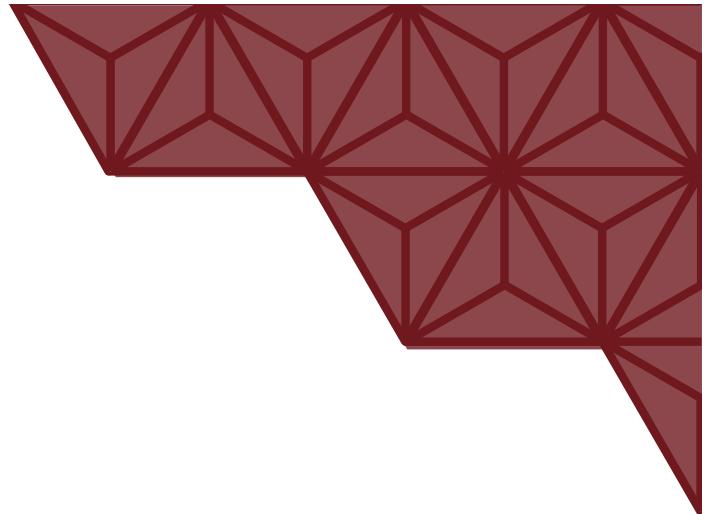


お茶の水女子大学におけるデータサイエンス教育の紹介

由良 敬

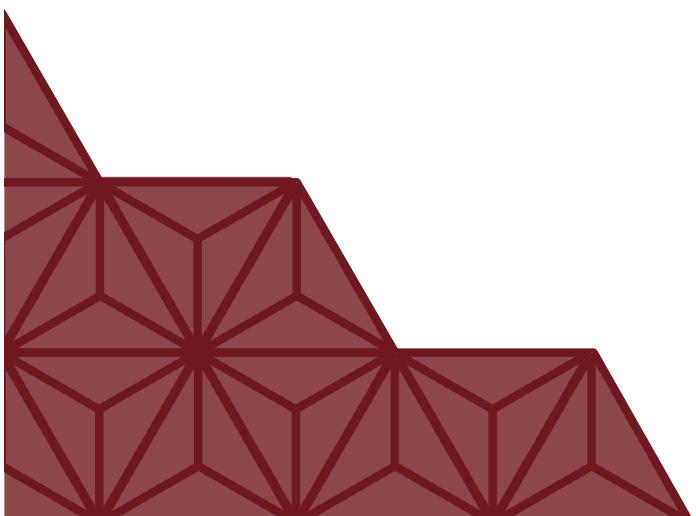
お茶の水女子大学
文理融合AI・データサイエンスセンター
(早稲田大学 先進理工学部)



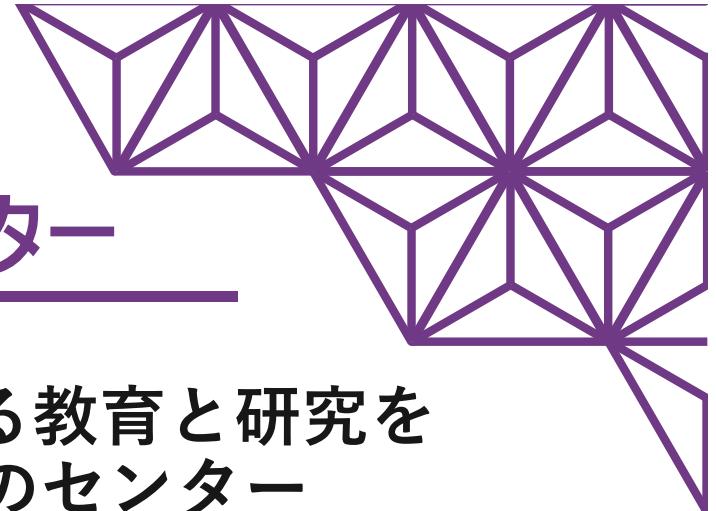
お茶の水女子大学は2019年

「文理融合AI・データサイエンスセンター」

を設立しました



文理融合AI・データサイエンスセンター



- AIおよびデータサイエンスに関する教育と研究を全学的に推進するための組織横断型のセンター

教育部門

文理融合データサイエンス
科目の全学的展開

データサイエンス科目群の
学際カリキュラム策定

データサイエンス教材の
協定校・附属校への提供

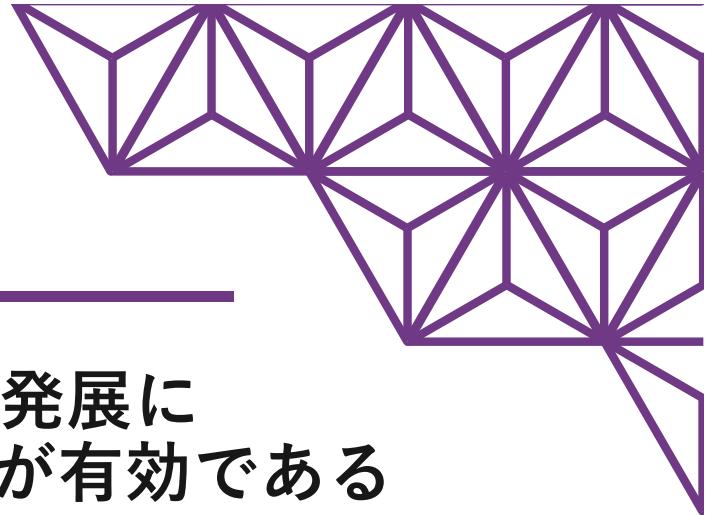
研究部門

学内共同研究や
产学連携の推進

AI関連研究に従事
する学生の支援

データ管理環境・
知的財産の強化





なぜ文理融合でAIなのか(1)

- ・文学・社会・芸術・心理などの学問の発展にツールとしてのAIやデータサイエンスが有効である

文学作品の作風の違いは
どこに起因するのか発見したい

災害時にどんな意外な人間行動が
あったのかを発見したい

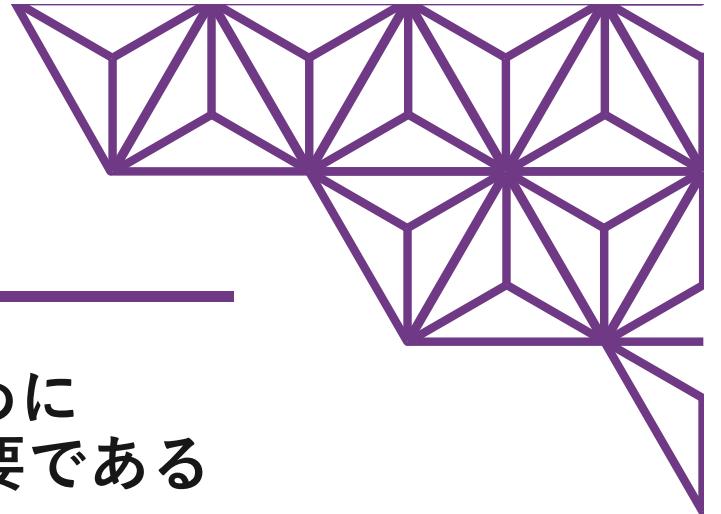
初学者と熟達者の芸術作品は
何が違うのか究明したい

言語ごとの日常生活の
会話の違いを発見したい

急な不景気と因果関係のある
未発見の事実を突き止めたい

どんな要因でどういう心理が
生じるのか実験結果から発見したい





なぜ文理融合でAIなのか (2)

- AIやデータサイエンスの正しい発展のために
法律・経済・倫理・心理などの議論が必要である

AIの普及に先がけて
どんな法規制が必要か

AIの普及によってどんな
新しい産業が生まれるか

AIの活用において倫理的に
何が許され何が許されないか

AIが引き起こす新たな犯罪を
どのように未然に防ぐか

AIの普及は企業経営を
どのように変えるか

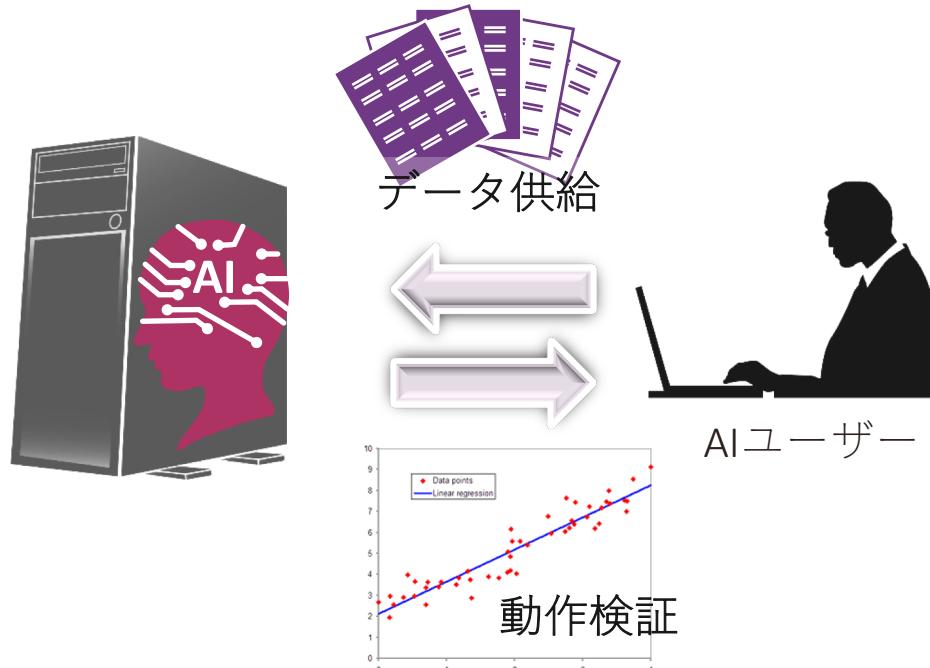
AIの普及は日常生活に
どんな新しいストレスをもたらすか



なぜ女子大学でAIなのか



- AI社会の新たな不平等（例えば男女格差）の解消を目指す



AIユーザには幅広い教養と
公平な視点が求められる



朝日新聞 2019年3月9日の記事より

(Dear Girls) 男女格差、AIが広げる?
村上明子さん、江田麻季子さん、津田大介さん

● 有料会員限定記事
2019年3月9日05時00分

シェア 0 ツイート 1181 ブックマーク 3 メール 印刷 面一覧

人工知能（AI）が社会で急速に広まる中、「AIは男女格差を広げる」という指摘がでている。開発者や、入力データに女性の視点が少ないのが一因だが、本当に格差を広げるのか。

■データ学ばせる側に偏り 村上明子さん
(日本IBM Watson開発アーキテクト)

開発者や利用者が
男性ばかりだとデータに
偏りが出る恐れがある

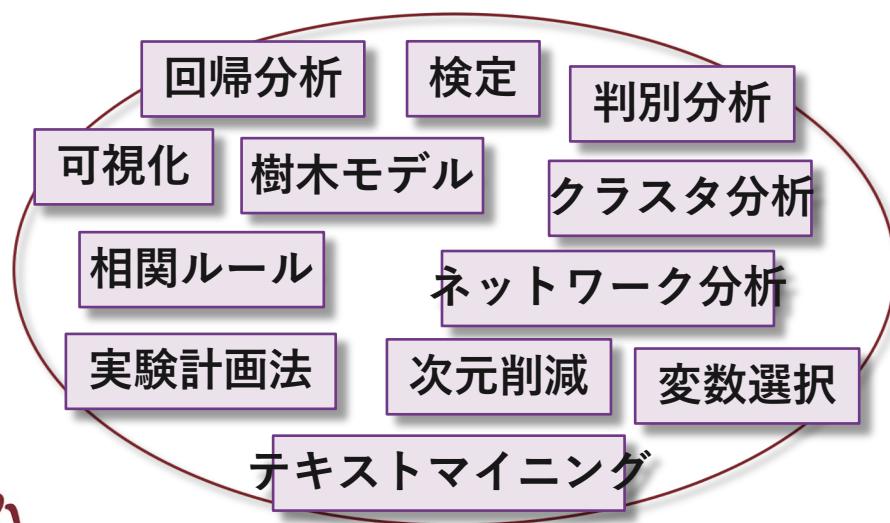
情報学を学ぶ女性は
男性より圧倒的に少なく
格差が広がりつつある

センターの活動(1) 「数理・データサイエンス教育協力校」として



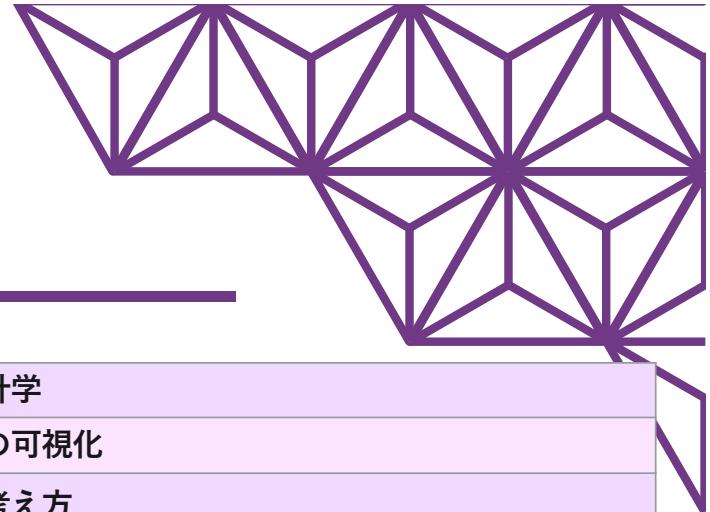
- 新設科目「文理融合データサイエンスI,II」
 - 主に学部1,2年生の全学部を対象
- データサイエンス相談室（週2回）
- データサイエンス教材の協定校・附属学校への提供

文理融合データサイエンスI,II



「文理融合データサイエンスI」

コア科目(1年後期)



本授業の主題はデータサイエンスの手法を学ぶことに加え、文化現象を対象にデータサイエンスの手法を用いた事例を概観し、文理融合型の研究におけるデータ分析の有効性と重要性を学ぶことです。

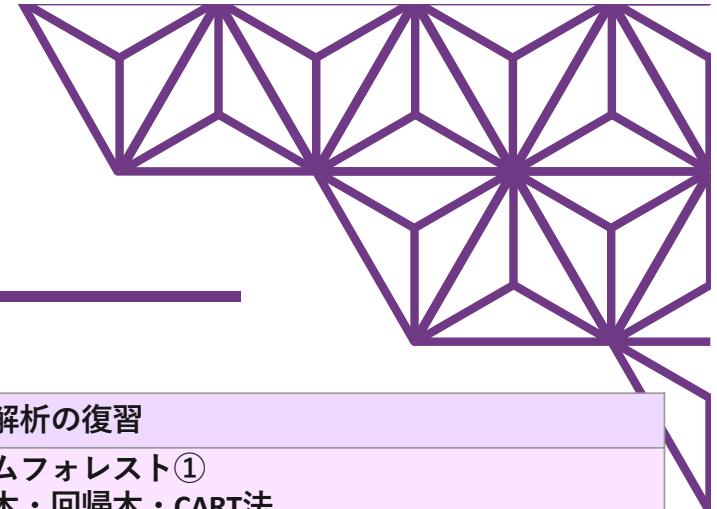
それ故、本授業ではデータサイエンスの手法を理解することだけではなく、文化に関するデータに対しデータサイエンスの分析手法をどのように応用するのか理解することが重要になります。

第1回	記述統計学
第2回	データの可視化
第3回	確率の考え方
第4回	推測統計学
第5回	さまざまな検定
第6回	事例(1)：区間推定による薬師寺の建立年代の推定
第7回	回帰分析1 事例(2)：シミュレーションデータの線形回帰分析
第8回	判別分析 事例(3)：日本近代文学における自然主義作家の識別
第9回	回帰分析2 事例(4)：ロジスティック回帰分析
第10回	主成分分析 事例(5)：主成分分析を用いた夏目漱石の小説の分析
第11回	対応分析 事例(6)：クロス集計と対応分析
第12回	クラスター分析 事例(7)：小説のクラスタリングによる著者識別
第13回	樹木モデル 事例(8)：決定木を用いたシミュレーションデータ分析 事例(9)：決定木を用いた短い文章の書き手の判別
第14回	樹木モデルの応用 事例(10)：『源氏物語』と『うつほ物語』の特徴語の抽出
第15回	多変量解析の復習



「文理融合データサイエンスII」

コア科目(2年前期)



本授業の主題は文理融合データサイエンス I において学んだ基礎的な多変量解析の手法を復習し、新たに機械学習の手法を学ぶことです。

これに加えて、文化現象を対象とするデータサイエンスの手法を用いた研究事例を概観することで、文理融合型の研究におけるデータサイエンスの有効性と重要性を理解することが目標となります。

第1回	多変量解析の復習
第2回	ランダムフォレスト① 決定木・回帰木・CART法
第3回	ランダムフォレスト② ブートストラップ・分類問題と回帰問題・変数重要度
第4回	ランダムフォレストの演習
第5回	ネットワーク分析 有向グラフと無向グラフ・隣接行列・中心性
第6回	ネットワーク分析の演習
第7回	これまでの復習および中間レポートの説明
第8回	ニューラルネットワーク
第9回	ニューラルネットワークの演習
第10回	SVM ソフトマージンとハードマージン・カーネル法
第11回	SVMの演習
第12回	相関ルール
第13回	相関ルールの演習
第14回	深層学習の紹介
第15回	これまでの総復習





データサイエンス教材の提供

- 多くの大学への波及を目指してデータサイエンス教材を制作中

【教材1】 対面授業向けの教材一式

- 本学コア科目「文理融合データサイエンスI,II」の内容に準拠したスライド、課題、R言語によるサンプルプログラムなど

【教材2】 ビデオ形式のe-learning教材

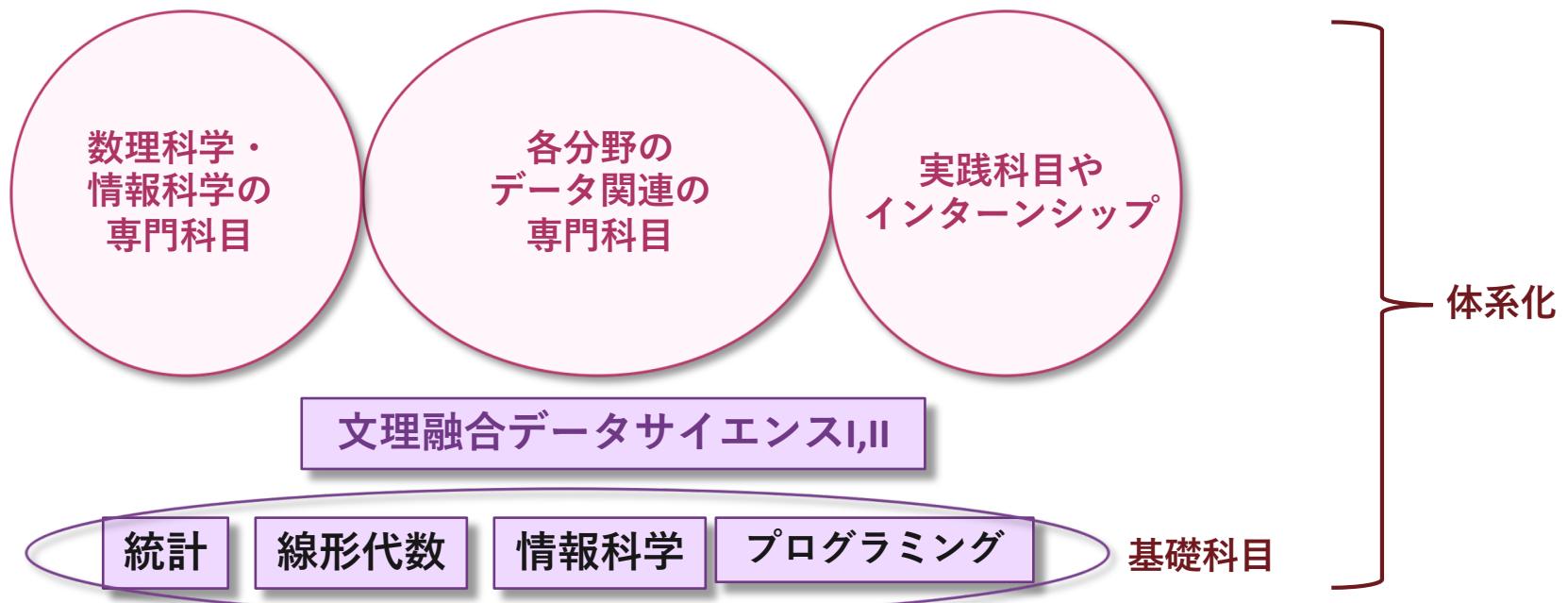
- データサイエンスを構成する25項目を各々20分以内のビデオで制作（変数、標本、推定、検定、回帰、判別、主成分分析、クラスタ、ネットワーク、アソシエーション、決定木など）



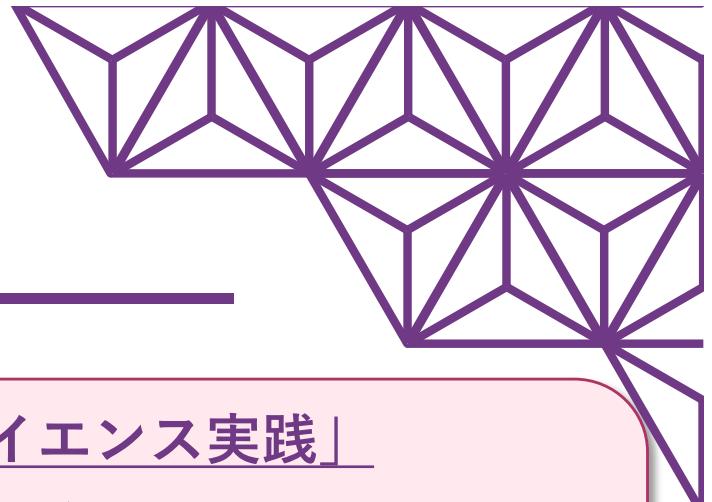
センターの活動(2) 高度なデータサイエンス教育



- 卒業研究でデータを使いこなすための新規科目
- 企業との連携による実践型の新規科目
- 学内のデータ関連科目の体系化（学際カリキュラム）



企業との連携科目



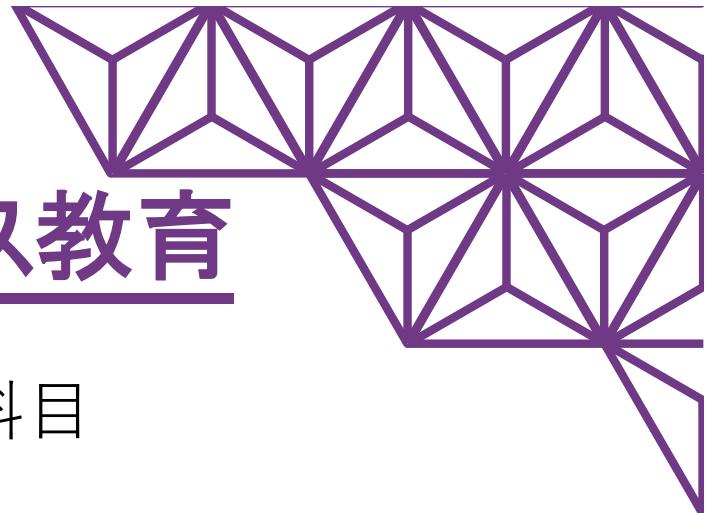
「デジタル概論」

- 全学コア科目
- 「デジタル人材」として社会で活躍できるスキルの体得
 - 上期：デジタルの活用による企業の変革状況を実例を交えて解説
 - 下期：社会課題にアラインしたテーマに沿って実際のモノづくりを通して学生自身が実践

「データサイエンス実践」

- 大学院博士前期科目
- **データサイエンスが求められる背景、データサイエンスを構成する技術と活用についての解説**
 - 社会および企業での関わり
 - 中核となる技術と活用
 - データサイエンスの活用による新しいビジネスの創出方法
 - 社会および企業での影響・課題
 - キャリアデザインとの関連性



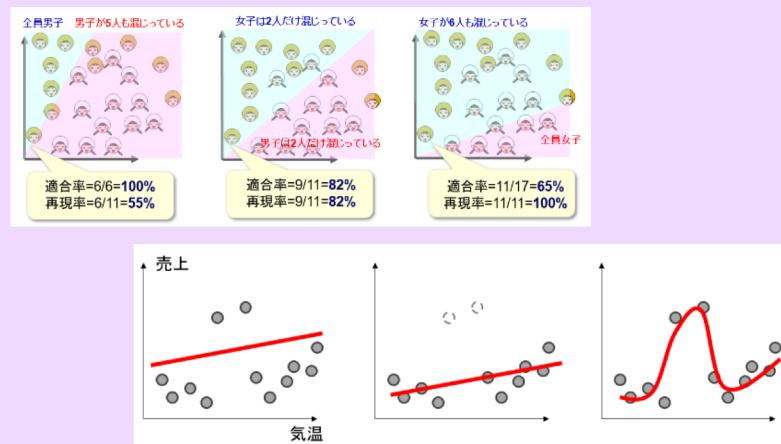


附属高校でのデータサイエンス教育

- SSH (Super Science High school) 1年生科目
「課題研究基礎」での講演と実習

データサイエンスの講演

判別・回帰・可視化の概念を
実例や例題を交えて解説



データサイエンスの実習

音楽の判別分析をウェブ上の
アプリケーションで体験実習

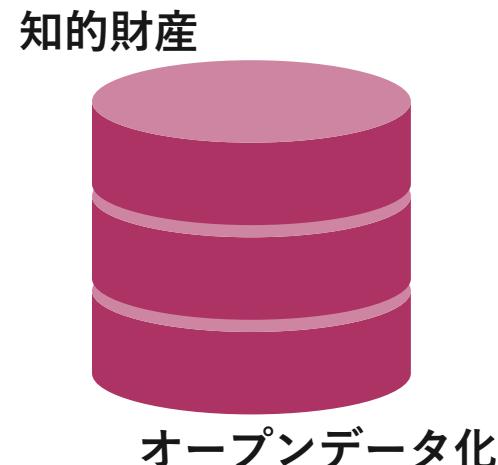


※2021年度以降 2年生「課題研究I」 3年生「課題研究II」 も支援予定

センターの活動(3) AI・データサイエンス関連研究の推奨



- 学内共同研究・产学連携研究の推進
- 知的財産・研究データ管理などとの連携
- 学生研究の支援（研究費助成・PC貸出など）
- 大学院生の進路支援（学生キャリア支援課らと連携）



第5回DSワークショップ@お茶大の宣伝

日時：12月23日(水)

プログラム概要：

土山 玄（お茶の水女子大学文理融合AI・データサイエンスセンター）

「文学作品のテキストデータを題材としたデータサイエンス演習」

松浦 司（中央大学経済学部／お茶の水女子大学非常勤講師）

「政府統計を用いたデータサイエンス演習の現状」

伊藤貴之（お茶の水女子大学文理融合AI・データサイエンスセンター／理学部情報科学科）

「附属高等学校でのデータサイエンス実習事例」