



データサイエンティスト育成プログラム カリキュラム

株式会社データミックス

2017年5月版

プログラムのゴールと特徴

プログラムのゴール

データ分析職(データサイエンティストでなくとも)に求められる高度な統計学や機械学習の理論とスキルを体系立てて理解し、ビジネス課題に適用できる人材の育成を目指します。具体的には以下のことができるようになることを目指します。

- 高度なデータ分析技術を用いてビジネスの課題の解決ができる
- 「なぜこの分析手法を使うのか」を技術的な側面とビジネスの側面から判断できる
- PythonやRを使って求められている課題に対し適切な分析手法を実行できる(やってはいけないことも理解している)
- 分析結果をクライアント(社内・社外問わず)にわかりやすく説明できる

特徴

- 体系立てて学習することで、データ分析技術の引き出しを増やします
- 小規模クラスのため、クラス内での議論や講師とのやりとりを多くすることで内容の理解を深めることができます
- 理解を深めるためのクイズやハンズオンでの演習(宿題含む)を数多く実施し、理解するだけでなく手を動かせることを目標にします。

プログラムの想定受講者

本プログラムは以下のような受講者を想定しています。

キャリアの指向性

- ビジネス課題を解決することに興味がある方(ビジネス課題を解決するための道具として高度なデータ分析手法を駆使したい方)

数学分野の条件

原則、数式にアレルギーがなければ受講可能です。

<望ましい条件>

- 高校または大学で微分・積分、線形代数の授業を受講したことがある方
- 文系・理系問わず大学で定量系の研究室・ゼミに所属していた方(金融・経済、心理学、商学系などデータ分析に関係していれば問題ありません)

プログラミングの経験

原則、不要です。ただし、初学者の方は学習開始時は宿題に週20時間ほどの時間がかかる場合がありますので、お時間の確保をお願いします。

<望ましい条件>

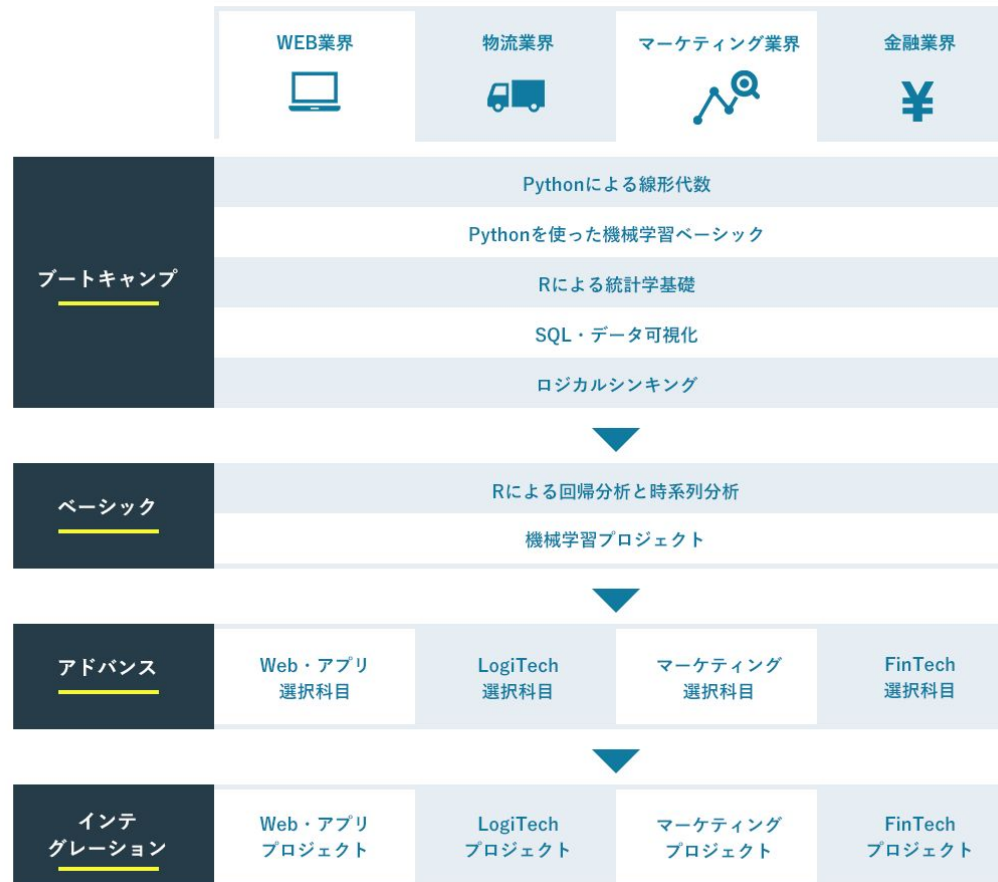
- Python, R, C, Java, Rubyなどのプログラミング言語を業務または個人プロジェクトで触ったことがある方

受講条件

- ノートPCは各受講者で用意できること
 - どうしても用意できない場合はご相談ください
- 持参するノートPCにPythonやRなどソフトウェアをインストールできること
- また、必要に応じて各ソフトウェアのライブラリをインストールできること
- 復習や宿題を実施する時間をとることができること
 - 目安として講義とは別に週に5~10時間程度の時間を確保できること

カリキュラムの全体像

大きく4つのステップに分かれています。「ブートキャンプ」「ベーシック」の2つのステップは業界に関わらず必要なデータサイエンスの基礎を学びます。後半の「アドバンス」「インテグレーション」ではそれぞれの業界でより求められるスキルを優先的に学びます。



ブートキャンプステップ

Bootcamp step

ブートキャンプステップの構成

コース	概要	主なトピック
Pythonによる 線形代数レビュー (講義: 3時間)	基本的なPythonでのコーディングを習得することに加え、機械学習や統計モデルを理解するために必要最低限の線形代数をPythonを使いながら学習します。	<ul style="list-style-type: none">● プログラミング基礎● ベクトル・行列● NumPy● 行列分解
Rによる統計学入門 (講義: 3時間)	R言語を用いて、統計学の基本的な考え方と確率モデルについて学びます。今後の学習をスムーズにするため理論と実践をバランスよく行います。	<ul style="list-style-type: none">● 確率分布● 最尤推定● 点推定、区間推定● 仮説検定
Pythonによる 機械学習ベーシック (講義: 3時間)	Pythonの中でもデータ分析で用いられるNumPy, pandas, scikit-learnといったライブラリの使い方を学習します。また、機械学習の中でも教師あり学習を中心に理論を学びます。	<ul style="list-style-type: none">● pandasを用いたデータ加工● 教師あり学習● 勾配降下法
SQLによるデータ加工とデータ可視化 (講義: 3時間)	大量データを扱ううえではデータベースを思い通りにできるようになることが重要です。数千万レコードのデータをSQLを用いて集計・加工するスキルを学習します。また結果を可視化する方法を学びます。	<ul style="list-style-type: none">● SQL● データ可視化● ダッシュボード
データ分析のための ロジカルシンキング (講義: 3時間)	本コースではこのビジネス課題から分析課題への落とし込むための定量的なロジカルシンキングとデータプレゼンテーションをケースを通じて学習します。	<ul style="list-style-type: none">● KPI● ロジックツリー● 分析結果から施策の提案● データプレゼンテーション

ブートキャンプステップの進め方

ブートキャンプステップは基礎的な知識・スキルを短期間で獲得するため講義は毎週行われます。また「わかったつもり」にならないよう毎週、コーディングを含めた課題が出されます。その課題を解くことにより手を動かす力を身につけます。

Week1	Pythonによる 線形代数	個人課題	課題提出日
Week2	Rによる 統計学入門	個人課題	課題提出日
Week3	Pythonによる 機械学習 ベーシック	個人課題	課題提出日
Week4	SQLによる データ加工と データ可視化	個人課題	課題提出日
Week5	ロジカル シンキング		

ベーシックステップ

Basic step

ベーシックステップの構成

コース

概要

主なトピック

Rによる 統計モデリング

レクチャー
(3時間)

統計モデルの中では一般的な線形回帰モデルを含む一般化線形モデルを学習します。また、モデルの推定に伴う検定やモデル選択、モデル評価を学習します。

プロジェクト

約2週間かけて以下の3つのテーマに取り組みます（具体的なプロジェクトは変更になる可能性があります）。

- 線形回帰モデルを用いたモデル構築
- 線形回帰モデルのチェック
- ロジスティック回帰を用いたモデル構築

- 線形回帰モデル
- 回帰係数のt検定
- F検定
- 決定係数
- AIC
- モデル選択
- ロジスティック回帰モデル
- オッズ比
- Z検定
- 回帰診断

機械学習 プロジェクト

レクチャー
(3時間)

教師あり学習と教師なし学習のアルゴリズムを速習します（事前にリーディング資料をお渡しします）。また、PythonとRでの実行方法や気をつけるべきことをお伝えします。

プロジェクト

約2週間かけて以下の2つのテーマに取り組みます。

- 予測モデル構築
- クラスタリング

- Lasso, Ridge, ElasticNet
- SVM
- 決定木
- アンサンブル学習
- ニューラルネットワーク
- KMeans法
- 階層型クラスタリング
- スペクトラルクラスタリング
- クロスバリデーション
- パラメータチューニング
- 特徴量作成・選択

ベーシックステップの進め方

ベーシックステップは統計モデリングと機械学習を中心に実践スキルを鍛えていきます。それぞれのコースで、レクチャーを3時間、その後約2週間かけてグループワークと個人プロジェクトに取り組めます。さらに2週間後に取り組んだ課題の成果を各自発表することでデータプレゼンテーションのスキルを高めます。



アドバンスステップ

Advance step

アドバンスステップの構成

トラック名	主な対象者・対象業界	推奨選択コース	コース概要
AdTech トラック	モバイルアプリや B to CのWebサービス、 アドテクサービスを 展開する企業での データサイエンティストを 目指す方	レコメンデーション アルゴリズム	クリックやコンバージョンを高めるための高度なレコメンデーションアルゴリズムを学びます。
		自然言語処理とAPI	記事のクローリングや収集した記事の処理(自然言語処理)、それをサービス化するためのAPI作成について学びます。
		BigData (Spark)	ログデータなど数億単位の大規模データを扱うためのフレームワーク(Spark)の使い方を学びます。
LogiTech トラック	機械学習・人工知能を用いて、これまでにない物流業界におけるデータサイエンティストを目指す方	組み合わせ最適化と動的計画法	組み合わせ最適化、動的計画法など物流倉庫や配送など物流は最適化問題に取り組むスキルを身につけます。
		需要予測(時系列分析)	ARIMA、VAR、状態空間モデルなど時系列分析を学習し需要予測を行うアルゴリズムを学びます。
		Deep Learning	画像認識(CNN)と時系列分析(RNN)をディープラーニングを用いて行う方法を学習し、複雑な課題への対応力を高めます。
Digital Marketing トラック	デジタルマーケティング、マーケティングサイエンスを強みとしたデータサイエンティストを目指す方	統計的因果推論と実験デザイン	施策の効果を検証したり、意思決定を支援するために実験デザイン、統計的因果推論を学習します。
		マーケティングミックスモデリング	広告予算の最適化のためのマーケティングミックスモデリングを学習します。
		ベイズモデリング入門	消費者行動を理解したり、one to one マーケティングなどで使われたりする、ベイズモデリング手法を学びます。

※ 金融業界向けのコース(FinTechトラック)は現在準備中

アドバンスステップの進め方

アドバンスステップはベーシックステップと同様、それぞれのコースで、レクチャーを3時間、その後約2週間かけてグループワークと個人プロジェクトに取り組みます。また、2週間後に取り組んだ課題の成果を各自発表することでデータプレゼンテーションのスキルを高めます。

Week1	選択コース1 レクチャー	グループワーク + 個人課題			
Week2		グループワーク + 個人課題			
Week3	選択コース1 成果発表	事前予習課題(個人課題)			
Week4	選択コース2 レクチャー	グループワーク + 個人課題			
Week5		グループワーク + 個人課題			
Week6	選択コース2 成果発表	事前予習課題(個人課題)			
Week7	選択コース3 レクチャー	グループワーク + 個人課題			
Week8		グループワーク + 個人課題			
Week9	選択コース3 成果発表				

インテグレーションステップ

Integration step

これまで学習したスキルを統合する インテグレーションステップ

インテグレーションステップのゴール

約1ヶ月間かけてビジネス課題をデータサイエンスを用いて解決するためのプロジェクトを行います。具体的には、分析計画の策定、実際の分析作業、結果のレポートまでの一連の流れを行います。また、毎週個別メンタリングを通じて進捗確認と技術的なサポートを行いながら、より実務で使うテクニックの習得を行います。

本コースの受講後にできること

- ビジネス課題を分析課題へ「翻訳」しプロジェクトに仕立てることができる
- 分析計画を策定することができる
- データの抽出、前処理、分析、アルゴリズム選択、評価など一連の作業を一人で実行できる
- オーディエンスに合わせたデリバリーができる
- 新しいスキルを学び続ける方法を身につける

本コースで使用するソフトウェア：R, Pythonなど何でも構いません

インテグレーションステップの進め方：

Week1	プロジェクト発表			個人プロジェクト			
Week2	個別メンタリング			個人プロジェクト			
Week3	個別メンタリング			個人プロジェクト			
Week4	成果発表 プレゼンテーション						

本講座・カリキュラムについてのお問い合わせ



株式会社データミックス info@datamix.co.jp