



# 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

2018年4月版

# プログラムのゴールと特徴

## プログラムのゴール

- 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
- 生徒数に対して講師とTAを適切に配置しますので、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(1)

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)

## プログラムの想定受講者(2)

本プログラムは以下のような受講者を想定しています。

### 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。入学に際して簡単な「数学の試験」を実施いたします。

\* 入学までに指定される予習を実施していただきます。

### プログラミングの経験

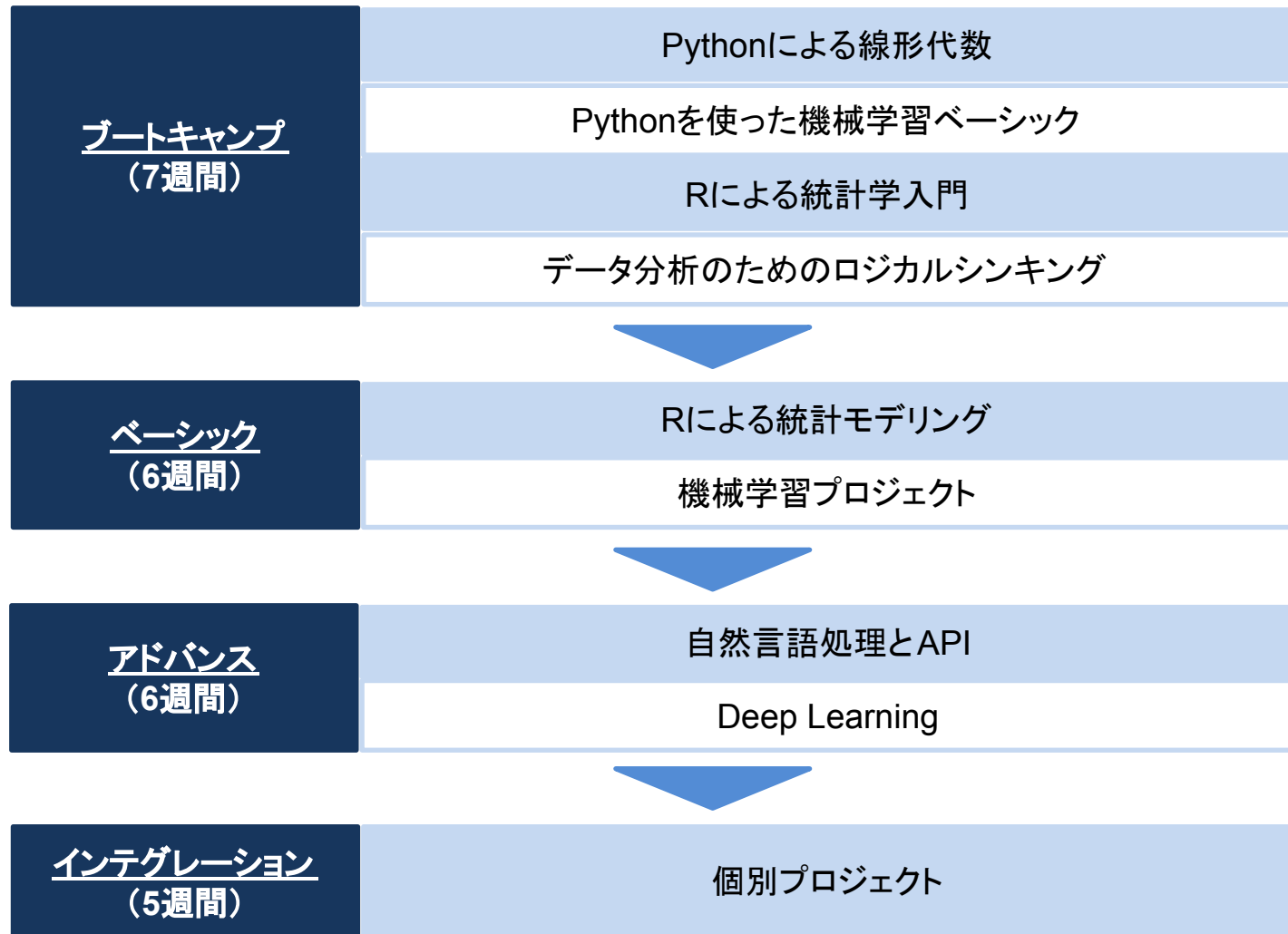
現時点で経験の必要はありませんが、入学に際して簡単な「コーディングの試験」を実施いたします。

また、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。

## 受講条件

- ノートPCは各受講者で用意できること
- 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



# データサイエンティスト育成コース 修了条件

- インテグレーションステップまで受講していること
- インテグレーションステップにて、勤務先企業や社会的課題等を分析課題に翻訳し、分析計画を立てられること
- 分析計画に沿って、時間内にデータサイエンス技術を用いて課題解決を行えること。
- 上記を示すレポートの提出または発表会にて外部専門家を含むオーディエンスの前で20分間の発表および10分の質疑応答を行っていただくことで修了と皆します。

# 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 (講義: 2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。ビジネスでも使えるページランクアルゴリズムを学習し、線形代数のビジネス適用シーンも学習します。	<ul style="list-style-type: none"><li>● プログラミング基礎</li><li>● ベクトル・行列</li><li>● NumPy</li><li>● 行列分解</li><li>● PageRank</li></ul>
Pythonによる機械学習ベーシック (講義: 3時間 x 2日)	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。また決定木、ランダムフォレストなどの機械学習アルゴリズムの概要を理解します。	<ul style="list-style-type: none"><li>● 勾配降下法, SGD</li><li>● pandasを用いたデータ加工</li><li>● 線形回帰分析、決定木、ランダムフォレスト</li><li>● クラスタリング</li><li>● ハイパーパラメーター</li></ul>
Rによる統計学入門 (講義: 3時間 x 2日)	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。特に、確率分布、最尤推定、仮説検定といった統計学の基礎に始まり、ABテストの策定などで必須となるサンプルサイズの決め方までを学習します。	<ul style="list-style-type: none"><li>● 確率分布</li><li>● 最尤推定</li><li>● 点推定、区間推定</li><li>● 仮説検定</li><li>● サンプルサイズの決め方</li></ul>
データ分析のためのロジカルシンキング (講義: 3時間 x 2日)	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、KPIツリーの作り方を学習します。その後、予測モデリング、ABテストの策定方法をケーススタディを通じて勉強していきます。	<ul style="list-style-type: none"><li>● KPI</li><li>● ロジックツリー</li><li>● 分析結果から施策の提案</li><li>● 予測モデリング</li><li>● ABテストの策定</li></ul>

# 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	Pythonによる 機械学習 ベーシック1	個人課題	課題提出日
Week3	Pythonによる 機械学習 ベーシック2	個人課題	課題提出日
Week4	Rによる 統計学入門1	個人課題	課題提出日
Week5	Rによる 統計学入門2	個人課題	課題提出日
Week6	ロジカル シンキング1	個人課題	課題提出日
Week7	ロジカル シンキング2	個人課題	課題提出日

# 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  
×2回)

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  
(3時間  
×2回)

課題解決型学習(PBL)スタイルで、クラスの中で以下の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
- 一般化線形モデル

### 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  
(3時間)

課題解決型学習(PBL)スタイルで、クラスの中で以下の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。統計モデリングはレクチャーとクラス内でデータ分析のケースを実践的に解いていく授業(プロジェクト)を2つのテーマに分けて各2週間ずつ2回実施します。また、機械学習についてはレクチャーとクラス内で実施するプロジェクトを通して2週間かけて学びます。それぞれのプロジェクトでは、各自成果を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	RIによる 統計モデリング レクチャー1	個人課題				課題提出日
Week2	RIによる 統計モデリング レクチャー2	個人課題				課題提出日
Week3	RIによる 統計モデリング PBL1	個人課題				課題提出日
Week4	RIによる 統計モデリング PBL2					
Week5	機械学習 プロジェクト レクチャー	個人課題				課題提出日
Week6	機械学習 プロジェクト PBL					

## 過去のプロジェクト

- Walmartの店舗別・商品カテゴリ別  
・週別の売上予測
- AirBnBの初回ユーザーの行き先予測

# アドバンスステップ

Advance step

# アドバンスステップの構成

## コース

## 概要

自然言語処理とAPI	レクチャー (3時間 ×2回)	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)、それをサービスにするためのAPI作成について学びます。  学んだスキルと知識を使ってプロジェクトに取り組みます。
	プロジェクト (3時間)	
Deep Learning	レクチャー (3時間 ×2回)	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。  学んだスキルと知識を使ってプロジェクトに取り組みます。
	プロジェクト (3時間)	



# アドバンスステップの進め方

アドバンスステップはレクチャーを 3時間×2週間実施し、その後より実践的なスキルを身につけるためにプロジェクトに取り組む3時間の授業を実施します。ベーシックステップ同様、取り組んだプロジェクトの結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	自然言語処理とAPI レクチャー 1	個人課題			課題提出日
Week2	自然言語処理とAPI レクチャー 2	個人課題			課題提出日
Week3	自然言語処理とAPI プロジェクト				
Week4	Deep Learning レクチャー 1	個人課題			課題提出日
Week5	Deep Learning レクチャー 2	個人課題			課題提出日
Week6	Deep Learning プロジェクト				

# 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。

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	個別メンタリング	個人プロジェクト			課題提出日
Week5	成果発表 プレゼンテーション				

# 過去の卒業生のテーマ

## テーマ

トレイルコースのクラスタリングとレコメンデーション

Azure ML 異常検知

Deep Learningの金融時系列データ予測への活用

Deep Learningを用いた自動キャプション生成と金融業界への応用

旅行会社のログデータを活用した顧客クラスタリング

自然言語処理の活用「トレンド型商品訴求の自動化」

探索的データ分析によるペルソナの理解と予測モデルの活用

# 受講に関してよくある質問

## FAQ

# 正直なところ予習・復習のワークロードについて

**Q:学習時間はどのくらい確保したらよいですか？**

A: 講義への出席以外にも予習・復習、課題プロジェクトのための時間として1週間で5~10時間程度(1日1~2時間程度)の自習時間の確保をお願いしております。特に、プログラミング初心者の方は、開講直後は10時間以上の予習・復習時間が必要になった例もあります。

## 振替授業・欠席した場合の対応

**Q:申し込んだ曜日のクラスに出席できない場合、どうしたらよいですか？**

A: 同じ内容の授業を同じ週の他の曜日に開講しておりますので振替受講していただくことが可能です。ただし、各クラス受講人数の定員がありますので、満席の際および振替授業への出席も出来ない場合は、授業を録画したeラーニングをご覧いただくこととなります。また、次回ご来校時にテキストをお渡しします。