

データサイエンス基礎演習

必修 4単位

城戸 隆

1. 授業の概要(ねらい)

データサイエンス基礎概論に対応した基礎演習を行う。RもしくはPythonを用いた演習によりデータ解析の基礎となる実践的スキルを身に着けながら、データを読み解くことの面白さ、分析の大切さを知る。2つのコースワークを行う。

●コースワーク1:「<AIと人>に関する参加型ディスカッション」(グループワーク)

「AIと人」に関わりのあるシナリオを与え、グループディスカッションを行う。ディスカッションを通して得られた気づきをレポートにまとめる。シナリオは、近未来の社会課題、倫理的課題など。

ディスカッションテーマの対象: AIがもたらす変化、仕事の未来、公平性、データバイアス、Basic Income、ブロックチェーン、パーソナルゲノム、倫理的課題等。「近未来の社会課題に対して、AIと人が協働して、どのような未来を創造できるか?」

例えば、マルチステークホルダーで、役割を与えたシミュレーションのディスカッションを行う。グループで、意見をまとめてプレゼンテーション、ディスカッション、最後にレポートを提出する。

シナリオの例1: 日本のIT企業X社が万引き犯によくみられる疑わしい特徴や挙動等のある顧客を抽出してリアルタイムで従業員に通知する顔認識、挙動探知技術を搭載した店舗内カメラを開発、民間施設への導入を進めている。政府、IT企業、市民、警察の視点から、このサービスを推進していくべきか?

シナリオの例2: 個人の遺伝子情報をもとに、個人の病気のリスクや薬効を予測し予防医療に生かそうとするサービス、さらには、医療の枠を超えて個人の潜在能力や才能発見という分野に生かそうというサービスが登場してきている。今後、どのような未来がありうるか? 推進すべきか、規制すべきか?

●コースワーク2:「オープンデータ解析」(データサイエンス教育 パブリック データセット)

データサイエンス教育のために開発されたパブリックデータセットを用いて、データ解析を体験しGood Practiceとの比較を行う。レポート作成、プレゼンテーションを行う。

データサイエンス教育のために開発されたパブリックデータセットとして、例えば、政府統計(市町村統計)のSSDSE(教育用標準データセット)を用いる。(あるいはCOVIDに関するPublic Databaseを用いる。)

☒ Good Practiceとして、例えば、過去の統計データ分析コンペティションの表彰論文を用いる。(あるいはCOVIDに関するPublic Databaseを用いる。)

☒ 過去の統計データコンペティションと同じ課題を与え、データサイエンスの一連の流れにそって、オリジナルな課題設定、解析計画、データ解析、考察を行いレポートを提出する。(あるいはCOVIDに関するPublic Databaseを用いる。)

☒ 提出レポートについてのプレゼンテーションを行い、得られた気づきをグループで共有する。

☒ 政府、金融経済、画像、感情分析、自然言語処理、自動運転、などのオープンデータセットとGood Practice、データサイエンスのオンライン教材等を紹介し、今後の学習(研究)の目標設定(テーマ設定)につなげていく。

2. 授業の到達目標

- ・現代のデータ社会に対して主体的な問題意識とパースペクティブを育むこと。
- ・「これからのAI時代に本当に必要なスキルは何か?」について、自らの進路にとって必要な目標を設定できること。
- ・データとAIがもたらす未来について、ありうる未来の姿を想像し、色々な考え方や価値観がありうることを知り、多様な価値観を尊重する態度を身につけること。世の中の誤解や間違った思い込みに気づき、データをもとに世界を正しくみる習慣の重要性を身近な実例を通して理解すること。
- ・データを読み解くことの面白さ、分析の大切さを知る。
- ・問題解決に必要なデータを適切に解析し、仮説や検証、予測や評価を行う力。
- ・必要最小限の情報処理、基礎プログラミング力。(Excel + R or Python)
- ・オンライン教材で自発的にスキルを獲得し、データコンテスト(Kaggle など)などに参加できる。

3. 成績評価の方法および基準

授業中に行われるディスカッションやグループワークへの参加、小テスト(30%・LMS上で実施)、期末レポート(70%)によって評価します。

4. 教科書・参考文献

教科書

竹村彰通、姫野哲人、高田聖治 (編) データサイエンス入門(データサイエンス大系) 学術図書出版社

参考文献

ハンス・ロスリング et al. ACT FULNESS 10の思う巨海を乗り越え、データを基に世界を正しく 日経BP

統計でウソをつく法—数式を使わない統計学入門— ブルーバックス

データサイエンス講義 オライリー・ジャパン

ネイト・シルバー シグナル&ノイズ 天才データアナリストの「予測学」 日経BP

revor Hastie (著), Robert Tibshirani (著), Jerome Friedman (著), 杉山 将 (翻訳) 統計的学習の基礎—データマイニング・推論・予測— 共立出版

C.M. ビショップ (著), 元田 浩 (監訳), 栗田 多喜夫 (監訳), 樋口 知之 (監訳), 松本 裕治 (監訳) パターン認識と機械学習 上下 丸善出版

黒滝 紘生 (翻訳), 河野 慎 (翻訳), 味曾野 雅史 (翻訳), 保住 純 (翻訳), 野中 尚輝 (翻訳), 富山 翔司 (翻訳) 深層学習 KADOKAWA

<https://www.nstac.go.jp/SSDSE/> SSDSE教育用標準データセット

<https://www.nstac.go.jp/statcompe/past.html> 統計データ分析 コンペティション 受賞論文

中山浩太郎「監修」、松尾豊 [協力]、塚本邦尊、山田典一、大澤文孝 [著] 東京大学のデータサイエンティスト育成講座 Pythonで手を動かして学ぶデータ分析 マイナビ

有賀友紀、大橋俊介 RとPythonで学ぶ実践的データサイエンス&機械学習 技術評論社

5. 準備学修の内容

- ・コンピュータの基本操作(起動終了、文字入力、ファイル操作など)を習得していること。
- ・高校1-2年程度の数学の知識、スキル、Excelで基本的なことができる程度の素養。
- ・専門領域に関する具体的な課題や問題意識をもっていることが望ましい。(必須ではない。)

6. その他履修上の注意事項

7. 授業内容

- 【第1回】 時代背景と本コースの位置づけ、事前アンケート、演習環境の確認。
- 【第2回】 データの取得と管理1:実習
- 【第3回】 データの取得と管理2:実習
- 【第4回】 データ分析の基礎 (1):実習
- 【第5回】 データ分析の基礎 (2):実習
- 【第6回】 統計データ分析 実習
- 【第7回】 データ分析で注意すべき点, Good Practice. 実習
- 【第8回】 コースワーク1: グループディスカッション
- 【第9回】 データサイエンスの手法 (1) 実習
- 【第10回】 データサイエンスの手法 (2) 実習
- 【第11回】 データサイエンスの手法 (3) 実習
- 【第12回】 コースワーク2: オープンデータ解析1
- 【第13回】 コースワーク2: オープンデータ解析2
- 【第14回】 コースワーク2: オープンデータ解析3
- 【第15回】 課題発表。より進んだ学習のために。
※各回の内容には、授業の進行ペースや皆さんの関心の傾きに依りて調整を加えることがあります。