

ビジネス現場で活躍する

# Python

## 分析と機械学習

中井 研 著

**お勉強で終わらない。仕事で使えるデータ分析プログラミング。**

こんなことができるようになります

- Pandasで簡単デモグラフィック分析
- matplotlibで高速グラフ作成
- scikit-learnとkerasで機械学習に挑戦



# ビジネス現場で活躍する Python分析と機械学習

---

Machine Learning for Bussiness with python

中井 研 著



本文中の製品名は、一般に各社の登録商標、商標、または商品名です。  
本文中ではTM、®、© マークは省略しています。

# はじめに

昨今の機械学習やAIの普及はめざましく、様々な業種・分野に活用されています。また機械学習の専門家やデータサイエンティスト、エンジニア以外でもプログラミングをせずにAIを使える環境も登場しています。しかし、現実の複雑な課題や要件を満たすAIを開発するためには、プログラミングの意識と技術が必要な段階だと私は考えています。そこで、その機械学習プログラミングをより多くの人、特に普段はプログラミングをしないような人にも使えるようになる内容を目指してこの本を執筆しました。

対象はもちろんプログラミング未経験の方から、プログラミングは知っているけれどもPythonで機械学習プログラミングをしたことがない方まで広く対象としています。また、他の書籍や無料講座などで学習をしたけれども、いまいち理解できなかった、業務に繋がられなかった方も大歓迎です。本書では、図説を用いながら段階をコード実行し理解を深めていき、実際の業務に近い課題設定で例題をご紹介します。

機械学習を活用できる場面はこの世の中にまだまだ存在すると思っています。使えば使う分だけ、機械学習の面白さや奥深さを知ることになると思います。機械学習プログラミングを学ぶことで、分析業務を効率的に改善したり、より精度の高い分析結果を導出できることを願っています。

2019年初夏

中井 研

## 本書の表記法

本書では、次に示すように文字の書体を使い分けている。

### ゴシック(サンプル)

新しい用語を示す。

### 固定幅(sample)

プログラマリストに使うほか、本文中でも変数、関数、データ型、環境変数、文、キーワードなどのプログラムの要素を表すために使う。

実行した結果やグラフなど何らかの出力の説明が必要な場合は、コードと合わせて出力結果の画像を記載しています。



このアイコンは一般的な注記を示す。

# 目次

はじめに	v
<b>1章 Pythonの基礎と開発環境</b>	<b>1</b>
1.0 はじめに	1
1.1 Pythonとは	1
1.2 Pythonのバージョン	2
1.3 開発環境の用意：Google Colaboratory	3
1.4 基本構文	4
1.5 FizzBuzzのコードの解説	8
1.6 Pythonのチュートリアル	9
1.7 基本データ型	9
1.8 まとめ	16
<b>2章 データ分析に必要なライブラリ</b>	<b>17</b>
2.0 はじめに	17
2.1 NumPy	17
2.2 Pandas	23
2.3 Matplotlib	28
2.4 ライブラリを活用した基礎分析	33
2.4.1 性別比	33
2.4.2 年齢構成	37
2.5 まとめ	40
<b>3章 Scikit-Learnを用いた機械学習</b>	<b>42</b>
3.0 はじめに	42

---

3.1 機械学習の基本 .....	42
3.2 Scikit-Learnで重回帰分析 .....	46
3.3 自動車の価格予測モデルを作る .....	47
3.4 応用：SVM(サポートベクターマシン) .....	51
3.4 まとめ .....	52
<b>4章 ディープラーニング入門 .....</b>	<b>53</b>
4.0 はじめに .....	53
4.1 機械学習とディープラーニング .....	53
4.2 機械学習とニューラルネットワークの学習フェーズ .....	54
4.3 ディープラーニングと特徴量 .....	57
4.4 ニューロン .....	57
4.5 ニューラルネットワーク .....	60
4.6 TensorFlow .....	63
4.7 Keras .....	64
4.8 パーセプトロン .....	65
4.9 多層パーセプトロンで分類する .....	67
<b>5章 ディープニューラルネットワーク .....</b>	<b>74</b>
5.0 はじめに .....	74
5.1 ディープラーニングと画像認識 .....	74
5.2 畳み込みニューラルネットワーク(CNN)とは .....	76
5.3 CNNの応用とその他のディープラーニング .....	80
5.4 CNNによる画像の分類 .....	80
5.5 まとめ .....	87

# 1章

## Pythonの基礎と開発環境

### 1.0 はじめに

機械学習を進める上で必要なPythonの基礎を学びます。Pythonを活用することで、業務分析や機械学習を効率良く進めることができるようになります。まずはPythonの歴史や特徴から始め、基本的な挙動や関数を学び、簡易的なスクリプト記述をしていきましょう。

### 1.1 Pythonとは

Pythonは汎用的なプログラム言語で、世界中で広く使われています。実行前にコンパイルが不要なインタプリタ言語で、LinuxなどのUNIX環境、macOSでは標準搭載、Windows環境でもインストールすれば実行可能です。

インデント（行頭の位置揃え）でプログラムのブロックを表現するなど、コードがシンプルになるように設計されており、様々なプログラムをわかりやすく書けるのが特徴です。

またPythonには、数学関数、正規表現、ネットワーク通信などのライブラリが標準搭載されています。他のプログラミング言語を触った経験がある場合は、標準搭載の機能で出来ることの幅広さに驚かれるかも知れません。

PyDataという言葉があるように、データサイエンスや人工知能の分野でPythonは幅広く活用されています。その理由は、Pythonに数学関数のライブラリが標準搭載され、さらに高度な数学のライブラリが開発されたことが理由といえます。

## 1.2 Pythonのバージョン

Pythonは1991年にVer. 0.9.0がリリースされ、30年ほどの歴史があります。メジャーバージョンとしてはPython 1系、2系、3系があり、現在は3系がメインで使用されています。

Python 3系は2008年から開発が始まりました。2019年4月時点の最新バージョンは3.7.3で、3.8系はベータ版がリリースされています。

Python 2系と3系の違いはいくつかありますが、大きく異なるのは文字列の扱いです。Python 2系では、文字列には非Unicode文字列のstr型と、Unicode文字列のunicode型の2種類があります。日本語を扱う際はunicode型を使用するため、型を意識してコードを書く必要があります。一方、Python 3系では文字列はすべてUnicodeで扱うようになり、日本語を扱うプログラムが書きやすくなりました。

本書では、Python 3.6を使用します。

それではPythonを利用できる環境を用意していきましょう。

### 1.3 開発環境の用意: Google Colaboratory

Colaboratoryは、Googleが提供するクラウドで実行できる Jupyter Notebook環境です。Jupyter Notebookとは、ブラウザ上でPythonを対話モード(インタラクティブモード)に近い状態で実行し、実行結果を記録できる環境です。主にデータの分析作業を進めるためのツールとして利用されています。

設定不要で、コードの記述と実行、解析の保存や共有、強力なコンピューティング リソースへのアクセスなどをブラウザからすべて無料で利用することができます。さらに機械学習に使う各種ライブラリが最初からインストールされています。

それではColaboratoryを利用するために、Googleのアカウントを取得しましょう。

#### Google アカウントの作成

<https://accounts.google.com/signup>

アカウントが作成できた方、もしくは既にアカウントをお持ちの方は、Colaboratoryに以下のURLからアクセスします。

#### Colaboratory スタート画面へのアクセス

<https://colab.research.google.com/notebooks/welcome.ipynb?hl=ja>

以下の画面が表示されます。



## 1.4 基本構文

基本構文を学ぶためFizzBuzzを書いて手を動かしながら、学習を進めていきます。

FizzBuzzとは、1～100までの数字を言い合うゲームで、このようなルールがあります。

- 3で割り切れれば、「Fizz!」という
- 5で割り切れれば、「Buzz!」という
- 3と5で割り切れれば、「Fizz Buzz!」という
- 上記以外の場合は、そのまま数字をいう

これをPythonで書いてみましょう。

それでは、Colaboratoryで新しいノートブックを作成します。