

開講科目名	データサイエンス演習				
成績入力担当	大谷 亨、山口 智子、中橋 龍			開講区分	単位数
				第4クォーター	1.0単位
ナンバリングコード		曜日・時限等	金3、金4(対面)	時間割コード	4MM01

授業のテーマ

データサイエンスは、スマートフォンやウェアラブルデバイスを用いて多数より継続的に取得されるバイタルサインなどのモニタリングデータや、医用画像と診断結果のデータなど、膨大な量のデータから特徴を抽出したり、分類したりするための、統計学やAIを基盤とする一連の手法である。本演習では、データサイエンスのプログラムについて、具体的な仕様の決定からプログラムの設計・作成・実装までのプログラミング技術修得を目指す。

授業の到達目標

ツールを用いて適切にデータの管理・解析を行うことができるようになること、プログラムを使ってデータ解析を行うための過程（要件定義・外部設計・詳細設計・実装）を理解し、自ら実装できる技術を修得すること、膨大なデータを適切に統計処理でできること。

授業の概要と計画

①授業形態

対面で行う。

②授業の概要と計画

サンプルデータに対し、R、Python、SQLなどのツールを用いた管理・解析を行う。データサイエンスの基礎（基礎的なツールの使い方、データの入出力、機械学習を含めた各種アルゴリズムの仕組み）と大量のデータ（時系列データ、画像データ）を解析するためのプログラムの実装方法を講義・演習を通して学習する。

第1回 インTRODクシヨン、Python、Rの基礎

第2回 データ前処理（データ収集、Webスクレイピング、データクレンジング、データベース技術）

第3回 確率統計（Python: Numpy, Pandas, Matplotlib等）、基本統計量、t検定、 χ^2 検定

第4回 アルゴリズム組合せ爆発、探索問題、回帰分析 最乗法、単回帰分析、重回帰分析

第5回 分類問題 決定、アンサンブル学習、クラスタリング問題、階層的クラスタリング、K-means法

第6回 連関分析、K近傍法、SVM、主成分分析、モデル選択

第7回 時系列データ分析、然語処理

第8回 画像処理、描き字識別

成績評価方法

授業中に提示する課題に対する演習結果を、作成したプログラムを含め、まとめたレポート（複数回）により評価（100%）

成績評価基準

- ・プログラムのソースコードやフローチャートにて自身のプログラムを説明できているか
- ・適切な体裁のグラフや図表により解析結果をまとめられているか
- ・結果に対して考察を行い、課題抽出ができていないか

履修上の注意（関連科目情報）

- ・「医療機器・システム設計概論」、「医療機器・システム設計演習」、「計測技術概論」とリンクさせて学習することを推奨する。
- ・演習は学生個人のPCを利用して行う。

事前・事後学修

【予習】 次回の演習内容について、あらかじめ演習の手順を確認しておくこと。関連する講義と演習の内容を再確認しておくこと。

【復習】 各課題に対する演習終了後のレポート提出に向けて結果をまとめ、各レポート課題を完成させること。
本学では1単位あたりの学修時間を45時間としています。毎回の授業にあわせて事前学修・事後学修を行ってください。

オフィスアワー・連絡先

基本的に、各授業の前後で時間をとりますが、メールで随時質問を受け付けます。

連絡先：中楯龍

E-mail:nakadate@people.kobe-u.ac.jp

学生へのメッセージ

不明な部分がある場合は躊躇せずに質問すること。

今年度の工夫

教科書

BEEF+で資料を配布します。

参考書・参考資料等

授業中及びBEEF+で指示します。

授業における使用言語

日本語

キーワード

データ解析、プログラミング、分類、機械学習