

1. 授業の概要(ねらい)

この授業は、八王子キャンパスで学ぶ学生を対象としたデータサイエンス入門科目の一つとして提供するものです。可能であれば、後期開講の「データサイエンス・AI入門(2単位総合基礎科目)」とセットで受講することをおすすめします。なお、教養科目のデータサイエンス入門科目としてどちらか一つを選択する場合は、後期開講「データサイエンス・AI入門(2単位総合基礎科目)」を優先して受講してください。さて、ネットワークの高速化、コンピュータ処理能力の飛躍的向上が、デジタル方式のビッグデータを使った合理的な分析による業務の効率化・高度化を促進しています。近年のAI技術進展も、その追い風になっています。しかしながら、現時点でわが国におけるデータの大規模な活用が効果的になされているとは言えません。例えば、国の組織でも各省庁間で漢字表記の統一がされていないため(日本経済新聞2020年9月13日「省庁データ表記統一へ」、結果として統合検索ができない」という記事がそのことを象徴しています。その意味でも、社会構成員のデータサイエンス対応能力向上は喫緊の課題です。データサイエンティストは各人の専門分野を反映していくつかの類型に分けることができます。理系であれば、データを整理加工する技術的基盤を構築する者、AIに代表される機械学習の開発に従事する者。文系の場合は、各自の専門領域をベースに、取得したデータの分析を行う者、業務への応用の仕組みを考える者、目的に合わせていかなるデータを取得してどのように分析すべきかという制度設計を提案する者が考えられます。この授業では、後者に焦点を合わせた内容を重点的に扱います。

2. 授業の到達目標

知識・理解の観点…データサイエンスの全体像を把握し、情報検索・整理・分析・活用で必要とする基本的事項および情報倫理や法的側面との関連性を理解する。
思考・判断の観点…ビッグデータを含む各種の情報について、正確に情報の意味を理解し法的・倫理的観点も含めて合理的に整理できる。
関心・意欲の観点…ビッグデータを含む各種の情報を取得・整理・活用する局面で積極的な対応ができる。
態度の観点…ビッグデータを含む各種の情報を統合し、データ活用について自身の専門領域で初歩的な提案ができる。
技能・表現の観点…ビッグデータを含む各種の情報を検索・整理・可視化し、情報活用について内在するリスク対応も含めて他者に説明できる。

3. 成績評価の方法および基準

評価方法: 1. 授業最終回提出の最終レポート(10%) 2. 毎回提出するワークシート(小テスト)の点数合計(90%)
評価基準: 1. 毎時間のワークシート(小テスト)…知識、理解、検索および分析技能を測定 各種のデータに対する知識理解と、例題レベルの検索および分析を行うことができる。
2. 最終レポート…思考、判断、関心・意欲、検索分析能力の形成を測定 講義(予習復習を含む)を通して獲得した知識とスキルを統合し、初歩的なデータ分析とそれを利用した仕組みの提案ができる。

4. 教科書・参考文献

教科書
なし…LMSからスライド・資料を配信します
参考文献
なし

5. 準備学修の内容

①事前に、下記ホームページをブックマークして、各自でアカウント登録(トップページの右上部分)を済ませるとともに、サイトの全体像を把握してください。どのようなデータが収録されているかざっくりと閲覧するだけで結構です。詳しくは授業で説明します。
政府統計の総合窓口(e-Stat)
<https://www.e-stat.go.jp/>
②ntt docomo が提供するモバイル空間統計のサイトで、何ができるか検討してみてください。
<https://mobaku.jp/>

6. その他履修上の注意事項

データ検索はスマホやタブレットでもできるように配慮しますが、可能であれば授業時にパソコン持参をお勧めします。
※成績評価は、授業最終回提出の最終まとめレポートと、毎回LMSで提出するワークシート(小テスト)の点数を合計したもので評価します。

7. 授業内容

【第1回】 講義内容の説明
文系学生を対象とする当該科目の概要と成績評価方法の説明。
最近のビッグデータやAI活用事例の代表例を紹介し、授業後半でビッグデータを使ってどのようなことができるのか検討します。ntt docomo が提供する「モバイル空間統計」サイトで帝京大学周辺、新宿周辺、八王子周辺地域の一日の人口密度の時間帯による変化を確認、その後このデータから読み取ることができる内容を議論します。また、船舶位置情報(マリントラフィック)を利用した同様の検討を行います。

【第2回】 各種のビッグデータ取得方法や特性を検討する
現在利用可能となっている様々なビッグデータの流通形態、取得方法、特性等を説明するとともに、各学生の専門領域の観点からこれらのビッグデータ利用方法について議論します。地図情報、交通系ICカード情報、携帯電話位置情報、衛星データ情報、商標情報、音楽著作物情報等々…を説明し、あわせて活用の成功例として渋谷駅の人の動線改良工事を検討します。

【第3回】 ビッグデータ活用の前提条件を検討する
ビッグデータを統合活用する際に必要となる要件を説明し、文字情報を例に検索に支障が出る事例と対処方法を検討します。文字による検索で支障が発生するモデル事例を提示して、なぜ検索ができなかったのかその原因の探索と対処方法を議論するとともに、コンピュータにおける文字コードの仕組みを簡単に説明します。また、日本経済新聞2020年9月13日記事「省庁データ表記統一へ」も使って、統一した検索を可能にする制度上のしくみも議論します。

- 【第4回】 AIって何だろう
ビッグデータ活用手段として期待されるAIについて、これまでの開発の経緯と現時点の技術開発の到達点、今後の技術予測を説明します。AIの基本的な仕組みを理解し、その仕組みから推測される現時点の限界を議論します。ここでは、人事評価AIシステムおよび銀行の融資信用力判定AIシステム利用で実際に生じた問題を扱います。あわせてGoogleが提供するwebベースシステムのTeachable Machineを使いその限界を実感することとします。演習(人と猫の顔判定)で利用する写真群は担当教員から提供しますが、学生が保有する任意の写真による独自検証でも構いません。
- 【第5回】 ビッグデータを分析するシステム設計時の前提条件を検討する
ここで使う「前提条件」検討とは、システムを駆動するコンピュータプログラム開発をすることではありません。例えば、特定のwebニュースサイトのアクセスランキングを作成するシステムを構築する際に、そもそも論として、ここで必要とするランキング自体の定義、ランキング利用目的の整理から始まり、その目的に合致したデータ生成の方法等を決定することです。例えば、各記事の一定期間における絶対的アクセス数ランキングか、一定期間の増加率か、計上する各記事のジャンルや文字数等で重みづけをするのか、これらの指標を組み合わせるのか、システム開発の前には目的に合わせてこれらの前提条件を確定することが必要です。担当教員が指定する特定のニュースサイトを使って、アクセスランキング作成の前提条件確定の演習を行います。
- 【第6回】 ビッグデータやAIシステム上でデータを扱う場合の留意事項(1)
個人情報に紐づいたビッグデータ利用の可否及び利用可能するための情報整理について、個人情報保護法(制定から現時点での改正情報を含む)、EUのGDPR(一般データ保護規則)を詳しく説明する。JR東日本の交通系ICカードの利用履歴情報販売の事例、顧客のレンタルビデオ貸し出し履歴の提供事例等で、あるべき姿を議論します。
- 【第7回】 ビッグデータやAIシステム上でデータを扱う場合の留意事項(2)
前回に引き続き、オプトアウト(個人データ提供を自身でコントロールするしくみ)の説明と具体的なオプトアウトの例、クッキーの利用可否と代替技術として台頭しているDF(デバイスフィンガープリント)に対する考え方。DFとGDPR(一般データ保護規則)の関係を検討します。
- 【第8回】 ビッグデータやAIシステム上でデータを扱う場合の留意事項(3)
いわゆる「忘れられる権利」について、ネットワーク上で生じている現状分析を行うとともにEUと日本でそれぞれについて規則制定や判例の経緯を説明します。これについて、今後のあるべき姿をいくつかのモデルケースを使って議論します。授業後半では、第6回から第8回の授業で修得した内容を踏まえてAI社会原則(公平性、説明責任、透明性、人間中心の判断)を検討します。
- 【第9回】 ビッグデータやAIシステム上でデータを扱う場合の留意事項(4)
情報セキュリティ(気密性、完全性、可用性)の基本的な概念を説明します。あわせて、暗号化の基本的な手法、匿名加工情報の定義、セキュリティ事故の事例も説明します。後半部分で、不正競争防止法の営業秘密に関する事案(通信サービス会社の社員移籍とそれに伴う4G・5G基地局情報等の漏洩事件)を検討します。併せて、サイバー犯罪の現状と対処方法、その他の法的側面から検討する内容も扱います。
- 【第10回】 【オンライン授業…LMSによるオンデマンド授業】
ビッグデータ利用・AI技術進展に伴う光と影
完全自立型を目指すAIを利用した自動運転車の開発、それに伴う社会制度の提案とリスク管理を検討します。また、AIによる自立型兵器の開発運用の現状を把握し、国際機関での議論の推移、今後のルール策定等についても議論します。その他にも、生命倫理の視点から議論するテーマを扱い、後半部分では、ビッグデータを駆使して真実に迫る「デジタルハンター」の活動を紹介します。
※この回の授業は、LMS上に配置した音声ファイルとスライド等の資料を利用したオンデマンド形式で実施します。なお、同じ内容の動画による配信を下記から実施します(2022年5月15日から配信)。ご自宅からの利用で回線容量が無制限の環境であれば、下記の動画サイトからの視聴をお勧めします。
https://www.kim-lab.info/domescon/t2022soinfo1/t2022soinfo1_no10.html
- 【第11回】 政府統計の総合窓口(e-Stat)サイトの利用(1)
流通するビッグデータの活用演習として、日本政府が提供する統計データ取得の総合窓口サイトを利用して情報取得から分析、そして分析を応用した仕組みの提案を行います。第11回目は、取得できるデータの概要と、設定したいいくつかのテーマに基づくデータ取得を扱います。
- 【第12回】 政府統計の総合窓口(e-Stat)サイトの利用(2)
前回に引き続き、e-Statサイトを利用した演習です。第12回目は、コロナウイルスが蔓延する中での人々の行動変化を、2020年月度統計データの推移を使い事実に基づいて客観的に分析します。都市・地方間の人口移動、国内外の人口移動、輸送量関連情報等々の統計データを元に検討します。
- 【第13回】 政府統計の総合窓口(e-Stat)サイトの利用(3)
第11回目と第12回目授業で修得したスキルを元に、各学生の専門分野にあわせた最終レポートのテーマ設定とデータ取得を行います。
- 【第14回】 政府統計の総合窓口(e-Stat)サイトの利用(4)
第11回目と第12回目授業で修得したスキルを元に、各学生の専門分野にあわせた最終レポートの内容を決定しあらすじを作成します。
- 【第15回】 ビッグデータを利用した最終レポート作成とまとめ
これまでの授業で修得した知識とスキルを使い、受講学生が独自に設定したテーマで最終レポートを作成します。