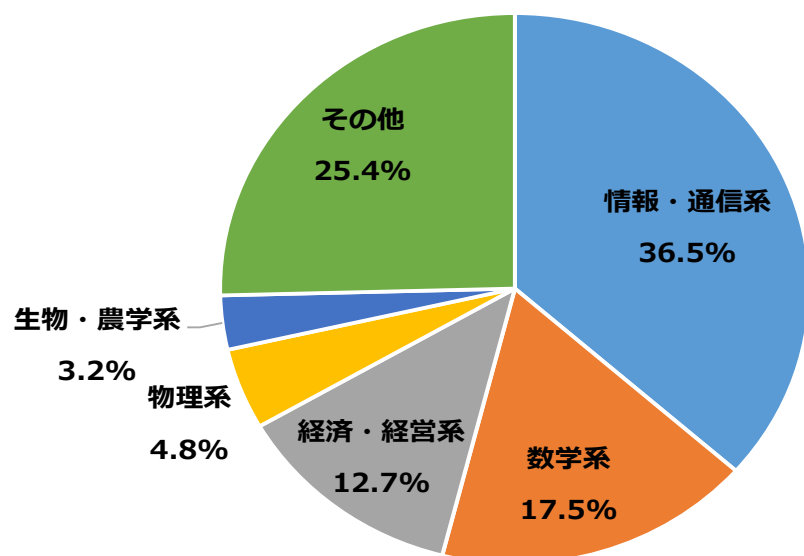


図2. 参加学生のドメイン【再掲】

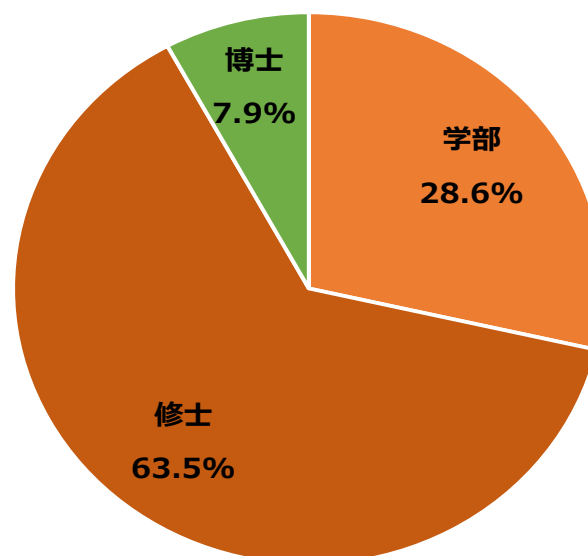


図3. 参加学生の修学状況【再掲】

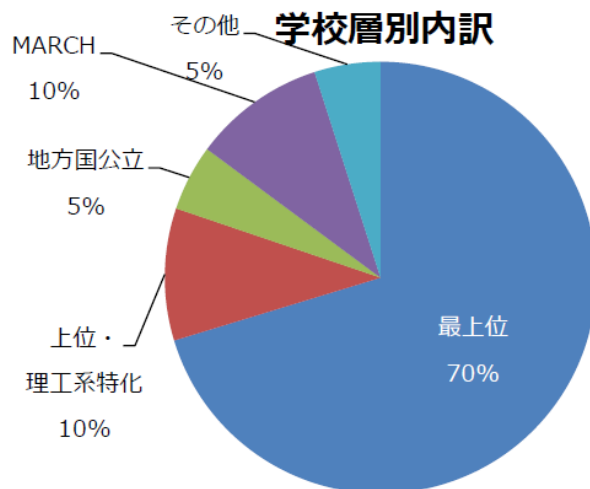
2014年6月15日（日）開催 データサイエンティストインターンシップ合同説明会 in 日吉 実績

<イベント概要>

開催日	2014年6月15日（日） 13:00～17:00
開催場所	慶應義塾大学日吉キャンパス 来往舎2階 大会議室
対象	研究や授業で何らかのデータ解析経験がある学部生～博士
当日スケジュール	13:00～13:05 統計数理研究所からの全体説明 13:05～14:00 全体プレゼン（11社×5分） 14:10～17:00 合同企業説明会

来場学生総数

63



<参加学生大学内訳>

大学院	学校名	人数	割合
最上位	慶應義塾大学大学院	15	24%
	東京大学大学院	7	11%
	東京工業大学大学院	6	10%
	一橋大学大学院	2	3%
上位・理工系特化	筑波大学大学院	3	5%
	横浜国立大学大学院	1	2%
	横浜国立大学大学院	1	2%
地方国公立	東京農工大学大学院	2	3%
	金沢大学大学院	1	2%

<参加企業一覧>

参加企業（五十音順）	
1	株式会社アイ・エム・ジェイ
2	日本アイ・ビー・エム株式会社
3	NTTコム オンライン・マーケティング・ソリューション株式会社
4	株式会社構造計画研究所
5	株式会社オプト
6	株式会社サイバーエージェント
7	有限責任監査法人トーマツ
8	富士通株式会社
9	株式会社フリークアウト
10	株式会社ブレインパッド
11	ヤフー株式会社

<参加学生の属性>

性別	人数	割合
男性	55	87%
女性	8	13%

属性	人数	割合
学部	18	29%
修士	40	63%
博士	5	8%

インターンシップ参加学生の声

- チームで働くことの難しさと重要性を学んだ
- ビッグデータを扱うときにデータの前処理が重要であることが分かった
- 時間をかけて分析しても価値ある結果が得られないことがある一方で、思いもよらないところから価値を見いだせたりしたことから、対象とするデータに価値があるかどうかを事前に見積もることの難しさを学んだ
- 大学での研究内容は、あまり役に立たなかった
- 自分に足りないこと(統計に関する知識、プログラミング・スキル、分かりやすく伝える能力)が分かった
- 自分の将来を考える良いきっかけになった
- メンターがしっかりしていて勉強になった。ただし、メンターがいるところといないところが有り、メンターがいるところの方が当然勉強になった。
- これまでに4回インターンシップに参加したが、育成体制がしっかりしているところが少ない。
- 学生に資本投資してくれるのは驚いた。



データサイエンティスト活用事例調査 – 佐賀県

- 2014年3月、佐賀県CIO森本様よりデータサイエンティスト協会にデータサイエンティスト採用に関して協力依頼
- 2014年7月、アクセンチュアとの業務委託を発表
- 8月、佐賀県を訪問してヒアリング
- その後、1-2ヶ月ごとに電話会議にてフォローアップ

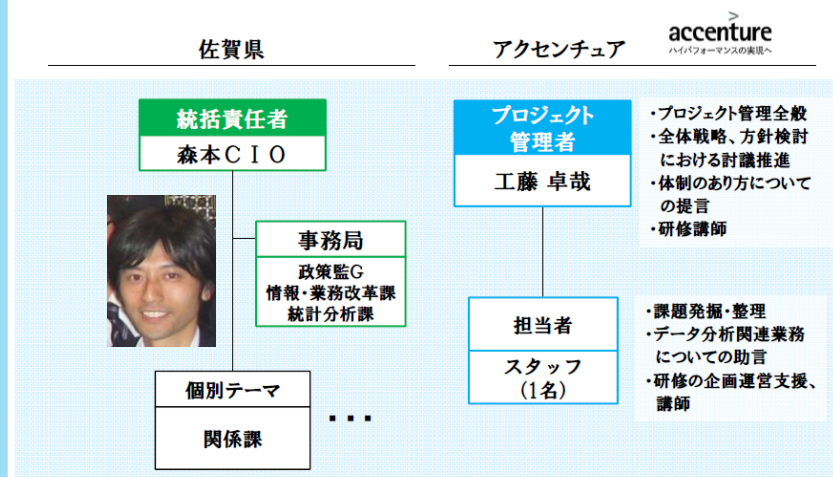


健康福祉本部 健康増進課

佐賀県
<http://www.pref.saga.lg.jp/>

佐賀県事例からの知見

体制



Copyright © 2014 Saga Prefecture. All Rights Reserved.

9

- 外部コンサルタントの利用
 - 複数ドメインでのノウハウの利用
 - 委託側のコンサル経験者
- 佐賀県庁内部での意識改革
 - データに基づく政策決定
 - データ分析が触媒となり、組織間の新しい繋がりが創出
- 研修によるスケールアップ
 - 施策決定者(課長・副課長クラス)、分析計画責任者(係長クラス)、分析実務者に分けて3回実施

フリーランスのデータサイエンティストの現状調査



クラウドソーシング
サイトで、データ分
析タスクを発注



10名の受注者

井川他、「クラウドソーシングにおけるデータサイエンティスト活用に関する
初期的調査」、第16回日本テレワーク学会研究発表大会、2014

受注者10名

- 年代
 - 20代4名、30代3名、40代1名、50代2名
- 性別
 - 男性7名、女性2名、未回答1名
- 職業
 - フリーランス2名、他は何らかの職業についている
- 全員登録時期
 - 2013年4月以降と経験は浅い
- 得点
 - 平均:16点、最大:22点、最少:10点



上位4名

- 40代女性、30代男性、20代女性、20代男性
- 全員が他の職を持っている
- 2名は今回が初の受注
- おもな統計手法は、
 - 相関分析、重回帰分析、因子分析、t検定、数量化II類
- 得点は、平均20点、19点から22点
- 発注金額:10,000円2名、15,000円2名

下位4名

- 30代男性、20代男性、50代男性、20代不明
- 学生が1名で、3名は他は職を持っている
- 1名は今回が初の受注、他は複数回受注経験あり
- おもな統計手法は、
 - 重回帰分析、集計、ソート
- 得点は、平均12.3点、10点から14点
- 発注金額:10,000円2名、15,000円2名

いかに受注者のスキルを見極めるか？

ネットワーク形成



スキルと実践を重視したビッグデータ・イノベーション人材育成プログラム

文部科学省委託事業



統計数理研究所

数学協働プログラム



公益財団法人 九州先端科学技術研究所

Institute of Systems, Information Technologies and Nanotechnologies (ISIT)



データサイエンティスト育成ネットワークの形成



統計検定

Japan Statistical Society Certificate



一般社団法人

データサイエンティスト協会



Japan Information Technology Services Industry Association

一般社団法人 情報サービス産業協会



一般社団法人

情報処理学会

Information Processing Society of Japan



「ビッグデータ利活用人材育成 ワークショップ」

2/16/2015 @ 国立情報学研究所

講演者	タイトル
丸山 宏 (統計数理研究所)	オープニング
山口 高平 (慶應義塾大学)	スキルと実践を重視したビッグデータ・イノベーション人材育成プログラム
大西 立顕 (東京大学)	文部科学省委託事業「データサイエンティスト育成ネットワークの形成」における人材育成の取り組み
美添 泰人 (青山学院大学)	統計教育大学間連携ネットワーク(JINSE)における人材育成の取組
井上 克郎 (大阪大学)	文部科学省情報技術人材育成のための実践教育ネットワーク形成事業「地域・分野を越えた実践的情報教育協働ネットワーク(enPiT)」における人材
橋本 武彦 (データサイエンティスト協会)	データサイエンティスト協会における人材育成の取り組み ～データサイエンティストの職種確立と地位向上に向けて～
橋本 大也 (データエクスチェンジ・コンソーシアム)	データエクスチェンジ・コンソーシアムにおける人材育成の取り組み
パネルディスカッション	ビッグデータ利活用人材の育成と活用に向けて
	司会: 丸山 宏(統計数理研究所)
	パネリスト:
	山口 高平 (慶應義塾大学)
	大西 立顕 (東京大学)
	美添 泰人 (青山学院大学)
	井上 克郎 (大阪大学)
	橋本 武彦 (データサイエンティスト協会)
	橋本 大也 (データエクスチェンジ・コンソーシアム)

H27年度の主要な成果

1. データサイエンティストのスキルレベル定義
 - データサイエンティスト協会のスキルレベル定義にあたり、日本学術会議 提言「ビッグデータ時代に対応する人材の育成」、その他関連情報提供
 - データサイエンティスト協会のスキルチェックリストの事前確認
 - データサイエンティスト協会のスキルチェックリスト公開 (11/13)
-  2. DS育成関連データベースの作成・公開
 - 教材271コース、コンテスト37
3. 我が国におけるDS人材育成施策に関する提言
 - 「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」報告書 (7/30)
4. 海外連携の検討
 - 英国Warwick大学訪問 (9/11)
-  5. インターンシップ・プログラムの展開
 - 説明会の民間への移管試行 (5/30)
 - 異業種交流会@東大 (1/31)
 - DSハッカソン (2/20-21)

提供されている関連 講座・教育プログラ ムのDB化

(平成27年12月 10日現在、
271講座)



「データサイエンティスト育成ネットワークの形成」

平虎 27 年 3 月

平虎 27 年 3 月



【再委託先】
東京大学大学院
情報理工学系研究科

事業期間/平成25年7月8日~平成28年3月31日

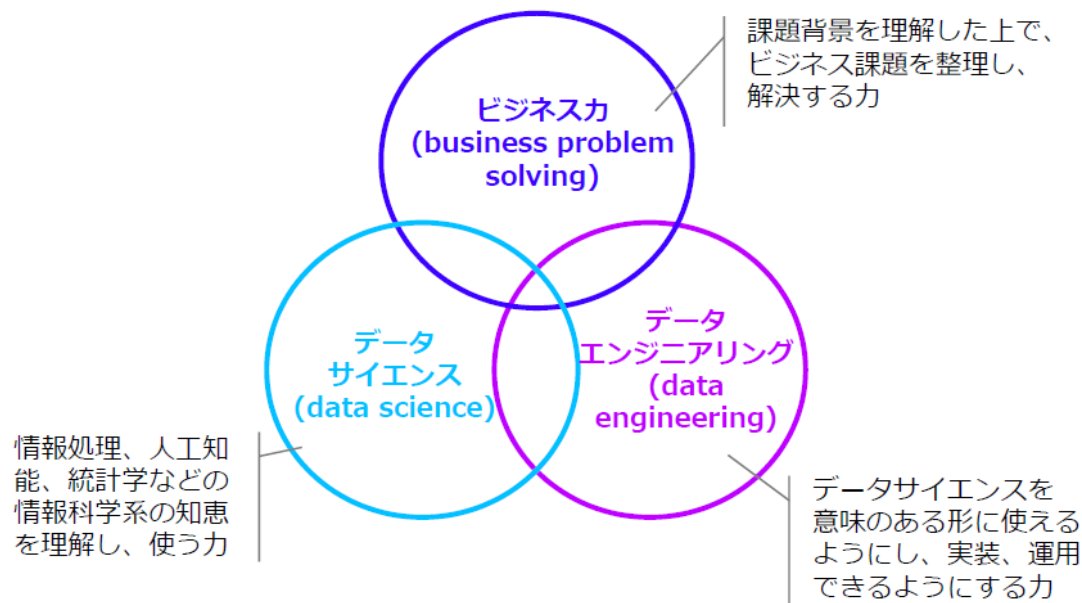
[illegible]

データサイエンス関連コンテストのDB化 (平成27年12月 10日現在、37コンテスト)

注) 各コンテストについて、本事業が推薦、あるいは支持するものではありません。また、内容の詳細についてまで調査しておりませんので、完全ではない部分、正確ではない部分や、最新ではない部分が含まれている可能性がございます。お気づきの点がございましたら、本事業事務局までお知らせください。

名称	主催団体	URL	概要	参加資格	参加費	開催頻度	問い合わせ先
1 Deep Analytics: Competition	オプト データサイエンスラボ	https://datasciencelab.jp/compelist	コンペティションを通じて、データ分析ニーズとデータサイエンティストのマッチングを目指している。	なし	なし	随時	OPT DataScienceLab 〒102-0081 東京都千代田区四番町6 東急番町ビル Email: 問い合わせ用のメールフォームを利用
2 データビジネス創造コンテスト	慶応義塾大学SFC研究所 データビジネス創造・ラボ	ビジネスパートナーになった企業がWebサイトをそれぞれ開設	学生がデータを広く深くDIGして新たな知の抽出や価値の創出を競う場を提供するとともに、才能の原石である自分自身をDIGする機会を提供することを目的とする。	高校生、高専生、大学生、大学院生（社会人経験者は除く）	なし	年2回	慶応義塾大学SFC研究所 データビジネス創造・ラボ事務局 Email: 問い合わせ用のメールフォームを利用
3 スポーツデータ解析コンペティション	日本統計学会 スポーツ統計分科会	http://estat.sci.kagoshima-u.ac.jp/sports/	野球やサッカーなど数多くのスポーツの場において、欧米諸国を中心として様々なデータ収集や高度な統計分析が行われ、科学的な見地から戦術解析や選手評価がされている。また、これらの国々では学術誌やコンgresなど学術的な場においてもこうした研究発表が盛んに行われており、統計学の専門家が積極的に関与している。 日本においては、2009年に日本統計学会スポーツ統計分科会を立ち上げるなどの活動が行われてきたが、スポーツの場における統計学の活用、および統計学専門家によるスポーツに関する研究はまだ十分であるとはいえない。そこで、実データに基づく応用研究の促進と、研究成果を現場に還元するきっかけを与えること、また研究者の裾野を広げることを目的としてコンペティションが実施されている。 コンペティションでは、野球およびサッカーの実データがデータスタジアム株式会社から貸与され、その結果を新規性や実用性など多様な観点から観る。学術的、実用的な観点から厳正に審査される。	①日本統計学会会員または同スポーツ統計分科会会員で、大学で統計学・統計科学を教える教員もしくは専攻する博士課程（博士後期課程）の大学院生を代表者とする個人またはチーム。 ②修士課程（博士前期課程）の学生や学部学生が参加する場合は必ず教員を代表者とすること。 ③チーム構成員には民間企業等に勤める者を含まないこと。 ④中学校・高校生のチームも参加可能（代表は教員）	なし	年1回	開催年によって異なる。
4 INSIGHT SIGNAL	野村総合研究所	https://www.is.nri.co.jp/contest/index.html	野村総合研究所が調査を行った消費者マーケティングデータを提供し、データ分析による斬新なビジネスの法則、マーケティング指標等を見つけ、その内容を競う。	なし。ただし、同業他社およびそれに付随するシステム開発会社、調査会社、コンサルティング会社に勤務する人のエントリーはできない場合がある。	なし	年1回	株式会社野村総合研究所 消費サービス・ヘルスケアコンサルティング部 マーケティング分析コンテスト事務局 〒105-0005 東京都千代田区丸の内1-6-5 丸の内北口ビル Email: mac2015@nri.co.jp TEL: 03-5533-2647
5 学生データサイエニティ日本決定戦 DATA LEAGUE	株式会社リクルートキャリア	https://careful.com/programDetails/?pi=K003&isc=cna0005h	リクルートグループが2014年から開催している。実データをもとに、統計学や機械学習などを活用し、予測モデルを構築するデータ分析が競われる。予測精度に力点が置かれている。	学生であること。	なし	年1回	問い合わせ用のメールフォームを利用する。
6 DATA LEAGUE～社会人から学ぶ！データ活用実践講座～データ分析	株式会社リクルートキャリア	https://careful.com/programDetails/?pi=K003&isc=cna0005h	その街の魅力を再発見し、プロモーション施策を考える実践型の「データ分析アイデアソン」。横浜 F・マリノスの活動データ×横浜市のデータ×総務省のオープンデータを素材に、各データの関係性から街の魅力を見つけ、プロモーション企画を考える。	学生であること。	なし	年1回	問い合わせ用のメールフォームを利用する。

データサイエンティスト協会によるスキル定義



Source: The Japan Data Scientist Society discussions

	ビジネス力 (business problem solving)	データサイエンス力 (data science)	データエンジニアリング力 (data engineering)
1. Senior Data Scientist 業界を代表するレベル	- 組織や市場全体にインパクトを出せる。 - 対象とする事業全体、産業領域における課題の切り分け、テーマ、論点の明確化ができる。	- 新しいアルゴリズムや分析手法の開発ができる。 - 複数のパラメータやアルゴリズムの選択など、適切な分析アプローチの設定ができる。	複数のデータソースを統合したデータシステム、もしくはデータプロダクトの構築、全体最適化ができる。
2. (full) Data Scientist 構築レベル	- 分析を通じオペレーション上の革新が実現できる。 - 仮説や可視化された問題がない中で（フレーミングさえしていなくても）、適切に問題を定義し、解き、価値を見出すことができる。 - 特定の課題領域において、課題と取組のテーマを構造的に整理し、見極めるべき論点をクリアにできる。 - 組織全体を見渡して、必要なデータの当たりをつけることができる。	- 多変量解析の概念を理解し、活用することができる。 - 機械学習、自然言語、画像処理のアルゴリズムを理解し、適切に活用、問題解決することができる。 - モデルを構築できる。	- 分析に必要なデータフォーマット、取得蓄積仕様等を設計できる（分析のためのデータシステム設計ができる） - 問題設定に応じた新規データマート設計ができる。 - 構造化データ／非構造化データを問わず、分析システムを設計できる - 構築したモデルを実装できる。 - データ分析を作ったシステムを自身で構築できる。
3. Associate Data Scientist 独り立ちレベル	- 仮説や既知の問題が与えられた中で、最適解・最大解を見出すことができる。 - 扱っている課題領域で新規の課題を切り分け、構造化できる。 - 当該プロジェクト・サービスを超えて、必要なデータの当たりをつけることができる。	- SPSS/SAS/R等が使える。 - 指示されなくてもサンプル抽出ができる（ともに内容を確認できる）。 - データクレンジング、分布、単回帰やP値の概念を理解し、活用することができる（二次元の分析はできる）。	大規模のファイルや、データベースにアクセスし、大量の構造化データを処理することができる（一般的なスプレッドシートで処理不能な規模感への対応力）。
4. Assistant Data Scientist 見習いレベル	- ビジネスにおける論理とデータの重要性を認識している。 - 仮説や既知の問題が与えられた中で、必要なデータに当たりをつけて、データを用いて改善することができる。 - 扱っている課題領域（例：配送の最適化）における基本的な課題の枠組みが理解できる。	基本統計量（平均、中央値など）の知識を有し、指示されればデータの抽出、グラフ作成を正しく行うことができる。	- 一般的なアクセス解析システムを使うことができる。 - 抽出されたデータがブツセットに対し、ExcelやAccess等の統合環境を用い、目的に応じた処理をすることができる。



大学共同利用機関法人 情報・システム研究機構

統計数理研究所

「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」報告書

現在

抜けている。スケールアウトしない原因

トップ研究教育機関が
小規模に育成

5～20人

3,400人

(2008年時。MGI report
Deep Analytical Talent)

US: 25,000人
中国: 17,000人
インド: 13,000人

世界的トップタレントの輩出

1～3人

棟梁レベル
(full) Data Scientist

500人/年

10倍以上

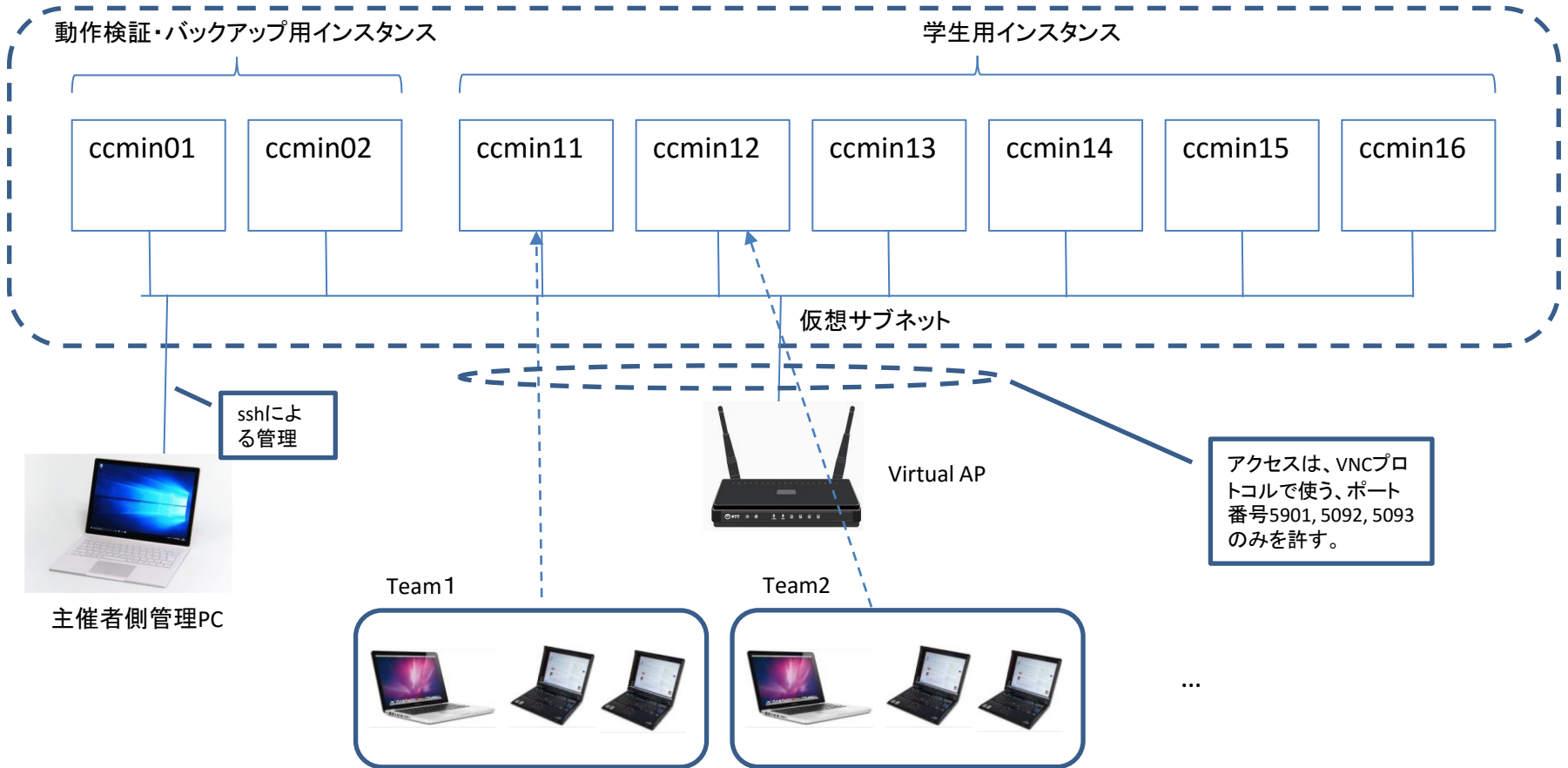
データ分析ハッカソン (2/20-21, 2016)

- 平成28年2月20日(土) 9:00 ~ 21日(日) 18:00
- 於: 統計数理研究所セミナー室1
- 対象: データ分析を学んだ学生
- 人数: 1チーム3名、6チーム
- データ: 「エンターテインメント系企業における販売データ」
(企業名は、提供元の要請により公開しない)
 - 1年分、およそ1千万レコード程度
- 課題: データを分析し、売り上げを向上させる施策を提案
- 参加チーム
 1. 東京大学: 修士・学部・学部
 2. 中央大学: 学部・学部・学部
 3. 同志社大学: 博士・学部・学部
 4. 東京大学: 修士・修士・修士
 5. 奈良先端科学技術大学: 修士・修士・修士
 6. 神奈川工科大学: 修士・修士・修士

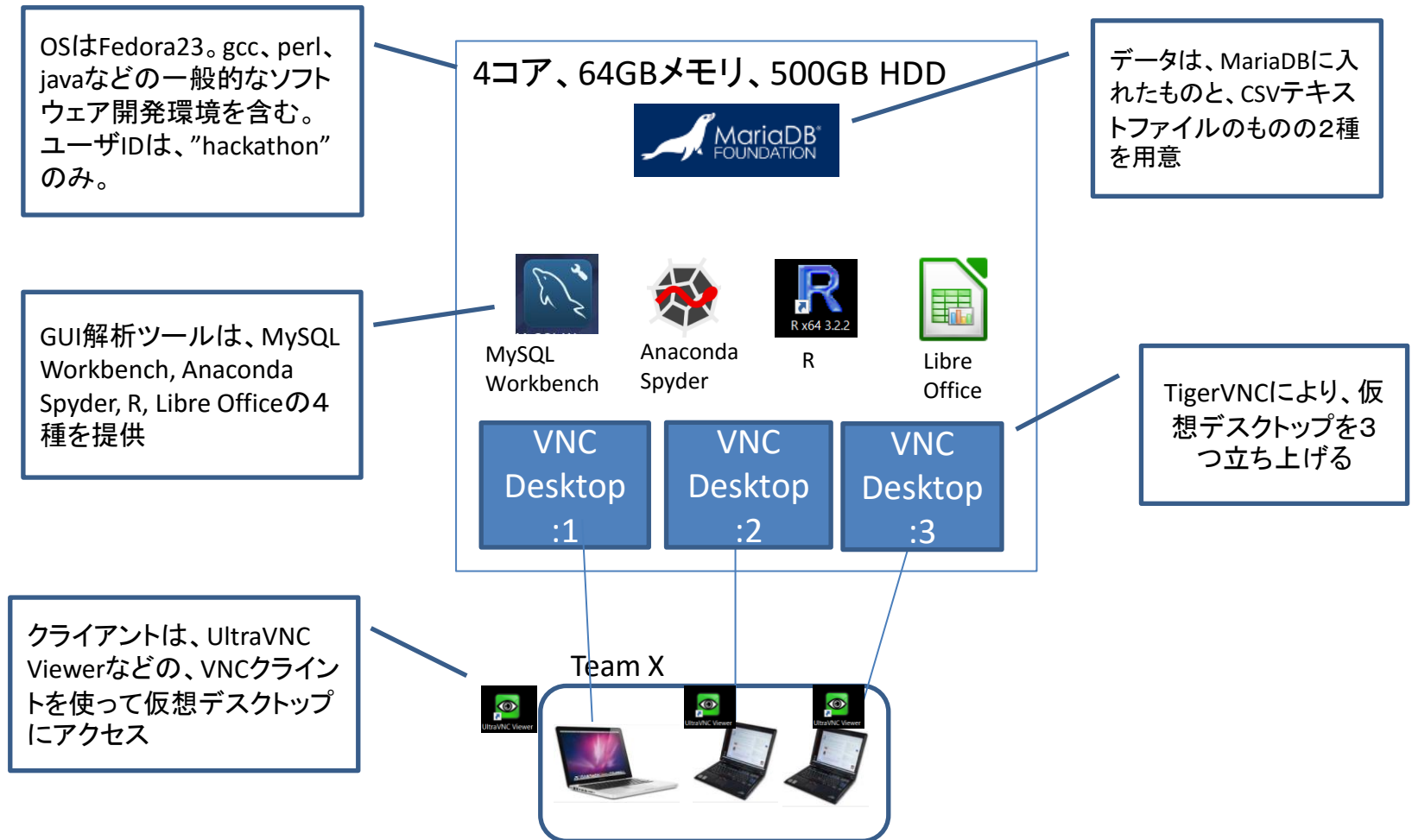




データ分析ハッカソンネットワーク構成 ～統計数理研究所共用クラウドを利用～



クラウドインスタンスの概要



スケジュール

1日目

9:00 開会挨拶

9:10 データ説明

9:40 ツール説明・ネットワーク接続

10:00 ハッカソン開始

|

22:00 サーバー停止

2日目

8:00 ハッカソン再開

|

15:00 各チームプレゼンテーション

16:30 懇親会

17:00 表彰

18:00 解散



参加学生の声

- データ分析は探索的
- 「手法ありき」ではうまくいかない
- 時間管理の重要性
- リーダーシップの重要性

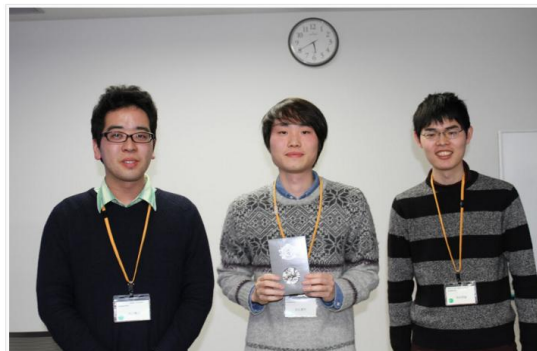
最優秀賞

同部 格明さん (同志社大学 文化情報学部)
土田 潤さん (同志社大学大学院 文化情報学研究科)
楠 和馬さん (同志社大学 文化情報学部)



優秀賞

渋谷 遼平さん (東京大学大学院 情報理工学系研究科 数理情報学専攻)
古山 貴之さん (東京大学大学院 情報理工学系研究科 数理情報学専攻)
松田 孟留さん (東京大学大学院 情報理工学系研究科 数理情報学専攻)



審査員特別賞

羽山 美優さん (中央大学 理工学部)
石原 渚さん (中央大学 理工学部)
望月 優志さん (中央大学 理工学部)



事業成果物の出口

- 現状調査
 - 日本学術会議提言に反映
 - BD利活用専門人材育成産学官懇談会における提言に反映
- 教材・コンテストDB
 - データサイエンティスト協会スキル委員会へ移管
- インターンシップ・プログラム
 - 民間(株式会社アカリク)への移管
- ハッカソン
 - 今後の統計数理研究所の人材育成事業へ

Thank you

