

# 2022年度Q 1 データサイエンス特論

## Lecture1: 講義ガイダンス - データサイエンス概観、logstics -

# 自己紹介

- 名前: 岩政 幹人 (いわまさ みきと)
- 現所属: (株)東芝 研究開発センター システム知能情報ラボラトリ
  - 機械学習の研究やビジネス応用を担当
- 経歴 :
  - 第五世代コンピュータプロジェクト (1989-1992)
  - スタンフォード大学客員研究員(知識科学、1994-1996)
  - スピンオフベンチャー出向(LSI高位設計技術,2000頃)

- 経歴(つづき)
  - ソフトウェア高信頼化
    - 形式的手法（北陸先端科学技術大学博士、情報科学）
  - 「AI白書2019-2020」編集(IPA出向、2018-2020)
  - 社内AI品質・ガバナンス
- 趣味
  - バイク

# コース概要:講義の目標

- データサイエンス領域を体系的に概観する
- 得られるもの:
  - 実務の入口としての、手法や、プログラミングの参照提供
  - (できれば) より深い研究テーマへの発展への糸口
- 目的ではないもの:
  - プログラミング(Python)を学ぶ
  -

# 講義の内容

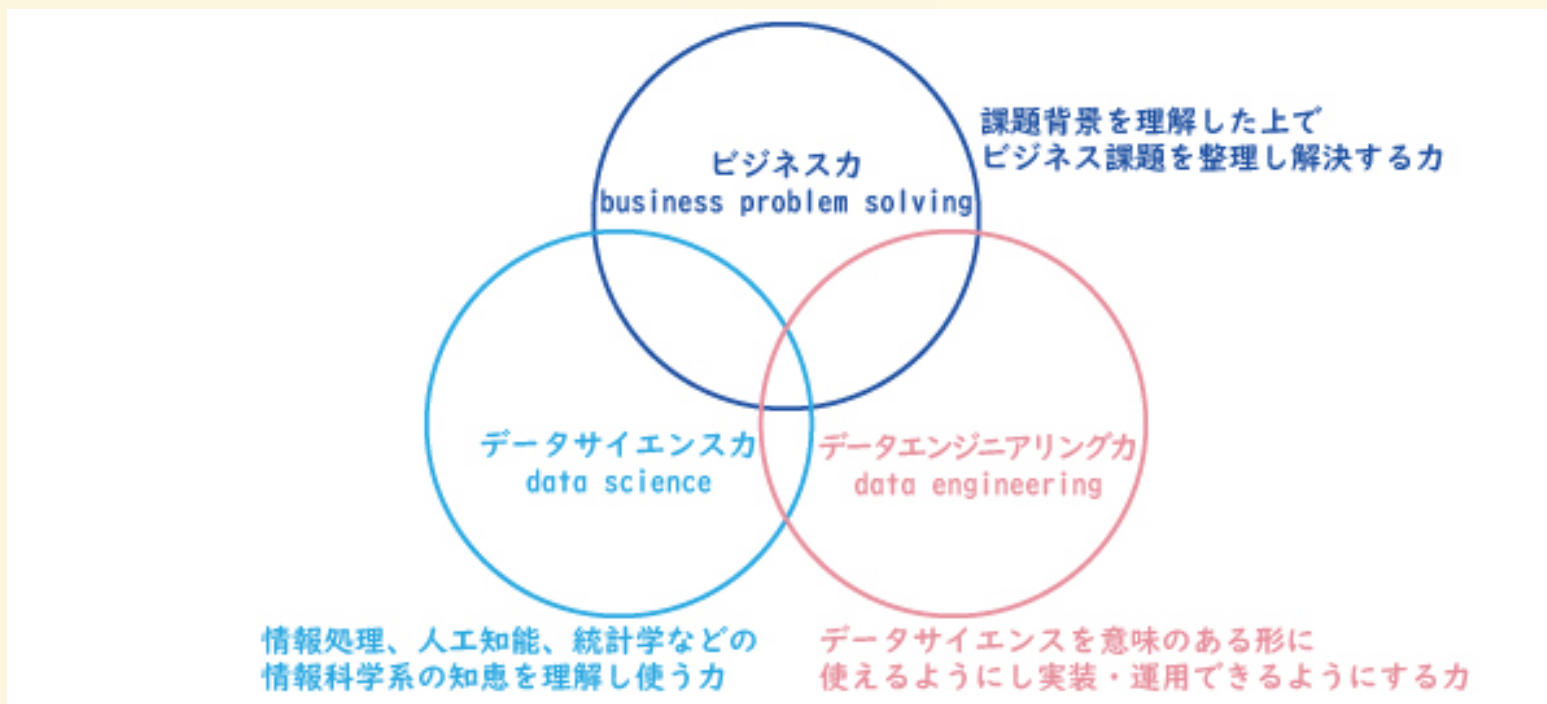
- データの取得・操作・視覚化(Lecture1-3)
- 統計分析の基礎(Lecture4)
- 回帰分析(Lecture5-6)
- パターン認識(Lecture7-8)
- 時系列・動的システム・周波数分析(Lecture9-11)
- 画像処理(Lecture12)
- メディア情報処理 (Lecture13,大久保先生)
- 位置情報処理 (Lecture14,大久保先生)
- センサデータ処理 (Lecture15,大久保先生)

# 第1回

- データサイエンスとは(short)
- 授業の内容（目次）
- 講義の進め方
- テストなど

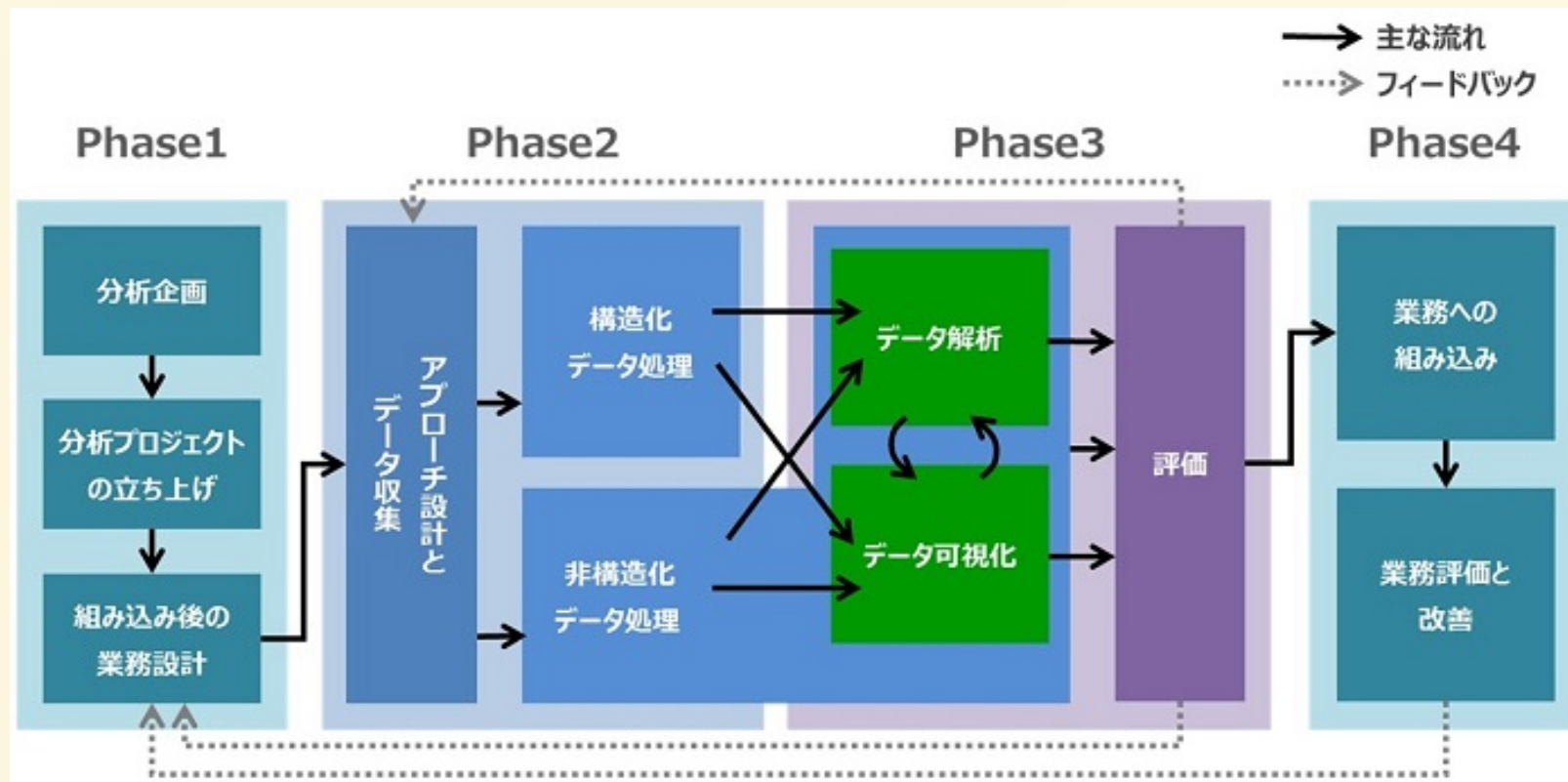
# データサイエンティストとは何か

- データサイエンティスト協会の定義
  - 「データサイエンス力、データエンジニアリング力をベースにデータから価値を創出し、ビジネス課題に答えを出すプロフェッショナル」



- スキルチェック by IPA(「参考」にURL)
  - 「企業等の業務において大量データを分析し、その分析結果を活用するための一連のタスクとそのために習得しておくべきスキル」

◦





# データサイエンティストに対する期待：産業界

- 「DX時代に求められる技術者育成施策－日立におけるデータサイエンティスト育成の事例を元に－」(2020情報処理、日立)

	データサイエンス力 情報処理，人工知能，統計学 などの情報科学系の知恵を 理解し，使う力	ビジネス力 課題背景を理解した上で ビジネス課題を整理し， 解決する力	データ エンジニアリング力 データサイエンスを 意味のある形に使えるようにし， 実装，運用できるようにする力
事業創造 レベル (Level6)	どのようなデータ（非線形， 疎など）でも解ける	データを駆使し，かつ実現性 のある新事業案を出せる	どのようなデータの量・ 形式の組合せにも 対応して実装できる
業務改革 レベル (Level5)	高度な機械学習の 手法を使って解ける	業務プロセス全体での 課題を定義できる	非構造データになっても， 分析できる環境を作り， 実装できる
作業改善 レベル (Level4)	多変量解析を使って解ける	個人作業範囲での 課題を定義できる	ツールだけでは難しくなった ときに，ライブラリ・プログラ ミングを活用して実装できる
ユーザ レベル (Level3)	可視化して考察できる 回帰分析を実行できる	定義された課題を 理解できる	手持ちの環境 (EXCEL，ツールなど) を 使って実装できる
導入 レベル (Level2)	データサイエンスを使うことで，社会，事業，仕事をより良くできると理解している		

# データサイエンス領域の概観： Data Science Landscape

- <https://github.com/dataprofessor/infographic>
- Programmingを中心として機械学習、可視化、統計、データ処理、ソフトウェア工学、数学の枝が伸びる
  - 機械学習のアルゴリズムの一つにニューラルネットワークをつかうものがあり、、、
  - 数学には、線形代数や確率理論や最適化問題がある

(c)2022 岩政幹人



# データサイエンスの領域についての概観

- 何をするのかを明確化する
- 基本方針を立てる
- データを集める
- データを処理する・加工する
- データをつかってモデルを作る
- モデルの妥当性の検証
- ちゃんと何をするのかが満たされているかを検証

# いろいろなデータ

- 分類データ
  - 鳥の分類、癌の種類、当
- 統計データ
  - 国勢調査、アンケート集計、等
- 画像データ
  - カメラ画像、Computer Graphics、等
- 音声データ
  - 音楽、合成音声、
- 文字データ
  - 書籍、報告書、議事録

# 機械学習の進化とデータサイエンス

- 古典的な機械学習
- 高次元の科学
- 深層学習の登場

# データ解釈の罠

- 再現性の問題
- HARKingとp-hacking
- 認知バイアス

# データ利活用の罫

- 状況・目的によって方法が異なる
  - 理解か予測の絵

## 参考文献

- データサイエンティスト協会
- IPAスキルチェック
  - <https://www.ipa.go.jp/jinzai/itss/itssplus.html>
- 日立
  - <https://www.ipsj.or.jp/dp/contents/publication/41/S1101-T04.html>
- 「AI白書2019」
  - [https://www.ipa.go.jp/ikc/our\\_activities/rs\\_01.html](https://www.ipa.go.jp/ikc/our_activities/rs_01.html)