

授業科目名	データサイエンス演習	担当教員	藤本 悠
必修の区分	選択		
単位数	1 単位		
授業の方法	演習		
開講年次	3 年 第 1 クォーター		
講義内容	<p>本授業は高度な情報技術を用いて情報を分析し、意思決定支援を行うための技能を習得することを目的とする授業であり、データを通して実世界における様々な課題を解決する方法について学ぶ。</p> <p>具体的には、データベースの設計と実装、SQL によるデータベースへの問合せ、多次元データ解析、ネットワーク分析、自然言語処理、地理空間データ解析の方法の基礎について学ぶ。また、一連の技能習得に加えて、プログラミングに関する基礎的技能やデータの可視化手法の習得も目指す。</p>		
到達目標	<p>本授業を通してはデータサイエンティストとして最低限のスキルとして以下の基本技能を習得することを目標とする。</p> <ul style="list-style-type: none"> <li>・ 他のプログラミング言語から RDBMS に接続できる</li> <li>・ 基本統計量について説明することができる</li> <li>・ 相関係数行列を計算し、出力結果の意味を適切に理解することができる</li> <li>・ 重回帰分析の出力結果の意味を適切に理解することができる</li> <li>・ クラスタ分析によってデータを分類することができる</li> <li>・ ネットワークを構築し、ネットワーク中心性を計算することができる</li> <li>・ 文章から単語を切り出し、単語ごとの頻出傾向を分析することができる</li> <li>・ 点分布を分析し、出力結果を適切に理解することができる</li> <li>・ オーバーレイ分析によって地理的事象を読み取ることができる</li> </ul>		
授業計画	<ol style="list-style-type: none"> <li>1. RDBMS の Dump と Restore の機能 → データベースの再構築</li> <li>2. RDBMS への接続方法 → R の基本操作と RDBMS への接続</li> <li>3. スクリプト言語上での SQL の実行 → R での SQL の実行とデータ取得</li> <li>4. 多次元データの構造 → 基本統計量の計算</li> <li>5. 分散共分散行列と相関係数の検定 → 相関係数の計算と T 検定</li> <li>6. 予測モデルと変数選択 → 重回帰分析とステップワイズ法</li> <li>7. クラスタ分析と分類基数 → 階層的分類と非階層的分類</li> <li>8. グラフ理論とネットワーク分析 → 中心性の計算</li> <li>9. テキストマイニングの基礎 → ワードクラウドの構築</li> <li>10. 空間分析手法と地理空間表現 → カーネル密度推定と空間分割</li> <li>11. 空間検索と域内集計を行う → オーバーレイとジオプロセッシング</li> <li>12. 分析レポートの作成 → 分析レポートの作成</li> </ol>		
事前・事後学習	<p>本授業では知識と技能を身につけることを目的としている。したがって、授業中に紹介された技術については必ず実践し、自分のスキルとして身につけるように心がけること。また、毎週のレポート課題を通して、授業で得られた知識をしっかりと振り返ること。</p>		
テキスト	特に指定しない		
参考文献	適宜紹介する		

成績評価の基準	成績については取り組み姿勢や課題の提出状況。授業貢献度などを総合的に判断する。
履修上の注意 履修要件	履修希望者が5名未満の場合には開講しない場合がある。
実践的教育	該当しない。
備考欄	ICTの流行は非常に速いスピードで変化するため、シラバスの内容に関わらず、旬となる情報については積極的に取り入れるので、授業内容の順番や内容そのものが変更となる可能性がある。また、受講人数によっては内容を受講生の興味、技術力、必要性の高さなどに合わせて変更する可能性がある。