



DataScientist Society

データサイエンティスト育成講座

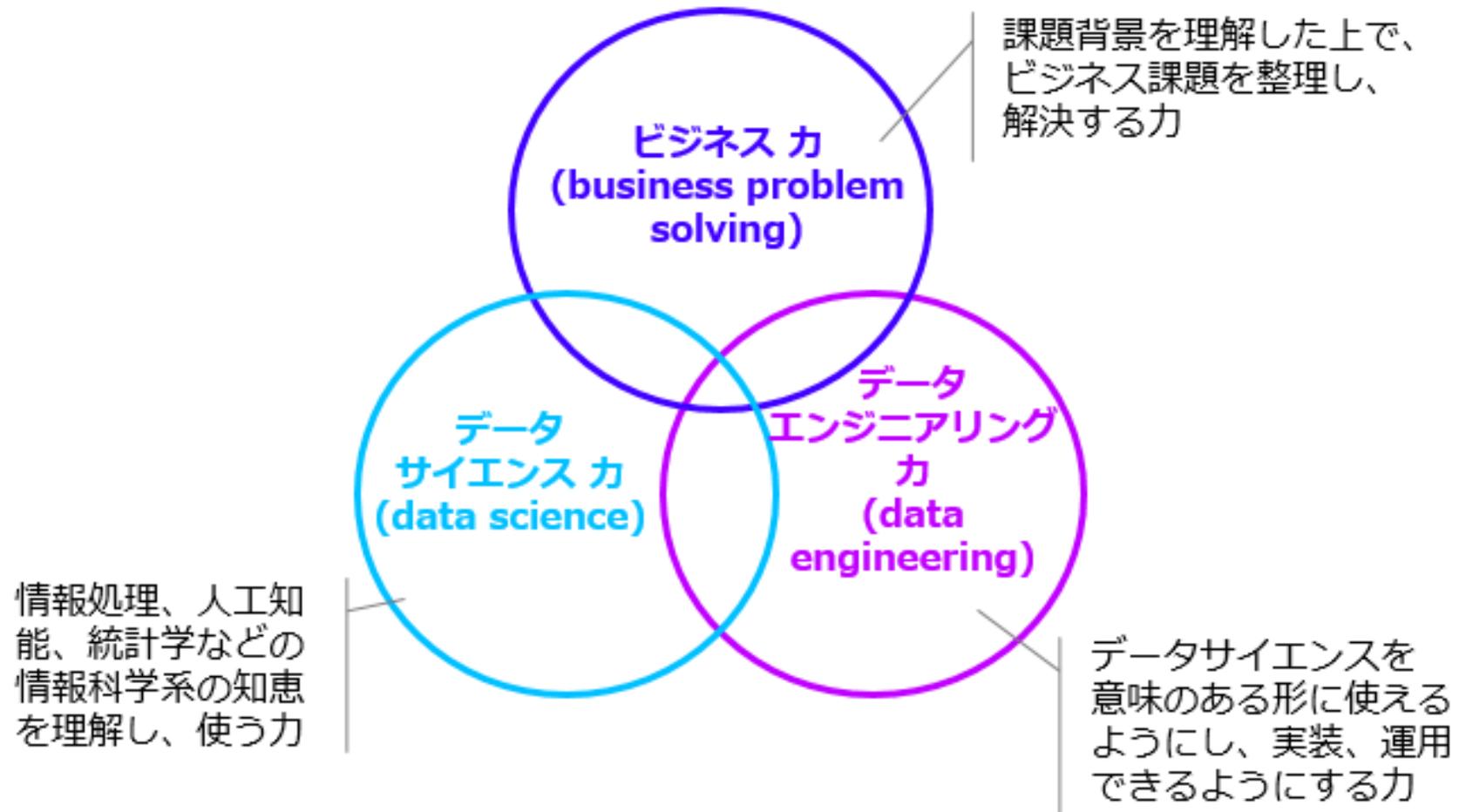
この講座の目標とオーバービュー

電気通信大学

斉藤 史朗

データサイエンティストに要求されるスキル

- データサイエンティスト協会では、以下のように定義しています



分析することと、モデリングの関係

- 「分析」とは人により様々なレベルを指すことがあります
 - ① Descriptive Analytics
 - ✓ 現在までのデータを集約加工して、過去の動き、現状を知る
 - ✓ 「分析」というより「集計」が中心、BIツールなどの適用
 - ② Predictive Analytics
 - ✓ 「予測する」ことが主目的となる(現況の把握は前提事項)
 - ✓ 各種のモデル構築手法を適用する
 - ③ Prescriptive Analytics
 - ✓ ビジネスにとって適切な戦略を「処方する」ことが主目的となる
 - ✓ モデルを使わないと浪花節になる

本講座の目的と「やらないこと」

- 今回の講座の目的はPredictive Analyticsを正確に行える「サイエンス力」のスキルを身に付けることが最大の目的です
 - ②Predictive Analytics
 - ✓ 「予測する」ことを可能とするモデルを構築します。
 - ✓ 「予測」において重要な判別モデル（「0、1を当てる系」）と回帰モデル（「数を当てる系」）の二つの種類のモデル作成の基礎を学びます。
 - 機械学習を3分して教師あり学習、教師なし学習、強化学習とするものをよくみますが、大元はビショップのPRMLのp.3ですが、よく注意する必要があります。
 - なので、エンジニアリング力とかビジネス力を教えることを目的とはしていません。
 - ✓ 実際にビジネスに使えるかどうかは別として、一旦は技術として「追い込む」スキルを身につけましょう。身につけたスキルを使うか使わないかは選択できますが、身につけていないスキルは必要になっても使うことができません！

グループワークのやり方など

- **基本的な事項は全員でやっていただきます**

- ✓ 実際にpythonでモデルを作るやり方や、パラメーター設定など、自分でやらないと身に付かないものは、自分でやりましょう。

- **グループワークの効用**

- ✓ アイデアの交換：特徴量（変数）を作る場合に、分析者本人のいろいろな物の見方により、違ったアプローチが可能です。他人のアプローチはとても参考になります。
- ✓ 技術の教え合い：コーディングなどは、得意不得意があると思います。得意な方は不得意な方に教えてあげてください。（もちろん、講師に質問もしてください）教えることが上達の一番の近道です。
- ✓ モデル作成の最後の追い込み：ここでは作業を分担すると効率的です。それぞれの得意な分野で協力して精度アップをはかりましょう。

本講座のねらい

- **データサイエンティストとして活動する上での必須スキルを身につけていただきます**
 - ✓ 先端のこと、「流行っている」ことよりも基本を教えます。
 - ✓ 「調べ方」もお教えしますので、先端のことは自分で調べてやってみてください。それができるような基礎をお教えします。
- **データサイエンティストとして必要なこと**
 - ✓ データをよく見ること
 - 質の良いデータサイエンス業務のエッセンスはこれに尽きます。
 - DSは決して「派手な」職業ではありません。地味な職人的な作業が大半を占めます。
 - ✓ アルゴリズムの性格をよく知ること
 - ただ実装できる、結果が出ましたではなく、そのアルゴリズムがどういう性質のもので、出てきた結果にはどういう特徴（偏り）があるのかを理解しておくこと。
 - ✓ 現在のデータサイエンス作業の多くは、数年内に自動化が可能です。
 - 人間でなければできないことに注目しましょう！