

文系における 文理融合型データサイエンス教育 真のデータ活用人材の育成に向けて

Rissho Data Science

初等中等教育改革
 小中:「データの活用」
 高校:共通必修
 「データの分析」数学
 「データの活用」情報Ⅰ
 「データサイエンス」情報Ⅱ

入試改革
 「記述・推測統計」
 数学Ⅰ・A・B
情報Ⅰ入試
 大学等高等教育改革
 社会人リスキリング

2022年創立150周年
 2021年4月DS学部開設
 定員:240名
 ◆学位:学士(データサイエンス学)

立正大学データサイエンス学部
 (公益社団法人私立大学情報教育協会
 情報教育研究委員会データサイエンス教育分科会主査)
 渡辺 美智子

9学部16学科7研究科
 経済学部
 経営学部
 文学部
 心理学部
 法学部
 仏教学部
 社会福祉学部
 地球環境科学部
データサイエンス学部

自動運転・商品の自動レコメンド・AIによる自動診断・不動産テック・スリープテック・牛の発情期検出予測・AI美空ひばり・スマートトイレ・野球サッカー等入場者数の予測に基づく価格運動性・・・
全てがデータでつながるSociety 5.0の社会/驚異的な入試倍率となっているデータサイエンス系学部
 MUSYC(アエラ2019.05):データサイエンス・AIの教育を全大学・全学部で基礎～応用基礎(発展)実施
 (AI戦略2019による教育改革)/U.S.News and World Report2021 Statistician(全6位、ビジネス2位),Data Scientist(全8位、技術職2位),Forbs: Best Job in America2019 Data Scientist(1位)、Data Analyst(31位)

内容

- 立正大学データサイエンス学部
文理融合カリキュラムと政策的な背景
 - 文系学生にとってのデータサイエンス教育
 - データサイエンス専門職能の拡大
 : データビジネスリーディングマネージャー
(Analytics Translators)
 (データ活用・分析の価値・コスト・リスクを組織で管理・展開し
 ビジネス価値につなぐ職能)
- 2026年までに、米国だけで
 200万人から400万人の需要
 (McKinsey Global Institute)

文系も、 理系も、

データサイエンス学部。

問題を解決に導き
 社会をデザインする...
 文理ではなく
 「ヒト」の話

科学技術が先行する
 社会、何のための、
 誰のための..を
 考える
 「ヒト」

ビジネスに
 データサイエンスを
 結びつける
 「ヒト」が必要



技術・数理の側面のみ

...理系

人間中心の社会デザイン

...文系こそ活躍
できる

効率化も問題解決も、
 そして課題発見も
 「ヒト」がいるからこそ

「モラリスト×エキスパート」

to tackle problems that really matter

**Data for Social
GOOD!**

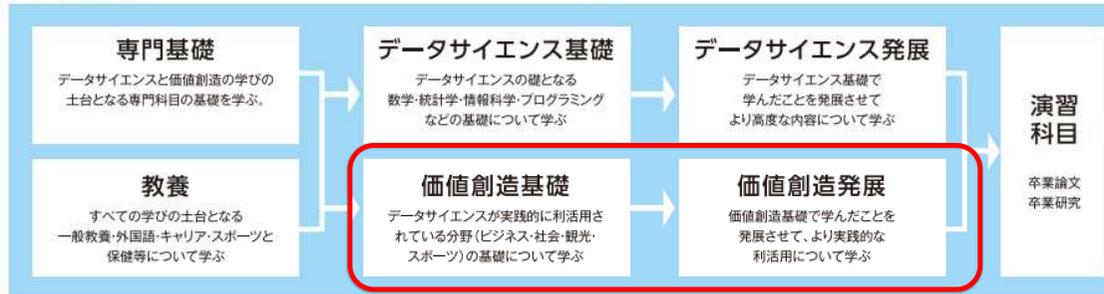
教育の特色

文理融合型の学びを通じて、実社会で活躍できるデータサイエンティストをめざす!

- 数理・統計、AI及びプログラミングを基礎からじっくりと学習
- データサイエンスの応用(ビジネス・社会・観光、スポーツ)に関する授業が充実
- インターンシップやフィールドワークなど、実社会におけるデータサイエンスの活用を意識した学び
- 学生全員に1人1台ノートパソコンを貸与し、プログラミング学習やオンライン授業などをサポート

AI・DSガバナンス
 + 倫理・法律

学びの流れ



世界時価総額ランキング2023.01

No.	前月比	会社名	時価総額	国
01	01 →	アップル Apple	2,332,313	アメリカ
02	02 →	サウジアラムコ Saudi Arabian Oil	1,862,633	サウジアラビア
03	03 →	マイクロソフト Microsoft	1,856,633	アメリカ
04	04 →	アルファベット (クラスA/クラスC) Alphabet		アメリカ
05	05 →	アマゾン・ドット・コム Amazon.com		アメリカ
06	06 →	バークシャー・ハサウェイ (クラスA/クラスB) Berkshire Hathaway		アメリカ
07	07 →	テスラ Tesla		アメリカ
..	..	シスコシステムズ Cisco Systems		アメリカ
48	↑	マクドナルド McDonald's		アメリカ
49	↑	エルメス・インターナショナル Hermes International		フランス
50	↑	リライアンス・インダストリーズ Reliance Industries		インド

平成元年
Top5を含め
50位中に
日本は32社

2018年6月総理大臣声明
第4次産業革命に突入
人材育成改革の政府方針

デジタル革命が急速に進展する中で、
価値を生み出すデータや人材をめぐる熾烈な争奪戦が
世界で繰り広げられている。

日本はこのまま手をこまねいてはならない。
Society5.0に向かって、我が国こそが、世界をリードし
ていかなければならない。この数年が勝負。

本年を**第4次産業革命元年**とし、生産性革命の実現を
あらゆる分野で推進(自動運転、ヘルスケア、デジタルガ
バメントなど)加えて、
こうした**社会変革の実現の基盤**となる、大胆な規制改革
に挑戦するとともに、**AI人材の育成を始めとした教育シ
ステムの改革、大学改革などイノベーションを生み出す
エコシステムづくりを進める。**

内閣府、文部科学省、経済産業省の認定
数理・データサイエンス・AI教育プログラム

変革の
後押し!

文部科学省「数理・データサイエンス・AI教育プログラム
(リテラシーレベル)」に認定されました!

2021年8月4日付けで...認定されました。(有効期限:令和8年3月31日) この認定制度は、令和3年度から開始されたもので、**数理・データサイエンス・AIに関する知識及び技術について体系的な教育を行うものを文部科学大臣が認定及び選定して奨励すること。**認定ロゴ(MDASH: Approved for Mathematics, Data science and AI Smart Higher Education)。

学校種別	区分	認定数(累計)	うち選定数(累計)
大学	国立	30	6
	公立	3	1
	私立	33	3
	小計	66	10
短期大学	公立	0	0
	私立	2	0
	小計	2	0
高等専門学校	国立	9	1
	公立	1	0
	私立	0	0
	小計	10	1
合計		78	11



- 尚美学園大学: データ・情報・メディア総合教育プログラム
- 亜細亜大学: データサイエンス副専攻
- 嘉悦大学: 嘉悦大学ICT・データサイエンスプログラム
- 工学院大学: 工学者のための数理・データサイエンス・AI教育プログラム
- 上智大学: データサイエンス専修
- 成城大学: データサイエンス基礎力養成・認定プログラム
- 玉川大学: 数理・データサイエンス・AI教育プログラム
- 東京都立大学: 数理・データサイエンス基礎教育プログラム
- 日本女子大学: 数理・AI・データサイエンスに関する教育プログラム
- 武蔵野大学: データサイエンス・AI入門
- 早稲田大学: GEC, CDSデータ科学教育プログラム
- 神奈川工科大学: データサイエンス・AIリテラシー教育プログラム 他 計35校

経済産業省
Ministry of Economy, Trade and Industry

申請・お問合せ English サイトマップ 本
ニュースリリース 会見・動静・談話



数理・データサイエンス・AI教育プログラム支援サイト

認定プログラムと就職の橋渡し: 企業と学生のマッチング・学修意欲喚起

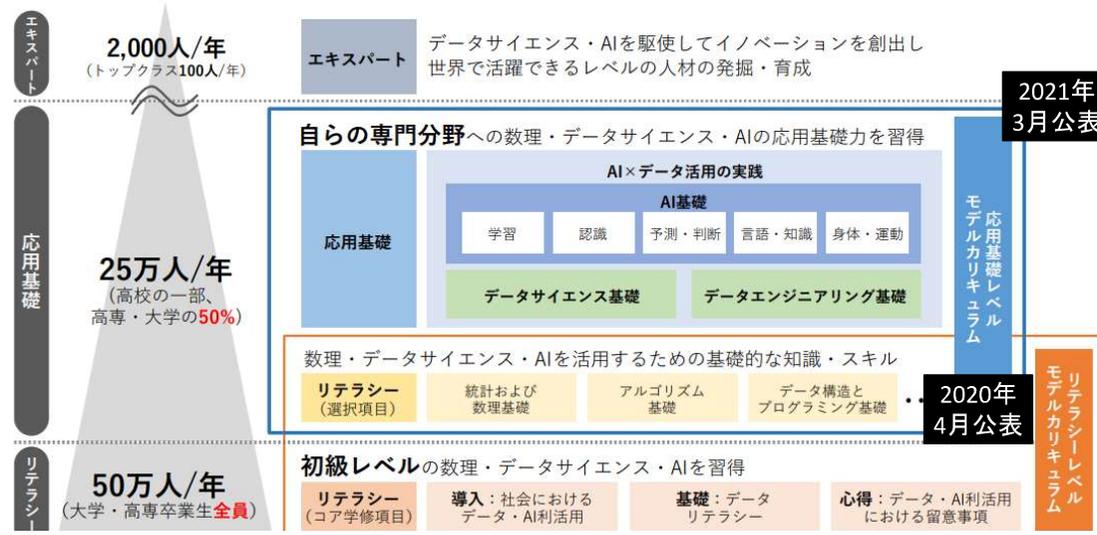
<MDASH SUPPORTERS>

本サイトは、「数理・データサイエンス・AI教育プログラム」を支援いただける企業等のサポーター
“MDASH SUPPORTERS”を掲載します。



数理・データサイエンス教育強化拠点コンソーシアム
大学教育の改革
モデルカリキュラムと認定

数理・データサイエンス・AI (応用基礎レベル) の位置づけ



- 背景

政府の「AI戦略2019」(2019年6月策定)にて、リテラシー教育として、文理を問わず、全ての大学・高専生(約50万人卒/年)が、課程にて初級レベルの数理・データサイエンス・AIを習得する、とされたことを踏まえ、各大学・高専にて参照可能な「モデルカリキュラム」を数理・データサイエンス教育強化拠点コンソーシアムにおいて検討・策定。
- 学修目標・カリキュラム実施にあたっての基本的考え方

今後のデジタル社会において、数理・データサイエンス・AIを日常生活、仕事等の場で使いこなすことができる基礎的素養を主体的に身に付けること。そして、学修した数理・データサイエンス・AIに関する知識・技能をもとに、これらを活用する際は、**人間中心の適切な判断**ができ、**不安なく自らの意志でAI等の恩恵を享受し、これらを説明し、活用**できるようにすること。
- 1. 数理・データサイエンス・AIを活用することの「楽しさ」や「学ぶことの意義」を重点的に教え、学生に好奇心や関心を高く持ってもらい魅力的かつ特色ある教育を行う。数理・データサイエンス・AIを活用することが「好き」な人材を育成し、それが自分・他人を含めて、次の学修への意欲、動機付けにつながるような「**学びの相乗効果**」を生み出すことを狙う。
- 2. 各大学・高専においてカリキュラムを実施するにあたっては、各大学・高専の教育目的、分野の特性、個々の学生の学習履歴や習熟度合い等に応じて、本モデルカリキュラムのなかから適切かつ柔軟に**選択・抽出し、有機性を考慮した教育**を行う。
- 3. **実データ、実課題を用いた演習**など、**社会での実例を題材**に数理・データサイエンス・AIを活用することを通じ、現実の課題と適切な活用方法を学ぶことをカリキュラムに取り入れる。
- 4. リテラシーレベルの教育では「**分かりやすさ**」を重視した教育を実施する。

- 背景

政府の「AI戦略2019」(2019年6月策定)にて、リテラシー教育として、文理を問わず、全ての大学・高専生(約50万人卒/年)が、課程にて初級レベルの数理・データサイエンス・AIを習得する、とされたことを踏まえ、各大学・高専にて参照可能な「モデルカリキュラム」を数理・データサイエンス教育強化拠点コンソーシアムにおいて検討・策定。
- 学修目標・カリキュラム実施にあたっての基本的考え方

今後のデジタル社会において、数理・データサイエンス・AIを日常生活、仕事等の場で使いこなすことができる基礎的素養を主体的に身に付けること。そして、学修した数理・データサイエンス・AIに関する知識・技能をもとに、これらを活用する際は、**人間中心の適切な判断**ができ、**不安なく自らの意志でAI等の恩恵を享受し、これらを説明し、活用**できるようにすること。
- 1. 数理・データサイエンス・AIを活用することの「楽しさ」や「学ぶことの意義」を重点的に教え、学生に好奇心や関心を高く持ってもらい魅力的かつ特色ある教育を行う。数理・データサイエンス・AIを活用することが「好き」な人材を育成し、それが自分・他人を含めて、次の学修への意欲、動機付けにつながるような「**学びの相乗効果**」を生み出すことを狙う。
- 2. 各大学・高専においてカリキュラムを実施するにあたっては、各大学・高専の教育目的、分野の特性、個々の学生の学習履歴や習熟度合い等に応じて、本モデルカリキュラムのなかから適切かつ柔軟に**選択・抽出し、有機性を考慮した教育**を行う。
- 3. **実データ、実課題を用いた演習**など、**社会での実例を題材**に数理・データサイエンス・AIを活用することを通じ、現実の課題と適切な活用方法を学ぶことをカリキュラムに取り入れる。
- 4. リテラシーレベルの教育では「**分かりやすさ**」を重視した教育を実施する。

1. 社会におけるデータ・AI利活用<スキルセット>

導入	1. 社会におけるデータ・AI利活用	1. 社会におけるデータ・AI利活用	キーワード (知識・スキル)
導入	1-1. 社会で起きている変化	1-1. 社会で起きている変化	ビッグデータ、IoT、AI、ロボット
基礎	2. データリテラシー	1-2. 社会で活用されているデータ	データ量の増加、計算機の処理性能の向上、AIの非連続的進化
心得	3. データ・AI利活用における留意事項	1-3. データ・AIの活用領域	第4次産業革命、Society 5.0、データ駆動型社会
選択	4. オプション	1-4. データ・AI利活用のための技術	複数技術を組み合わせたAIサービス

データサイエンスサイクル

数理・データサイエンス教育強化拠点コンソーシアム

導入	1. 社会におけるデータ・AI利活用	1-2. 社会で活用されているデータ	1-3. データ・AIの活用領域	1-4. データ・AI利活用のための技術	1-5. データ・AI利活用の現場	1-6. データ・AI利活用の最新動向
導入	1-1. 社会で起きている変化	1-2. 社会で活用されているデータ	1-3. データ・AIの活用領域	1-4. データ・AI利活用のための技術	1-5. データ・AI利活用の現場	1-6. データ・AI利活用の最新動向
基礎	2. データリテラシー	2-1. データを読む	2-2. データを扱う	2-3. データを扱う		
心得	3. データ・AI利活用における留意事項	3-1. データ・AIを扱う上での留意事項	3-2. データを守る上での留意事項			
選択	4. オプション	4-1. 統計および数理基礎	4-2. アルゴリズム基礎	4-3. データ構造とプログラミング基礎	4-4. 時系列データ解析	4-5. テキスト解析

<数理・データサイエンス・AI教育(応用基礎レベル)の学修目標>

数理・データサイエンス・AI教育(リテラシーレベル)の教育を補完的・発展的に学び、**データから意味を抽出し、現場にフィードバックする能力、AIを活用し課題解決につなげる基礎能力**を修得すること。そして、**自らの専門分野に数理・データサイエンス・AIを応用するための大局的な視点を獲得**すること。

数理・データサイエンス・AI 応用基礎レベルの学修内容と教育方法

- 応用基礎レベルの学修内容**
- データサイエンスおよびデータエンジニアリングの基本的な概念と手法、応用例を学ぶことで、**データから意味を抽出し、現場にフィードバックするための方法を理解**する
 - AIの基本的な概念と手法、応用例を学ぶことで、**AI技術を活用し課題解決につなげるとは何かを理解**する
- 応用基礎レベルの推奨される教育方法**
- データサイエンス、データエンジニアリング、AIを学ぶ過程において、**演習**や**課題解決型学習(PBL: Project Based Learning)**等を効果的に組み入れることにより、**実践的スキルの習得**を目指す



数理・データサイエンス・AI(リテラシーレベル)講座(全5講座)

放送大学 BS231
モデルカリキュラムに即した
キャリアアップ講座
インターネット配信講座
の充実

導入A データサイエンス基礎から応用

講座概要: 基礎技術として数理・統計に関する技術情報を提供し、応用領域として様々な事例を解説し...

出演講師: 渡辺 美智子 教授 (立正大学), 長谷山 美紀 教授 (北海道大学), 南和法 教授 (統計科学研究所), 北川 由紀恵 教授 (放送大学), 吉田 健 (筑波)

基礎B 数理・データサイエンス・AI リテラシー講座 基礎理論、社会での応用、Excel等を活用したデータ処理について分かりやすさ

出演講師: 渡辺 美智子 教授 (立正大学), 小野 陽子 准教授 (横浜国立大学), 大橋 流太郎 講師 (文政大学), 竹内 光悦 教授 (実践女子大学), 梅澤 友樹 講師 (岐阜聖徳学園大学)

導入B 数理・データサイエンス・AI リテラシー

講座概要: 社会におけるAI・データ活用の最新動向につ

出演講師: 竹村 彰通 教授, 齋藤 邦彦 教授, 前田 薫 教授

基礎A デジタル社会のデータリテラシー

講座概要: データ思考を幅広くデータリテラシーの内容を、身の回りの社会の実例に沿って分かり易く解説します。

出演講師: 渡辺 美智子 教授 (立正大学), 小野 陽子 准教授 (横浜国立大学), 大橋 流太郎 講師 (文政大学), 竹内 光悦 教授 (実践女子大学), 梅澤 友樹 講師 (岐阜聖徳学園大学)

自動運転・商品の自動レコメンド・AIによる自動診断・不動産テック・スリープテック・牛の発情期検出予測・AI美空ひばり・スマートトイレ・野球サッカー等入場者予測による価格連動性・
全てがデータでつながるSociety 5.0の社会／驚異的な入試倍率となっているデータサイエンス系学部
 MUSYC(アエラ2019.05)：データサイエンス・AIの教育を全大学・全学部で基礎～応用基礎(発展)実施
 (AI戦略2019による教育改革)／U.S.News and World Report2021 Statistician(全6位、ビジネス2位), Data Scientist(全8位、技術職2位), Forbs: Best Job in America2019 Data Scientist(1位)、Data Analyst(31位)

内容

- 立正大学データサイエンス学部
文理融合カリキュラムと政策的な背景
- 文系学生にとってのデータサイエンス教育
- データサイエンス専門職能の拡大
：データビジネスリーディングマネージャー
(Analytics Translators)
(データ活用・分析の価値・コスト・リスクを組織で管理・展開し
ビジネス価値につなぐ職能)

2026年までに、米国だけで
200万人から400万人の需要
(McKinsey Global Institute)

公益社団法人私立大学情報教育協会情報教育研究委員会
データサイエンス教育分科会主催
データサイエンス・AI授業実践研究ワークショップ
 多くの学系と多様なレベルの学生に対して、数理・データサイエンス・AIの全学必修化教育モデルカリキュラムの大枠を踏まえ、関連科目担当者間で教育プログラムの内容や教材開発等の具体的な授業設計の検討が必要。リテラシーレベルの授業支援を目的とし、教員間で意見交流を重ね、**データサイエンス・AIを体感する授業構成**、学生に**興味・関心を持たせる授業学修内容と工夫**などの観点からの理解促進を目指す教員コミュニティの形成

- 第1回：令和4年6月30日(木) 17:00～18:30 (Zoom会議, 参加無料)
 講演「**文系学生に配慮したデータサイエンス教育を考える**」
 辻智氏(大阪公立大学研究推進機構特任教授、成城大学非常勤講師、元IBM)
- 文系学生がデータサイエンスを学ぶ意義と動機
人文・社会科学系の学生にとって、データサイエンスを学ぶと何がよいのか?
 - 各大学におけるデータサイエンス教育に特色を出そう
平均的な授業ではなく、各大学の理念や将来ビジョンに立った特色ある授業群を構成
 - 学生に興味・関心を持たせ主体的な学習に誘う授業の工夫は?

データサイエンス×STEAM×アクティブラーニング Science, Technology, Engineering, Art and Mathematics

クオリティマネジメント・サービスデータサイエンス・多変量因果解析・デジタル社会のデータリテラシー・アスリートのためのデータサイエンス

『VIEW21』高校版 2019年度 2月号
【特集】「学校教育デザイン」を描く今と未来

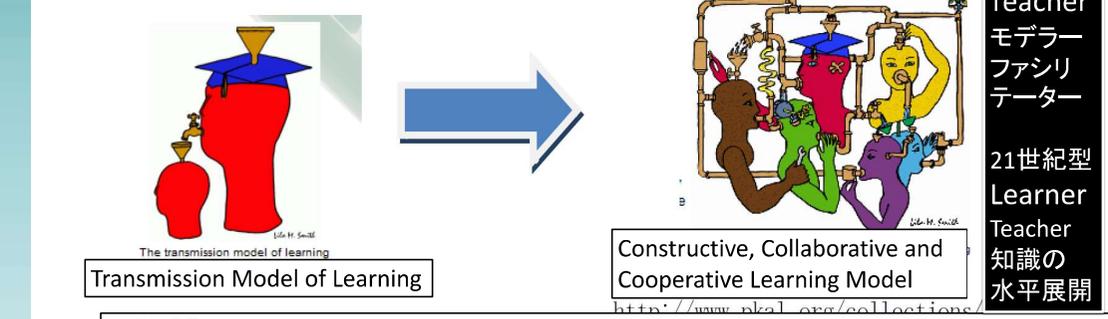
価値創造への情熱をデータとともに他者へ伝え、協働して未来を創る
大学院のSTEM/STEAM教育

分析手法を活用してサービスの価値を創造し、社会実装するPBL型授業
①聞いたことはすぐ忘れる
②見せてもらったことは覚えるかもしれない
③やってみたことは忘れられない

想いをデータで語り合い、人がつながる経験を生徒に積ませたい!

可視化(ダッシュボード) 記述統計・推測統計 重回帰分析(教師あり) 主成分分析(教師なし)・主成分回帰分析・選択ベクトル・コンジョイント分析 構造方程式モデリング 潜在クラス・トピック分析

協働的問題解決力(コンピテンシー)は どうやって育成するのか?



21世紀型FD; Project Kaleidoscope (PKAL) is AAC&U's STEM higher education reform center

Tell me and I will forget; Show me and I may remember; Involve me and I will understand.

- グローバル社会では **Adult Learning** を通して、世界共通の力量である21世紀型ワークスキル が身に着くと認識
- ① 思考力(創造性と革新性, 批判的思考・問題解決・意思決定, 学習能力等)
 - ② チームで働く力(コミュニケーション, コラボレーション(チームワーク))

21世紀型 Teacher モデレーター

21世紀型 Learner Teacher 知識の水平展開

大学学部レベルカリキュラムと教育方法の変革

米国統計協会2014

育成すべきコアコンピテンシー

- データに基づく思考力
Think with Data
- 現実の課題を統計的な仮説におとし、統計的に仮説の検証ができる
Pose and Answer
Statistical Questions

ポイント

- ① データサイエンスの要素の拡大
Rなどの統計ソフトスキル、データハンドリングスキル、データベースやプログラムスキル、問題解決スキル
- ② 現実の複雑なデータを扱う経験、データ取得デザイン
- ③ 予測モデル・要因分析
デザイン・交絡・バイアス
- ④ 統計コンサルティング・ビジュアライゼーション

21

米国(大学院)

* データサイエンスの需要は大きい。 * 多くの企業が求めるのは、知識ではなく活用経験・実績 * MOOC等オンライン授業の提供のみでは用を足さない

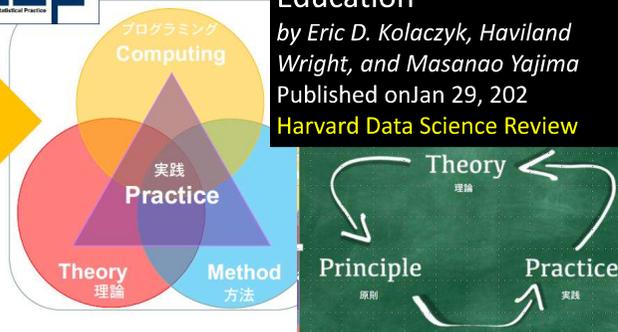
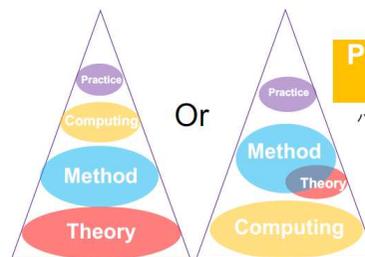
この5年間で、多くのデータサイエンス教育プログラムが溢れている

(500大学、830プログラム... 応用統計、データサイエンス、ビジネスアナリティクス)

経験・実績不足への対応に関心が移行

Data Science (DS) Education Program Design

Traditional DS program



Statistics Practicum: Placing 'Practice' at the Center of Data Science Education
by Eric D. Kolaczyk, Haviland Wright, and Masanao Yajima
Published on Jan 29, 202
Harvard Data Science Review

日本のAI戦略 VS. 中国

AI戦略【基本的考え方】

- > 「人間尊重」、「多様性」、「持続可能」の3つの理念を掲げ、Society 5.0を実現し、SDGsに貢献
- > 3つの理念を実装する、4つの戦略目標（人材、産業競争力、技術体系、国際）を設定
- > 目標の達成に向けて、「未来への基盤作り」、「産業・社会の基盤作り」、「倫理」に関する特定

戦略目標Ⅰ：人材
人口比において最もAI時代に対応した人材を育成・吸引する国となり、持続的に実現する仕組みを構築

理念（実現する社会）

- 人間の尊厳の尊重 (Dignity)
- 多様な人々が多様な幸せを追求 (Diversity)
- 持続可能 (Sustainability)

戦略目標Ⅲ：技術体系
理念を実現するための一連の技術体系を確立し、運用するための仕組みを実現

AI上の魔法の動物たち
科学的探究
実態(対象)を観察・特徴量をまとめ、機能に落とし込む

ネコの光る眼
しゃべる象
キリンの伸びる首

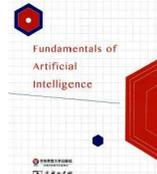
全国首套人工智能教材明年进入中小学

全国首套涵盖了从小学到高中的人工智能教材在沪亮相。



人工智能基础 (高中版)

人工智能基础 (高中版)



科学的探究能力の育成重視の転換点

・アメリカの科学教育スタンダード (NSES, 1996)

物理, 地学, 生命科学, 科学技術の社会への応用等の領域に加えて、科学的探究 (Scientific Inquiry) が独立の領域として設置。強調されたのは、「既存の科学的知識の習得を薄くし、科学の本性(の理解と探究能力の開発)により重点を置く」

- ・身近な諸現象に対する科学的な接し方、記述や説明などの科学的表現力、データや考えをまとめるための統計的処理の概念やスキル
 - ・教育の早期段階から学年を追って体系的に教育する枠組みを具体的に提示学校までの毎学年で統計的内容が体系的に示されている
- 数学教育との関連: イギリスのMath. Sciences Edu. Board(1990), アメリカの数学教師協議会(1989, 2000), OECDの生徒の学習到達度調査(PISA, 2003)などで、統計と確率、不確実性の数理の領域の相当重要な位置際の文脈に沿って活用する力、いわゆる統計的思考力の育成により重点を置くことが世界の主流

011 年全米共通コアカリキュラム数学:統計内容拡充、2012年、2013年全米学術研究協議会 (NRC) 次世代科学教育スタンダード (NGSS) ビッグデータを意識した科学的探究、データサイエンス教育の体系化・データサイエンティスト・プロフェッショナル育成の制度化

- ① 身の回りの現象を注意深く観察する。
- ② 現象に対する疑問や問題を見つけ、明確にする。
- ③ 問題点を具体化する。
- ④ 関係する既存の知識を洗い出す。
- ⑤ 仮説を明確にする。
- ⑥ 仮説検証のための観察・調査・実験等の研究を行う。
- ⑦ 結果を解釈する。
- ⑧ 結果を研究成果に反映させる。
- ⑨ 成果を社会と組織に向けて公表し議論する。



図1 科学的探究のプロセス

このような科学的探究プロセスの理解と探究する態度の育成は、学校教育に

「仮説: 科学的に検証が可能な仮説、また、「データ間の因果をつなぐ論理」または「現象間の連関を予想」とあるように、因果に関する仮説と(ターゲット指標の変動を説明するルール、パターン)

Data Investigation Process データに基づく探究プロセス InSTEP

Holistic, Dynamic and Nonlinear

Frame Problem

- Consider real-world phenomena & broader issues related to problem.
- Pose investigative question(s).
- Anticipate potential data and strategies.

Communicate & Propose Action

- Craft a data story to convey insight to stakeholder audiences.
- Justify claims with evidence from data and propose possible action.
- Address uncertainty, constraints, and potential bias in the analysis.

Consider Models

- Analyze and identify models that address the problem.
- Consider assumptions and purpose of the models.
- Recognize possible limitations.

Explore & Visualize Data

- Construct meaningful visualizations, static or dynamic.
- Compute meaningful statistical measures.
- Explore and analyze data for potential relationships or patterns that address the problem.

Consider & Gather Data

- Understand possible attributes, measurements, and data collection methods needed for the problem.
- Evaluate and use appropriate design and techniques to collect or source data.
- Consider sample size, access, storage, and trustworthiness of data.

Process Data

- Organize, structure, clean, merge, and transform data in efficient and useful ways.
- Consider additional data cases or attributes.

データの背景の文脈

Key Considerations & Dispositions

Make sense of data with respect to context	Take advantage of technology	Attend to variability & uncertainty	Seek expertise & information	Communicate & collaborate	Be curious creative, & intuitive	Persist & be resilient	Consider ethical issues & biases	Be a skeptic
--	------------------------------	-------------------------------------	------------------------------	---------------------------	----------------------------------	------------------------	----------------------------------	--------------

コミュニケーション・協同的・創造力・批判的な見方

Gamma Mojica, Hollylyme Lee, Emily Thrasher, Zachary Vaskalis, and Greg Ray. (2021). The data investigation process classroom poster. In *Invigorating Statistics Teacher Education Through Professional Online Learning*, Friday Institute for Educational Innovation: NC State University. Available at: <http://cod.instepwithdata.org/DataInvestigationProcessPoster.pdf>

ノースカロライナ州立大学 NSF 研究費

統計・DS教師の指導力向上教育のための統合オンラインプラットフォーム

InSTEP: Invigorating Statistics Teacher Education Through Professional Online Learning

ビジネスへの戦略的な見通し

ビジネスの文脈と優先順位で結果を可視化、実行可能なビジネス・ストーリーを一般向けに語る

統計モデルによる予測・分類 / モデルの評価・改善

多種多様な複数のデータセットの突合

データクレンジング

パターンと特徴量の抽出

THE DATA SCIENCE PROCESS

- 01 Frame the Problem
- 02 Collect Raw Data
- 03 Process the Data
- 04 Explore the Data
- 05 Perform In-Depth Analysis
- 06 Communicate Results

SKILLS REQUIRED

01. FRAME THE PROBLEM
 - Scoping Language
 - Domain Knowledge (needs)
 - Product/Service (metrics)
 - Business Strategy (priorities)
 - Teamwork (specific resources)
02. COLLECT RAW DATA
 - Database Management
 - Systems: MySQL, PostgreSQL, Oracle, MongoDB
 - Querying Structured Databases
 - SQL
 - Retrieving Unstructured Info
 - Information Retrieval / Text Mining
 - Distributed Storage
 - Hadoop HDFS, Spark, Hink
03. PROCESS THE DATA
 - Scoping Language
 - Python or R
 - Data Wrangling & Cleaning
 - Python "Pandas" Library
 - Distributed Processing
 - Hadoop MapReduce / Spark
04. EXPLORE THE DATA
 - Scientific Computing
 - Python: NumPy, Matplotlib, Scipy, Pandas
 - Inferential Statistics
 - Hypothesis Testing
 - Correlation vs. Causation
 - Experimental Design
 - A/B tests, controlled trials
05. PERFORM IN-DEPTH ANALYSIS
 - Machine Learning
 - Supervised (Unsupervised) algorithms
 - Controlled Experiments
 - ML Tools Library
 - Python: scikit-learn
 - Advanced Math
 - Linear Algebra & Multivariate Calculus
06. COMMUNICATE RESULTS
 - Business Acumen
 - Non-technical terminology
 - Data Visualization Tools
 - Tableau, D3.js, Google Visualize, matplotlib, ggplot, custom
 - Data Storytelling
 - Presenting & speaking
 - Reporting & writing

データサイエンスプロセス(サイクル)と各ステップの要素

<https://medium.com/the-mission/deconstructing-data-science-breaking-the-complex-craft-into-its-simplest-parts-15b15420df21>

AIとデータサイエンスの関わり

科学的方法論のパラダイムシフト

1600年頃

1950年頃

2000年頃

第1パラダイム 実験(経験)科学 小規模な観察や実験・調査によるデータから経験則を構築

第2パラダイム 理論科学 モデルを前提に数理理論を展開し公理公則を構築

第3パラダイム 計算科学 (コンピュータサイエンス) 理論(確率)モデルによる大規模シミュレーション

第4パラダイム データ科学 (データサイエンス) センサー、カメラ等から取得される(大規模)データから工学的な分析で法則導出

第1次AIブーム (推論・探索) 1956

第2次AIブーム (知識表現) 1970

第3次AIブーム (機械学習) 1980

1995

2010

2015

リアルワールドエビデンス

100%完全に成立するエキスパートルール → 確率的にしか成立しないデータから学習した傾向

Cyber Physical System



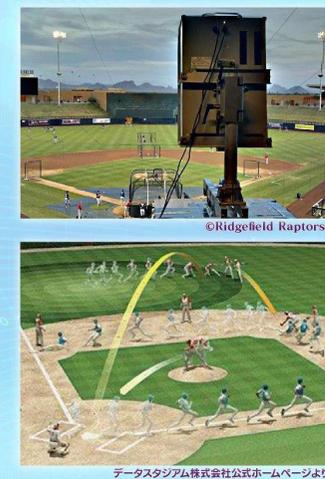
実世界とサイバー空間との相互関係(Cyber Physical System)が社会のあらゆる領域に実装され、大きな社会的価値を生み出していく社会

知識: データから統計的に推論(学習)される原因と結果のルール

データサイエンス技術

- ①データ収集:** リアルな(現実の)対象からデータを収集、対象を多次元の特徴量データを要素とするプロフィールで捉える。
- ②データの蓄積・解析:** データ化されたプロフィールの蓄積(プロフィールの大量観察)から関連性の解析を通して、社会課題の解決に資する知識(ルール)を抽出
- ③解析成果を現実世界へフィードバック:** ②のモデルを自動制御や自動レコメンド機能としてICT機器に搭載

球場での試合のデジタル化



- 計測技術の進化**
- ▶ レーダー式の弾道追尾システム (ボールのトラッキング)
 - ▶ 光学高精細カメラ (選手の動き)
- 取得データの質と量の飛躍的拡大**
- ▶ 投球データ
リリースポイントの位置・球速・回転数(回転速度)...
 - ▶ 打撃データ
打球の速度・角度・飛距離...
 - ▶ 選手
プレー、成績...

2015年 MLB 全球団が導入
2016年 本塁打...7年ぶりに記録更新 約5600本
2017年 本塁打記録の更新...史上最多 約6100本

フライボール革命 (バレルの法則)

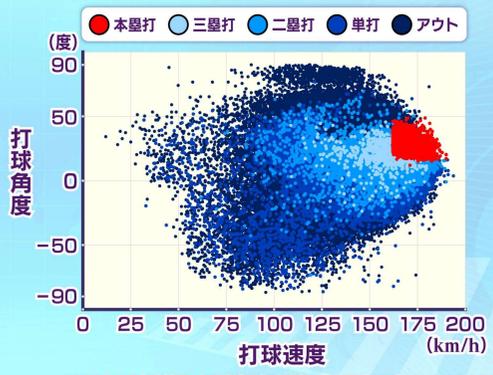
確率的な法則が結果を変える

本塁打になる確率が高くなる打球の条件をデータサイエンスで検出

打球速度 158km/h 以上、打球角度 30度前後

打球ID	打球速度	打球角度	打席結果
00001	194.0454	3.3	単打
00002	191.3101	7.5	単打
00003	191.3101	4.5	単打
00004	190.9883	15.7	単打
00005	190.3447	20.1	本塁打
00006	190.3447	19.8	本塁打
...
127020	20.1125	37.8	アウト
127021	18.5035	27.5	アウト
127022	12.2284	33.8	アウト

データ参照: Baseball Savant



対象のプロファイリング

対象: 打球

打球速度: 200km/h, 180km/h, 160km/h, 140km/h

打球角度: 50°, 30°, 3°, 0°

バレルゾーン

MLB平均
打球速度: 140.8km/h
角度: 11.6°

特徴 (プロフィール)

質的現象 (事象)

打席結果
選手
ホーム or アウェイ
...

量的現象 (事象)

打球速度
打球の角度
飛距離
...

データ表

打球ID	打球速度	打球角度	打球結果
00001	194.0454	3.3	単打
00002	191.3101	7.5	単打
00003	191.3101	4.5	単打
00004	190.9883	15.7	単打
00005	190.3447	20.1	本塁打
00006	190.3447	19.8	本塁打
...
127020	20.1125	37.8	アウト
127021	18.5035	27.5	アウト
127022	12.2284	33.8	アウト

データ参照: Baseball Savant

問題解決のストーリー

- 1 問題の定義
- 2 現状分析
- 3 原因・要因分析
- 4 改善法則(ルール)の規定
- 5 効果の検証
- 6 価値の創出

データサイエンス
データに基づく問題解決と価値創出

打球ID	打球速度	打球角度	打席結果
00001	194.0454	3.3	単打
00002	191.3101	7.5	単打
00003	191.3101	4.5	単打
00004	190.9883	15.7	単打
00005	190.3447	20.1	本塁打
00006	190.3447	19.8	本塁打
⋮			
127020	20.1125	37.8	アウト
127021	18.5035	27.5	アウト
127022	12.2284	33.8	アウト

データ参照: Baseball Savant

Step.2 要因分析 Yの分布を動かす要因 Xの探索

質的データ × 質的データ Xで条件を付けたときのYの(条件付き)分布の比較
割合(確率)の比較

表1 打球速度階級別打席結果の構成割合(%)

打球速度(km/h)	アウト	三塁打	単打	二塁打	本塁打	総計
8-33	96.0%	0.0%	5.0%	0.0%	0.0%	100.0%
33-58	73.0%	0.0%	27.0%	0.0%	0.0%	100.0%
58-83	81.4%	0.0%	18.5%	0.2%	0.0%	100.0%
83-108	80.6%	0.0%	17.6%	1.8%	0.0%	100.0%
108-133	79.2%	0.1%	18.3%	2.3%	0.0%	100.0%
133-158	74.0%	0.6%	18.6%	5.4%	1.4%	100.0%
158-183	40.4%	1.3%	26.4%	14.3%	17.6%	100.0%
183-208	19.9%	0.0%	31.6%	19.9%	28.6%	100.0%
全体	66.6%	0.6%	20.6%	6.8%	5.4%	100.0%

表2 打席結果別にみた打球速度階級の構成割合(%)

打球速度(km/h)	アウト	三塁打	単打	二塁打	本塁打	全体
8-33	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
33-58	0.3%	0.0%	0.4%	0.0%	0.0%	0.3%
58-83	2.3%	0.0%	1.7%	0.0%	0.0%	1.9%
83-108	6.9%	0.0%	4.8%	1.5%	0.0%	5.7%
108-133	26.9%	5.5%	20.1%	7.6%	0.0%	22.6%
133-158	47.1%	38.6%	38.2%	33.7%	11.0%	42.4%
158-183	16.3%	55.9%	34.5%	56.7%	88.2%	26.9%
183-208	0.0%	0.0%	0.2%	0.5%	0.8%	0.2%
総計	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

データ参照: Baseball Savant

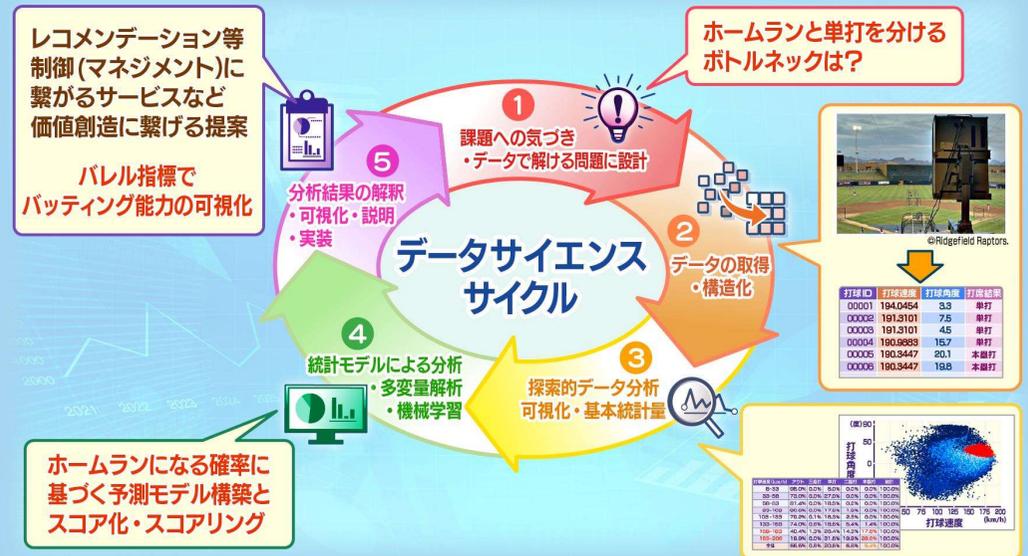
<https://www.chunichi.co.jp/article/328892>

大谷翔平の進化を示す『バレル率』 打球初速&射出角度の指標が大幅アップ…「今季は別の惑星上にいるレベル」

大リーグ公式サイト「2021年に最も向上した選手たち」
大谷は「昨季からバレル率12・8%アップはメジャートップ」

近年大きく扱われるようになった**指数「バレル」**とは、長打になりやすい打球初速と射出角度のコンビネーション。初速は98マイル(158キロ)以上で、角度は30度前後。初速に応じ、バレルに該当する角度も変化する。**バレル率23・5%**(この日の試合後は23・2%)はメジャートップというだけでなく、(大リーグ公式データシステムのスタットキャストが導入された)2015年以降のシーズン最高記録だった17年ヤンキースのアーロン・ジャッジの25・7%に次ぐ数字となっている」と報じた。また、「ところで、彼の“副業”はエリート級の投手だが、被バレル率はどうかと言えば、実は、3分の1以下の6・9%しかない」と伝えた。

統計: 統べて測る(個の大量観察から全体傾向把握; 個から全体)
データサイエンス: AI実装; 個から全体、個に戻す; 予測、レコメド、異常検知、評価指標・・・)



トラッキングデータによるスポーツアナリティクスの世界

試合(対象)の流れ => 複数のイベント(データ化した指標)のプロセスで記述 => 何が何に寄与するのか?
 指標間の関連性を分析 if **=??, then XX, else OO.
 If-then-else 型制御

and we try to track every move of a player

イベントストリーミング分析

事象と事象間の関連性を確率的な傾向で捉える (統計モデリング)

データとは？

データサイエンス データから知識(役に立つルール)を学習する方法とスキル

データ 分析対象(単位) → ex. 顧客、商品、企業、試合、選手、画像、音声、文章、動画 ... を特徴づける複数の変数(指標): **プロフィール**



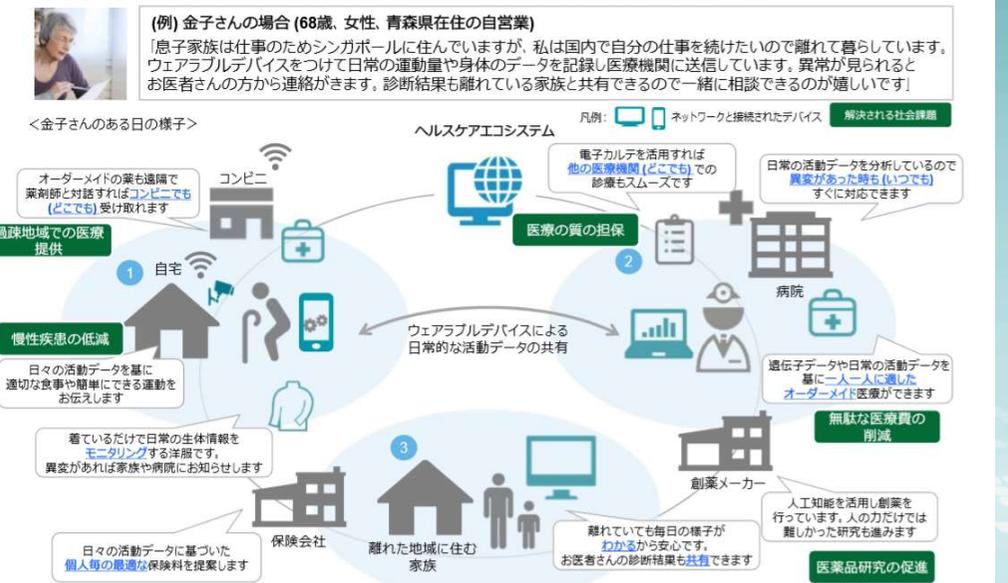
試合 => 複数のイベント(データ化した指標)のベクトル

分析 何が何に寄与するのか? 変数(指標)間の関連性を分析

制御 If-then-else 型制御 → if X=**, then Y=××, else Y=# #.

[3. 健康を維持する、生涯活躍する] データ駆動型社会の新たなヘルスエコシステム

[ヘルスケアエコシステムによる最適医療]が達成された時の社会のイメージ



59 平成26年度製造業技術実態等調査(製造業における“Connected Industries”の推進による付加価値の創出・最大化に関する調査)

データサイエンティストに求められる能力

- 経済産業省
我が国情報経済社会における基盤整備(融合新産業創出に向けた動向調査事業)
- データを扱う統計学の知識
 - 仮説を基にデータを解析する手順を組み立てる能力
 - データを解析するソフトウェアを使いこなす能力
 - データを解析した結果をビジネスに適用するストーリーを組み立てる能力
 - データを解析した結果を可視化して他人に伝える能力



官庁データサイエンティスト、ヘルスデータサイエンティスト、スポーツデータサイエンティスト、ビジネスデータサイエンティスト、気象データサイエンティスト ...

専門職能に繋がる

Real Data, Real Problem Real Learning, Transferable Skill (転用可能な力)

最も重要なスキル: Storytelling with data

自動運転・商品の自動レコメンド・AIによる自動診断・不動産テック・スリープテック・牛の発情期検出予測・AI美空ひばり・スマートトイレ・野球サッカー等入場者予測による価格連動性・・・
全てがデータでつながるSociety 5.0の社会/驚異的な入試倍率となっているデータサイエンス系学部
 MUSYC(アエラ2019.05):データサイエンス・AIの教育を全大学・全学部で基礎～応用基礎(発展)実施
 (AI戦略2019による教育改革) / U.S.News and World Report2021 Statistician(全6位、ビジネス2位), Data Scientist(全8位、技術職2位), Forbs: Best Job in America2019 Data Scientist(1位), Data Analyst(31位)

内容

- 立正大学データサイエンス学部
 文理融合カリキュラムと政策的な背景
- 文系学生にとってのデータサイエンス教育
- **データサイエンス専門職能の拡大**
 : データビジネスリーディングマネージャー
 (Analytics Translators)
 (データ活用・分析の価値・コスト・リスクを組織で管理・展開し
 ビジネス価値につなぐ職能)

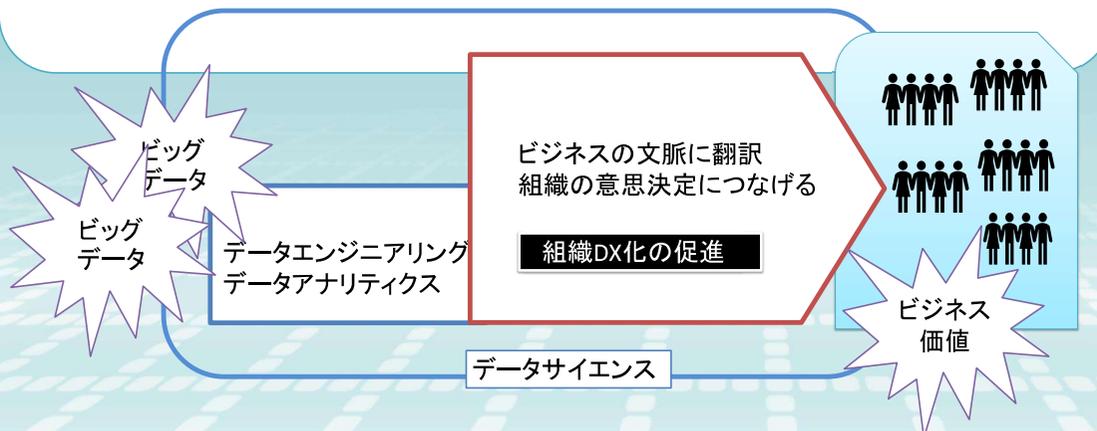
2026年までに、米国だけで
 200万人から400万人の需要
 (McKinsey Global Institute)

組織の中での
データサイエンス人材の機能分化
 : データビジネスリーディングマネージャー
 (Analytics Translators)
 (データ活用・分析の価値・コスト・リスクを組織で管理・展開し
 ビジネス価値につなぐ職能)
 2026年までに、米国だけで200万人から400万人の需要 (McKinsey Global Institute)

- You Don't Have to Be a Data Scientist to Fill This Must-Have Analytics Role – Henke, Levine, McInerney (HBR, Feb 2018)
<https://hbr.org/2018/02/you-dont-have-to-be-a-data-scientist-to-fill-this-must-have-analytics-role>
- [analytics] translators help ensure that the deep insights generated through sophisticated analytics translate into impact at scale in an organization. By 2026, the McKinsey Global Institute estimates that demand for translators in the United States alone may reach two to four million.

組織の中での
データサイエンス人材の機能分化
 : データビジネスリーディングマネージャー
 (Analytics Translators)
 (データ活用・分析の価値・コスト・リスクを
 組織で管理・展開しビジネス価値につなぐ職能)
 2026年までに、米国だけで200万人から400万人の需要 (McKinsey Global Institute)

データサイエンスの知識がある
 最高データ責任者 最高デジタル責任者
 ビジネス開発マネージャー
 プロジェクトマネージャー
 データコンサルタント BIエキスパート
 データアナリスト データビジュアライザー
 ...



「AI時代に対応した人材」とは、単一ではなく、

- ・最先端のAI研究を行う人材
- ・AIを産業に応用する人材
- ・中小の事業所で応用を実現する人材
- ・AIを利用して新たなビジネスやクリエイションを行う人材

などのカテゴリーに分かれるが、いずれにしても、各々のカテゴリーでの層の厚い人材が必要となる。

人材の増大には、女性も含む多様な人材や、海外から日本を目指す人々も含め、それぞれの層に応じた育成策、呼び込み策が重要である。そのため、今後、先進的な教育プログラムの構築が重要であり、さらに、これを海外にも提供できるレベルにまで充実させることも必要になる。

多様性に対する無意識を意識するシンボル

International Year of Women in Statistics and Data Science (IYWSDS)

とうけいあかりちゃん



フローレンス・ナイチンゲール
生誕200年祭

- In this section:
- [Events](#) »
 - World Statistics Congresses
 - Other conferences
 - Sponsorship or endorsement
 - Webinars
 - Courses



The ISI community celebrated the IYWSDS from May 2020 through July 2021. It was launched on 12 May 2020 with commemorating the *200th anniversary of Florence Nightingale's birth.*

統計学は世の中で最も重要な科学である。なぜなら、他のあらゆる科学、あらゆる芸術の実践的応用が統計学にかかっているからである。すべての政治・社会行政、すべての教育、すべての組織にとって不可欠な、経験に基づく科学であり、それは我々の経験によって得られる事実に関する結果のみを提供するからである

Statistics is the most important science in the whole world: for upon it depends the practical application of every other science and of every art: the one science essential to all political and social administration, all education, all organization based on experience, for it only gives results of our experience. Florence Nightingale