

## データサイエンス・AI 教育の実践報告 (2)

2021/03/11

応用物理学科 玉井良則

(3月5日発行のFD通信R2-22の続きです。あわせてお読みいただければ幸いです。)

### 4. データサイエンス・AI 入門

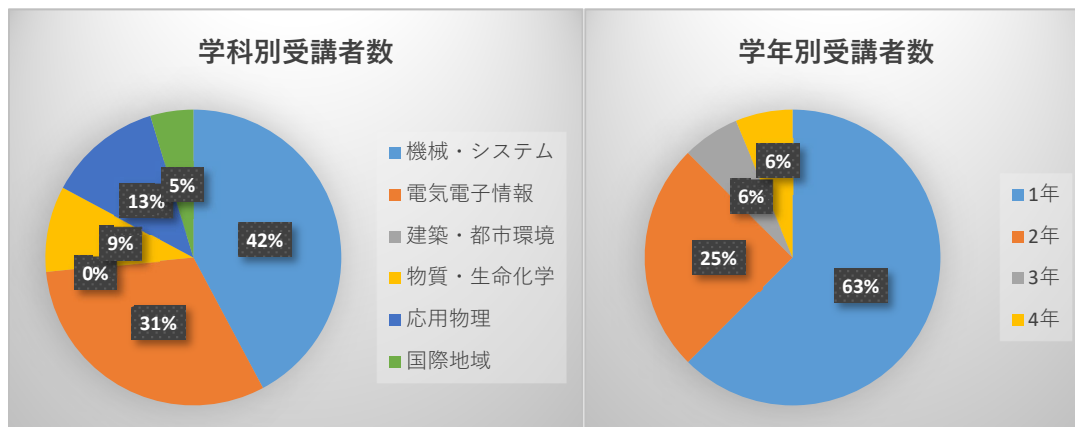
#### 4-1 教科書と授業の構成

私はデータサイエンスが専門ではありませんので、授業で何をすればよいのかよくわかっていませんでした。そのため、滋賀大学データサイエンス学部の教員の方々が執筆された「データサイエンス入門」 <https://www.gakujutsu.co.jp/product/978-4-7806-0701-7/> を教科書として使うことにしました。滋賀大学は「数理・データサイエンス教育強化拠点コンソーシアム」の拠点6大学の1つですので、間違いのないでしょう。また、教科書に合わせた動画がYouTubeで公開されていて、予習・復習に活用できます。この教科書の欠点は、記述が簡潔すぎることです。学生が理解できるようにするには、1コマで3ページ分の内容を説明するのに、約30枚のスライド作成が必要でした。資料を集めてデータサイエンス初心者の私が理解できるようにスライドを作成すると、結果として学生にもわかりやすいスライドになったようでした。

授業は、スライドを用いた解説とスプレッドシートやPythonを用いた課題を組み合わせ、クリッカーなども使いながら進めました。授業の当日中に解答する小テストで基礎知識の定着を図り、週末までに提出する課題で、授業で習ったことをすぐに実際に試してみる構成としました。この組み合わせは、学生にも好評でした。課題は私自身も楽しめるようなものという観点で、独自の物を作成しました。授業のトピックは表1 (FD通信R2-22) をご参照ください。

#### 4-2 受講者の構成

受講者は64名で、実習を行うには適正な人数でした。文系の学生にも多く受講して欲しかったのですが、ほとんどが工学部の学生でした。Pythonを使うので敬遠されてしまったのかとも思いましたが、国際地域学部の学生にはプログラミングが目的で受講した学生もいて「Pythonの実践ができて非常に良かった」という感想をもらいました。様々な需要があるものだ実感しました。工学部内ではほぼ学科定員に比例していますが、物質・生命化学科が少なめ、建築・都市環境工学科は0名でした。学年別では1年生の受講者が多く、データサイエンス・AIに対する興味が高いのかもしれませんが。



### 4-3 座学と実習の融合

この授業でこだわったことは、実際にデータサイエンス・AIをツールとして使えるようになることです。そのため、座学だけではなく、スプレッドシートやPythonによる実習を導入しました。当初は1~10回目を講義室で座学、11~15回目を情報処理演習室で実習という計画でしたが、コロナ禍でリアルタイム遠隔授業での実施となったため、学生は常にPCの前に居ることになり、毎週の授業で座学的内容と実習をミックスできるようになりました。結果としては、これが非常に良かったと思っています。

トピック「データ分析の基礎」では、G Suiteに付属のGoogleスプレッドシートを主に使いました。Microsoft 365 Web版のExcelよりも機能が豊富で、データサイエンスの授業に十分使えるものです。クロス集計機能はかなり強力でした。唯一、箱ひげ図だけはデスクトップ版のExcelを使いました。後半のトピック「データサイエンスの手法」と「AI実習」では、Pythonを使用しました。データサイエンス用のライブラリが充実しており、比較的簡単なプログラムでさまざまな解析ができます。

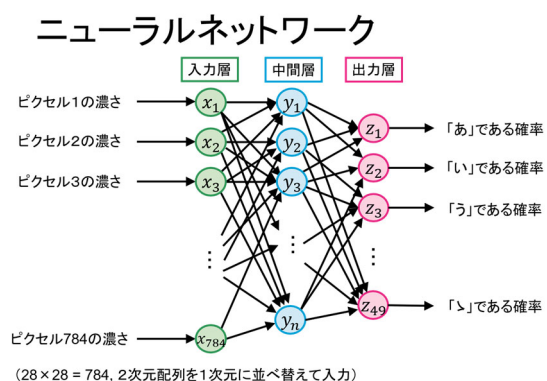
### 4-4 Colaboratory の活用

Pythonは、高度な使い方も簡便な使い方もできる柔軟性に富んだプログラミング言語です。今時、プログラミングは小学校でも必修になりましたし、大学生がプログラムを組めないわけにはいきません。初回の授業でアンケートを取ったところ、受講者の70%はプログラミング未経験で、多少焦りましたが、ほとんどの学生は十分に課題ができました。

初学者でも無理なく実習を行うためにGoogle Colaboratoryが非常に有効でした。Python学習環境としてよく用いられるJupyter NotebookをGoogleの仮想サーバ上で実行できるものです。面倒なインストール作業は不要で、ブラウザさえあれば、いつでもどこでも、気軽にプログラミングができます。Classroomと一体として使うことができ、Classroomの課題をクリックするとColaboratoryが起動してすぐにPythonの課題に取りかかることができます。Notebookの名の通り、課題の説明や解説とプログラムを混在して記述でき、プログラミング練習帳のような感じになります。学生が最初に取り組んだPython実習課題の例を、資料1に示しておきます。

### 4-5 AI実習

学生がAIにかなり興味を持っているのは実感していましたので、ニューラルネットワークを使った実習を用意しました。比較的簡単でかつ目に見えて実感できる題材として、枕草子の「くずし字」の文字認識を取り上げました。KerasというPythonのライブラリを使うと、Colaboratory上でGPUを用いたディープラーニングを簡単に実行することができます。くずし字の学習用データは、人文学オープンデータ共同利用センターでKMNISTデータセット <http://codh.rois.ac.jp/kmnist/> として公開されているものを使わせていただきました。古典文学の文字認識は結構ホットな話題のようです。

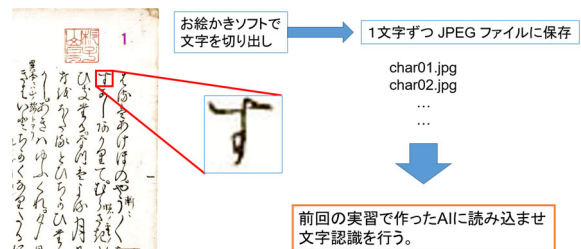


まず、いろいろな条件でくずし字を学習させ、あまり賢くない AI や賢い AI を何通りか作成します。これを使って、枕草子十三行古活字版 1 (九州大学附属図書館) から切り出した文字を認識させ、判定の精度を評価するという課題です。

中間層の数を変えたり、学習の繰り返し数を変えたりすると学習の精度が変わり、過学習なども実際に体験できます。AI の威力と限界の両方が実感できたのではないかと思います。学習したのは

ひらがな 49 文字でしたが、意地悪な課題として漢字の文字判定もしてもらいました。「うまく認識できない」という結果になると予想していたのですが、漢字の「乃」がひらがなの「の」に認識されたりして、学生にも私にもいろいろと発見があって楽しめました。共通教育の授業でここまでできるようになったのは、素晴らしい環境だと思いました。

## 文字を切り出してAIで判定



### 4-6 アンケート結果

半期の授業終了後に学生に回答してもらったアンケートの結果を、資料 2 に添付します。全体として学生はデータサイエンス・AI に非常に興味を持ち、将来役に立つと感じてくれたようです。後半のトピックでは難易度が高いと感じた学生が増えています。リテラシーレベル「+α」をターゲットとしていたので狙い通りと言えますが、TA を活用したサポート体制の充実などが必要だと思いました。なお、AI 実習を難しかったと回答した学生も、ほとんどが「興味を持った」と回答しています。

課題の分量については、毎週、フォームの問題と実習課題がありましたので、決して少なくはなかったのですが、興味を持って取り組んでくれたようで、負担に感じる学生は少数でした。

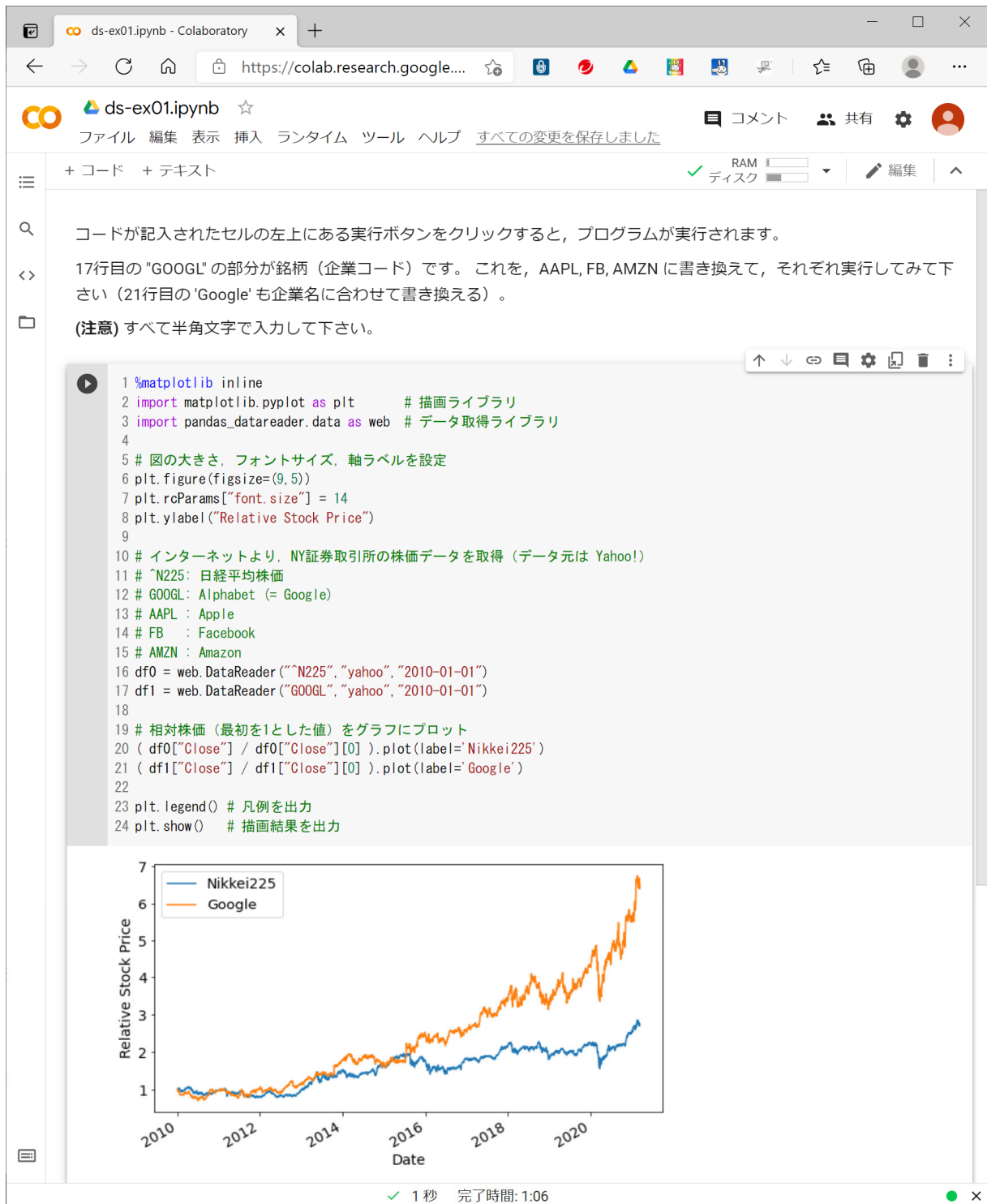
### 4. おわりに

今回、データサイエンスの授業を担当させていただく機会を得ましたが、学生の興味もさることながら、教育や研究のツールとして非常に可能性を秘めたものだ実感しました。文部科学省は「数理・データサイエンス・AI 科目」の全学必修化を推し進めていますが、文科省に言われるまでもなく、教育課程にうまく取り込んでいくと、学生も教員もハッピーになれるのではないかと思います。

学生の履修のわかりやすさや教育効果を考慮すると、1 年次に情報リテラシーの基礎とデータサイエンスのリテラシーレベル+αを必修科目として履修し、その後、学生の興味や専門にに応じて、より高度な科目を選べるような仕組みにするとよさそうです。そのうえで、既存の専門科目にもデータサイエンスの要素を組み込んだり、複数学部混成で、グループワークを主体とした PBL 科目などを用意したりすると、学生の興味をより引き出せると思います。

(長文にお付き合いいただき、ありがとうございました。)

## 資料1 Google Colaboratory の実行画面の一例



The screenshot shows a Google Colaboratory notebook interface. The browser address bar displays `https://colab.research.google...`. The notebook title is `ds-ex01.ipynb`. The code cell contains the following Python code:

```
1 %matplotlib inline
2 import matplotlib.pyplot as plt # 描画ライブラリ
3 import pandas_datareader.data as web # データ取得ライブラリ
4
5 # 図の大きさ, フォントサイズ, 軸ラベルを設定
6 plt.figure(figsize=(9,5))
7 plt.rcParams["font.size"] = 14
8 plt.ylabel("Relative Stock Price")
9
10 # インターネットより, NY証券取引所の株価データを取得 (データ元は Yahoo!)
11 # ^N225: 日経平均株価
12 # GOOGL: Alphabet (= Google)
13 # AAPL : Apple
14 # FB : Facebook
15 # AMZN : Amazon
16 df0 = web.DataReader("^N225", "yahoo", "2010-01-01")
17 df1 = web.DataReader("GOOGL", "yahoo", "2010-01-01")
18
19 # 相対株価 (最初を1とした値) をグラフにプロット
20 ( df0["Close"] / df0["Close"][0] ).plot(label='Nikkei225')
21 ( df1["Close"] / df1["Close"][0] ).plot(label='Google')
22
23 plt.legend() # 凡例を出力
24 plt.show() # 描画結果を出力
```

The code cell is followed by a line graph titled "Relative Stock Price". The x-axis is labeled "Date" and ranges from 2010 to 2020. The y-axis is labeled "Relative Stock Price" and ranges from 1 to 7. Two lines are plotted: a blue line for "Nikkei225" and an orange line for "Google". Both lines show an upward trend, with the Google line reaching a value of approximately 6.5 by 2020, while the Nikkei225 line reaches approximately 2.8.

At the bottom of the notebook, the execution status is shown as "1秒 完了時間: 1:06".

GAF A (Google, Amazon, Facebook, Apple) について, NY 証券取引所の株価データをインターネットで取得してグラフにする課題です。プログラムは完成した形で与え, 企業名を各自で書き換えて実行してもらいました。GAF A の台頭に関する解説とセットになっています。

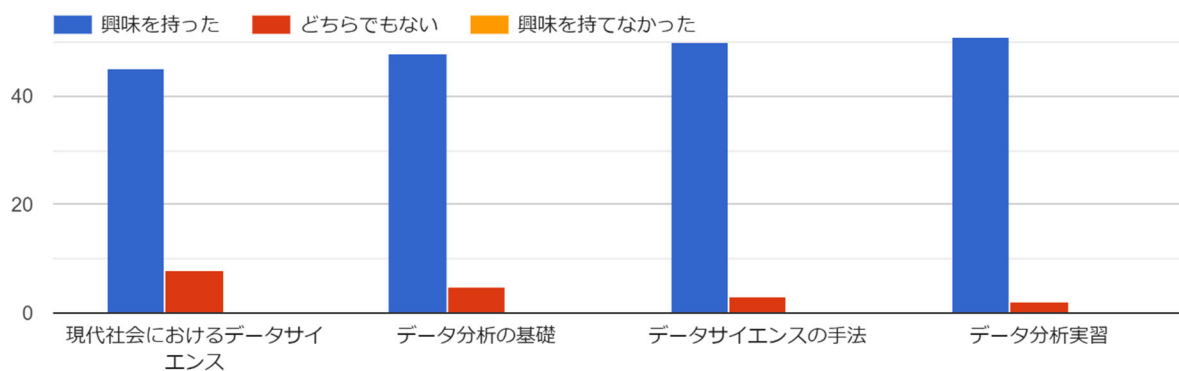
## 資料2 授業アンケートの結果

半期の授業終了後に学生にアンケートを取りました。棒グラフは授業の4つのトピック別になっています（AI実習は「データ分析実習」と表記）。回答数は53（回答率83%），縦軸は人数。

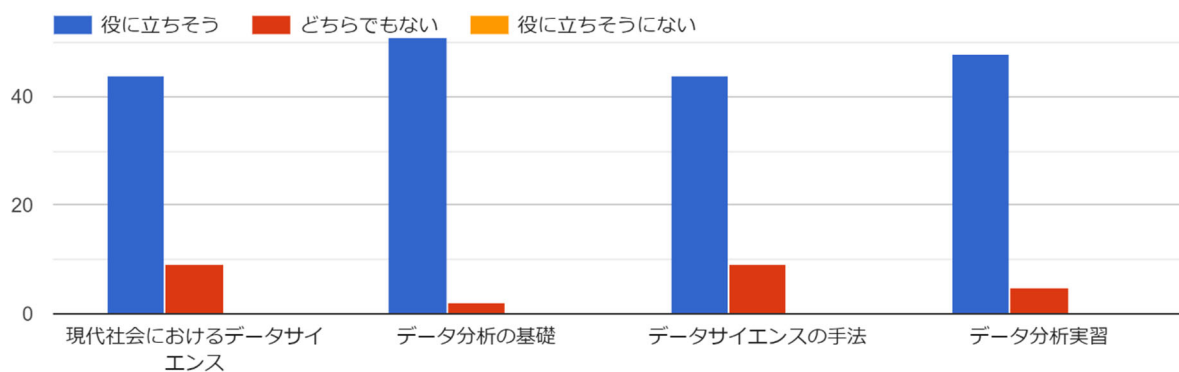
（表1の一部を再掲）

科目	トピック	内容
データサイエンス・AI入門	現代社会とデータサイエンス	データサイエンスとは ビッグデータ活用とデータ倫理，GAFAsの台頭
	データ分析の基礎	ヒストグラム，箱ひげ図，散布図，平均値と分散 回帰直線，データ分析の注意点
	データサイエンスの手法	クロス集計，回帰分析，ベイズ推論，クラスター分析 決定木，ニューラルネットワーク，機械学習
	Pythonを使ったAI実習	Pythonの基礎，COVID-19データの可視化 古典文学（くずし字）の文字認識

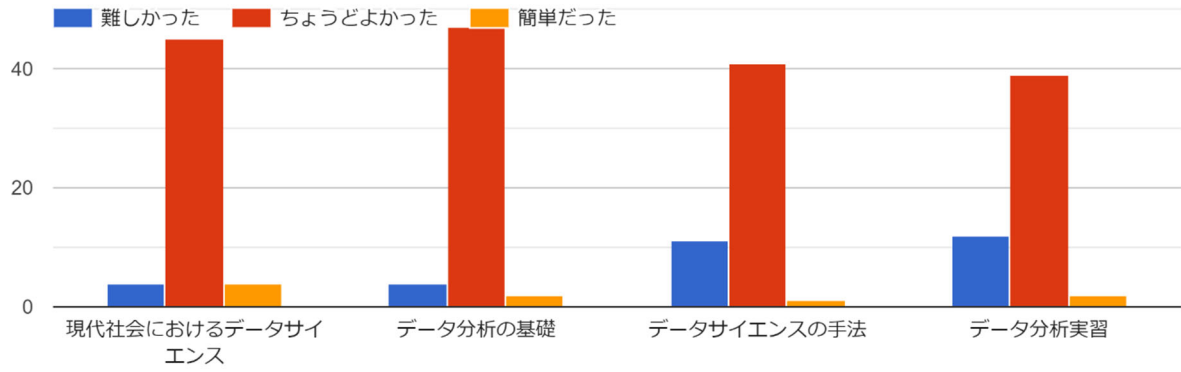
内容に興味を持ちましたか？



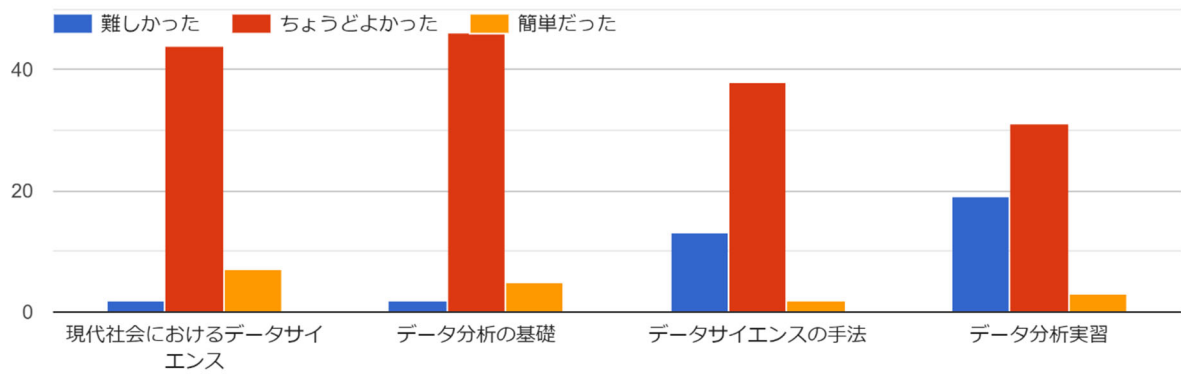
内容は今後、役に立ちそうですか？



講義（スライドなど）の難易度はいかがでしたか？



課題の難易度はいかがでしたか？



課題の分量は適切でしたか？

53件の回答

