

UEC データサイエンスプログラム

—2022年度社会人向け講座のご紹介—

2022年3月8日

国立大学法人 電気通信大学特任教授

斉藤史朗



1. 自己紹介

➤ 学問歴

- ✓ 1983年東京大学法学部を卒業：法制史のゼミにいた
- ✓ 2004年東京大学人文社会系大学院修士課程入学
- ✓ 2017年博士号取得：「日本社会学における家理論の形成と展開—その社会像と政治観」（学説史・文献学）

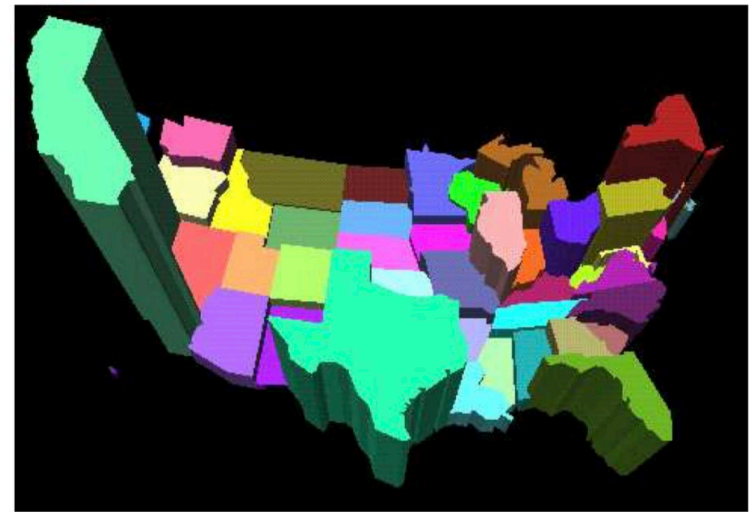
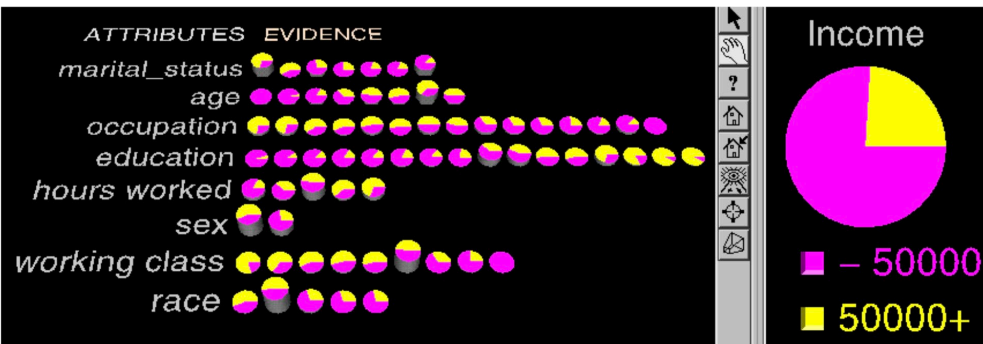
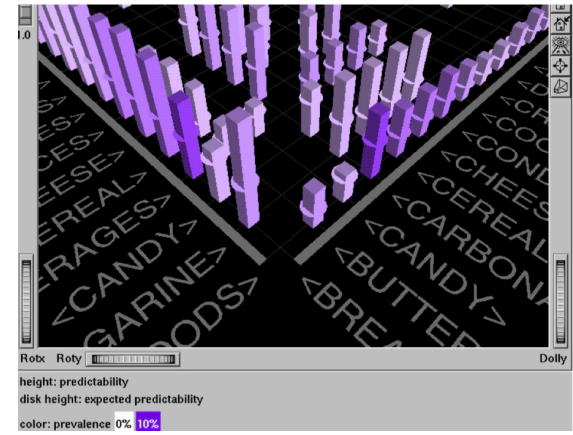
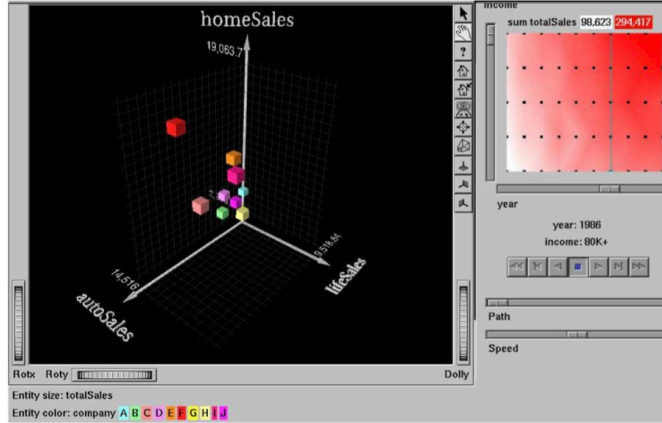
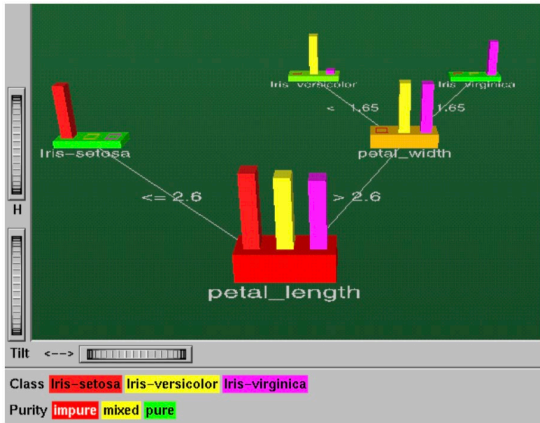
➤ 職歴

- ✓ 様々な仕事（舞台監督なども）を経て
- ✓ 2001年12月から株式会社金融エンジニアリング・グループ
- ✓ 2018年7月から株式会社ブレインパッド
- ✓ 2019年11月から電気通信大学特任教授

➤ データサイエンス歴

- ✓ 1999年にSGIのMineSetというツールに出会う
- ✓ 2000年に“Mastering Data Mining”の翻訳チームに参加
- ✓ 統計やデータマイニングは独学で学ぶ
- ✓ 2014年からデータサイエンティスト協会企画委員会の委員長、2019年から2021年までは事務局長

参考 : SGI(Silicon Graphics Inc.)のMineSet



こんな本出しています。



「超」文系の本です。博士論文が元になっています。漢字が多いです。数字はほとんど出てきません。数字よりもドイツ語の方が多いくらいです。



翻訳書です。共訳なので、他の人の担当のところでも、日本語が「アレ」なところも多いです。データ分析が実際にどう使われるのかを知るのにとても良い本です。



データサイエンティストの育成や管理の本です。人事関係の参考書と言っても良い本です。お腹を抱えて笑える事例が多いです。

2. 講座の枠組み

社会人向け講座の背景

➤ データアントレプレナーフェロープログラム

- ✓ 2015年度から始まったデータサイエンスの講座が母体となっています。
- ✓ 2017年度からは文科省の「データ人材育成」プロジェクトの助成を受けてきました。
 - 大学院生と社会人学生とが半々で学ぶ1年ものの講座
 - 現在も継続しています。（略称・DEFP）

➤ 社会人向けリスキリング（リカレント）講座の開始

- ✓ 産業界からの要請
 - データサイエンス人材が不足している！
 - 必要な即戦力を確保したい（社会人教育の方が早道！）
- ✓ 世の中の講座ものに不足が多い
 - 「机上の空論」が多い：会社に帰って「使える」スキルが欲しい！
 - データを使う仕事の「足回り」が脆弱
- ✓ 2021年度にテスト実施：「ゼロから始めるデータサイエンティスト養成講座」
 - 半年ものの講座を2回実施して何をどうすべきかをリサーチした

運営主体：産学官の連携・コンソーシアム

➤ 産学官の連携プロジェクト

- ✓ 「データアントレプレナー・コンソーシアム」という産学の連携組織から出発した団体が運営主体となっています。
- ✓ 2020年度から神奈川県の逗子市が連携組織として参加し、自治体との連携が開始されました。

➤ 産業界との連携

- ✓ 講師としての参加
 - トップレベルのデータサイエンティストとして活躍する方が講師を務めています
 - KaggleGrandmasterやKaggleMasterの講義もあります。
- ✓ データ・スポンサー
 - 実務に使われている実際のデータを講義用にご提供いただいています。

➤ 自治体との連携

- ✓ 課題やデータの提供を行っていただいています。

3. 教育の目標： どのような人材を育てたいのか

➤ 文科省のDDDRIVEに最初に採択された4拠点の名称

- ✓ 早稲田大グループ：Development of Data Analysis Talents
- ✓ 大阪大学グループ：独り立ちデータサイエンティスト人材育成プログラム
- ✓ 医科歯科大学グループ：医療・創薬データサイエンスコンソーシアム
- ✓ 電気通信大学グループ：データアントレプレナーフェロープログラム

➤ 「データアントレプレナー」って何？

- ✓ わかりにくい、特に学生にはわかりにくい
- ✓ あえて「アントレプレナー」の語を冠した意味がある
 - わかりやすくするなら、こんな表現は不要

- **データアントレプレナーとはUEC データサイエンス・プログラムの基軸**
 - ✓ 「データサイエンティストとしての素養を持ち、新たな価値を生むビジネスを創出できる人材」

- **データサイエンスを教えながらも、データサイエンスを自己目的としているわけではありません。**
 - ✓ このプログラムは、先端的なデータサイエンスの研究・開発をすることが目的ではありません。
 - ✓ あくまで目的は、「新たな価値」を創造する人材の輩出です。
 - ✓ いわばデータサイエンスはブレークスルーのための武器です。
 - ✓ それゆえ、データサイエンティストの実践力を鍛えることがプログラムの重要な目的となります。

▶ データサイエンスとは「使う」ためのものである

- ✓ 「データ」には意味はない。データから意味を取り出して有用な「情報」にするのが「データサイエンス」
- ✓ 適用されない「データサイエンス」 = 机上の空論は無意味

▶ データサイエンスの「使い方」

- ✓ 単に技術を教えるのではなく、どういう場面で、どのように「使う」のかを教えてください。

▶ 何のためのデータサイエンス

- ✓ データサイエンスの「使い方」の先にある、「何のための」データサイエンスなのかを意識してもらうように努めています。

▶ 実際のデータ、実際の課題に取り組みます

- ✓ 実データを使ったPBL(Project Based Learning)は7年間の実績がある
 - しかし、多くの他のデータサイエンス学部等でも行われている。
- ✓ 単にアルゴリズムを適用する対象としてではなく、価値を生み出す源泉としてデータを捉える観方を教えます

▶ データをよく見る

- ✓ 「データサイエンス」は「データ」の科学
- ✓ 「前処理」をしつこくやります。
- ✓ 「前処理」をやり切るだけの力を身につけていただきます。

▶ アルゴリズムの性質を良く知る

- ✓ 単に手法の使い方を覚えて、形式的に精度を上げるにとどまらない。
- ✓ なんのためにその手法を使うのかの吟味ができるようにする。

4. 講座の概要

➤フルオンライン

✓講義はZoomで行ない、専用のSlackを立てて議論・質疑を行います。

➤社会人用のプログラムですが、学生も混じります

✓学生は社会人から、社会人は学生から、様々な学びがあります。

➤グループワークが主となります

✓データサイエンスの実業務は数名のグループで行うことが多いです。

- 一人でデータサイエンス業務を行うことにはさまざまなリスクがあります。

✓データサイエンスの技術を教えるだけでなく、仕事のやり方も学んでもらいます。

➤実際のデータ、実際の課題を使って学びます

✓実際のデータを分析し、モデルを作って実際の課題を解決することを常に意識してもらいます

参加に必要なもの（ネットワークやマシンについて）

➤ マシンの要件

- ✓ WindowsでもMacでも構いませんが、平日、休日、業務時間外でも持ち帰って学習できるようなノートPCをご用意ください。
 - もしこの講習用に新たにPCを購入なさるのであればMacのノートブックをご購入ください。（Unix環境が簡単に手に入るのです。）
- ✓ Pythonの統合環境「Anaconda」をインストールして使っていただきます。
 - メモリ8G以上、空きディスクが200G以上が必要です。

➤ ネットワーク

- ✓ ZoomやMicrosoft Teams、Google meetなどのオンライン会議システムに入れるネットワークが必要です。
- ✓ 質疑等のコミュニケーションをSlackで行いますので、Slackが使えるネットワークが必要です。
- ✓ 共通した計算環境を得るためにGoogle Colaboratoryにもアクセスできるネットワークをご用意ください。

➤ 会議環境

- ✓ 1社から複数名ご参加の場合でも、会議室に集まってまとめてのご参加ではなく、個人個人でご参加ください。
 - 学習においては他社の方とチームを組む場合が多いので、マイクやスピーカーは個人に属している必要があります。

基礎レベル

コンピュータサイエンス

Pythonプログラミング

ITリテラシ

Python100本ノック

EDAとモデル作成の基礎

応用レベル

統計検定2級自習

回帰モデル作成

判別モデル作成

特別講義

エキスパートレベル

確率統計

DE実践論

特別講義

DS特論

深層学習の基礎

自然言語処理

レコメンド

因果推論

文科省

リテラシーレベル

応用基礎レベル

DS協会

見習いレベル

独り立ちレベル

棟梁レベル

到達目標

現在、データを活用して業務を行う上で必須とされるスキルを身につけた上で、データサイエンスのプロジェクトでリーダーに言われたことをきちんと実行できる。

DSプロジェクトでリーダーの指示の下、必要なことを自力で実行することができる。

DSプロジェクトでリーダーとしてデータ活用の目的をクライアントと相談して策定し、相応しい分析を行なった上で、結果を評価することができる。また、結果をシステム化するための要件を開発側に提示できる。

➤ DS基礎講座（総講義時間：80時間）

✓対象:社会人

✓開講期間

- 2022年4月から6月(応募者多数の場合には10月から12月に再度実施します。)

➤ 独り立ちデータサイエンティスト養成講座（総講義時間：134時間）

✓対象:社会人

✓開講期間

- 2022年4月から9月

➤ データアントレプレナーフェロープログラム（総講義時間：144時間）

✓対象:大学院生および社会人

✓開講期間

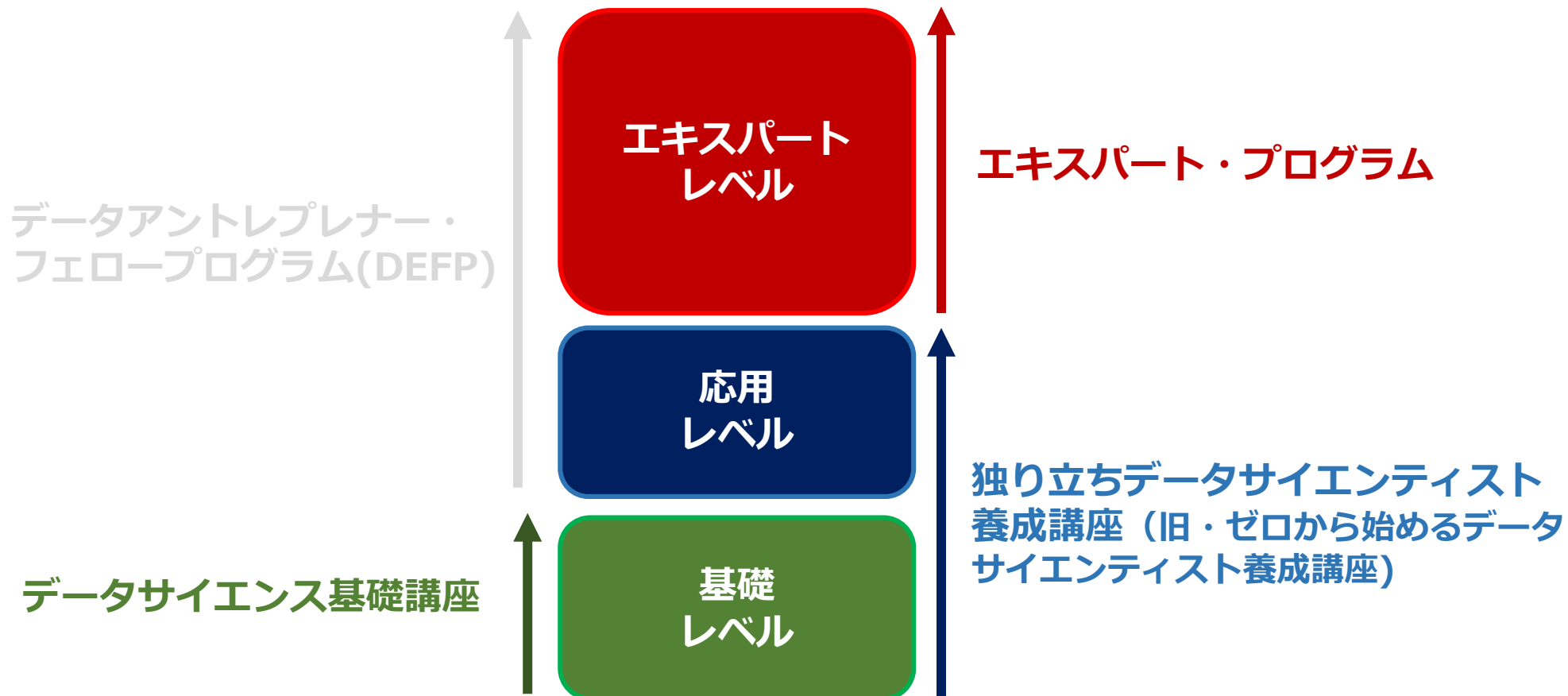
- 2022年4月から2023年3月

➤ エキスパートプログラム（総講義時間：90時間）

✓「独り立ちデータサイエンティスト養成講座」修了生で、特に許可を得たものは追加の費用をもって、エキスパートプログラムを受講することができます。

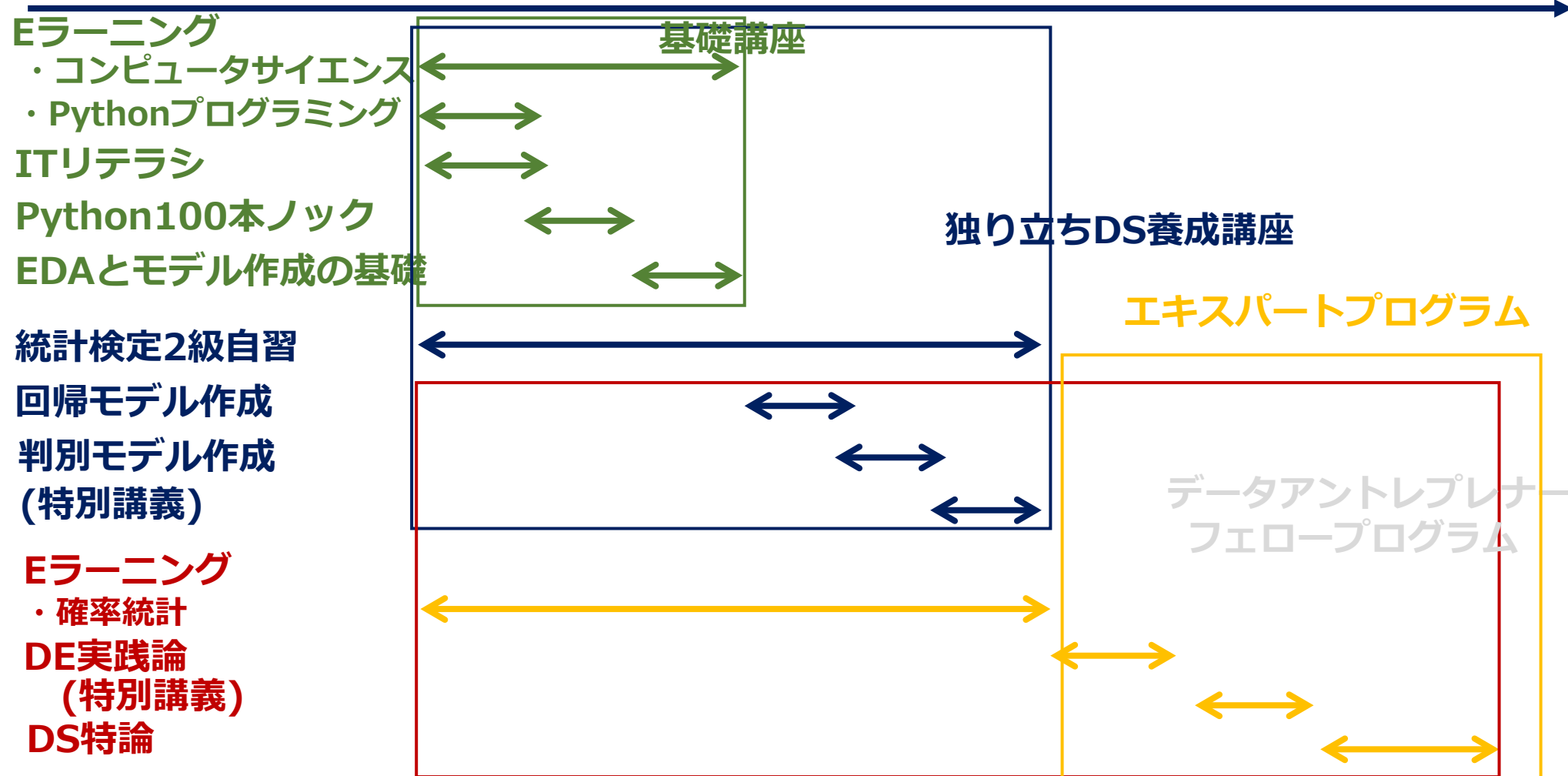
※毎週土曜日 13時から18時に講義を行います。

※講義時間の2倍から3倍の自宅学習時間が必要です。



UEC DSプログラム2022・開講スケジュール

4月 5月 6月 7月 8月 9月 10月 11月 12月 1月 2月 3月



eラーニング（DS基礎&独り立ち・共通）

➤ コンピュータサイエンス特論

- ✓ データを使って業務を行う際のベースとなる、「機械で計算すること」のエッセンスを学びます。
 - 機械としてのコンピュータの動き方・働き方の基本
 - 「計算すること」のベースにある数学的な考え方
- ✓ ここで身につける発想があるかないかでその後の「伸びしろ」が違います

➤ プログラミング言語特論

- ✓ プログラミング言語としてのPythonの入門講座です。
 - 100本ノックの入り口としての部分
- ✓ これから先のEDAや100本ノックでは扱わないような部分も含めて、重要な基礎を学びます。
 - 実際のデータ分析やモデル作成にあまり登場しないからと言って、不要なわけではない部分
 - 一歩上を目指すために必要な基礎
 - 後で「ああ、あれだったんだな」と思い出すことが重要

ITリテラシ講義の必要性（DS基礎&独り立ち・共通）

➤データサイエンス開始以前の問題

- ✓ビッグデータの時代、データや情報を扱う業務は、これまでとは仕事のやりかたやお作法が全く違ってきました。
 - データはクラウドにあるから「SQL書いて自分でとってきて」
 - 計算もクラウドでやるから、「GoogleのアカウントでColaboratoryを使って」
 - プログラムのバージョン管理は「GitHubでね」
- ✓こうした「新しい仕事のやり方」を身につける必要があります。

➤これまでの実績で積み重ねた経験を注入した講義

- ✓これまでの受講生がどこで詰まったのか、何がハードルとなっているのかを熟知した担当者が、受講生を指導した経験を教材化

cf. 受講生の声：「職場の若い人とのコミュニケーションがスムーズになりました。これまではSlackだのGitHubだのと言った若いメンバーの会話に入れなかったのですが、自分でも会話に入るだけでなく、実践できるので、若い人たちから「おおっ！」って驚かれています、前よりもコミュニケーションが取れて、とてもやりやすくなりました。」

ITリテラシ講義概要（DS基礎&独り立・共通）

➤ コンピューターで仕事をする

- ✓ コンピュータの構造：UNIXの基本コマンド
- ✓ 文字：文字コード、正規表現

➤ ネットワーク環境で仕事をする

- TCP/IP入門
- セキュリティの基礎
- HTMLとWebサイト

➤ チームで仕事をする

- ✓ GitとGitHub
- ✓ コミュニケーションツールの活用

➤ 大規模データの取り扱い

- ✓ データベース概論
- ✓ SQLの基礎

Python 100本ノック（DS基礎&独り立ち・共通）

➤ 目的

- ✓ データをもとに「自由に発想する」余裕を持つ
 - アイデアはすぐに蒸発するので、蒸発する前にデータにぶつける必要がある
 - リズムを持って分析する。
- ✓ 目を瞑っていても集計と可視化はできるようにする。

➤ やること

- ✓ データサイエンティスト協会で作った「データ分析100本ノック Python 篇」をやる。

➤ やり方：スパルタ・「虎の穴」方式

- ✓ 問題を全部、自分で解く
 - 教わっていないことは自分で調べる。
 - 「調べ方」を身につける。
- ✓ Zoomで画面共有をしてもらって進捗を確認する。
 - 苦手とか言わせない。
 - できるまでやる。

EDAとモデル作成の基礎（DS基礎&独り立ち・共通）

➤ 目的

- ✓ データを集計 & 可視化する方法を学びます。
 - データで何かを「語る」とはどのようなことか？
 - データで何かを「語る」ためには何が必要か？
- ✓ 予測モデルを作成することの基礎を学びます。
 - 精度とは何か
 - 現在のモデル作成の課題とは何か

➤ やること

- ✓ EDA
 - 社会科学の研究に使われた実データをもとに、データを集計・可視化して「論を立てる」やり方を学びます。
- ✓ モデル作成の基礎
 - 入門用コンペの課題にトライして精度をあげることを通じて、教師あり学習の基本を学びます。
 - 判別モデル：タイタニック
 - 回帰モデル：Jリーグの観客動員数予測

実データを使った本格的データ分析(「独り立ち」)

➤ 概要

- ✓ 実データを使って、モデルを作ってもらって、その精度をコンペ形式で評価
 - チーム戦：数名のチームで精度を競い合います。
 - 評価指標は講義の最初に明らかにしておきます。
- ✓ 業務でモデルを使う意義についても実践的観点から解説します

➤ 回帰モデル

- ✓ 使用データ：ワシントンD.C.の不動産価格データ
- ✓ 講師：G.A.テクノロジーズ（DEFPコンソーシアム参画機関）

➤ 判別モデル

- ✓ 使用データ：コンビニやスーパーなどの売上データ
- ✓ 講師：コニカミノルタ（DEFPコンソーシアム参画機関）

5. 受け入れ条件、受講条件や修了条件など

➤ アドミッション・ポリシー（受け入れ条件）

- ✓ 高校の数学II程度の数学のスキルがあること（必須）
- ✓ なんらかのプログラミング言語を触ったことがあることが望ましい
- ✓ なんらかのデータ分析を行ったことがあることが望ましい

➤ カリキュラム・ポリシー（受講条件）

- ✓ リアルタイムの講義には出席すること(毎週土曜日午後)
- ✓ 予習は不要ですが、復習に講義時間の2倍から3倍の時間が必要です。
- ✓ Eラーニングはサボっていると後で大変ですから計画的に受講してください。
 - 特に統計検定2級の試験は大変なので、支援プログラムを参考にして、計画的に臨んでください。

➤ ディプロマ・ポリシー（修了条件）

- ✓ 必要な試験に合格すること（「独り立ち」のみ）
- ✓ リアルタイムの講義に出席すること

6. 費用

➤ DS基礎講座

- ✓ 対象:社会人
- ✓ 費用:30万円+税

➤ 独り立ちデータサイエンティスト養成講座

- ✓ 対象:社会人
- ✓ 費用:50万円+税

➤ データアントレプレナーフェロープログラム

- ✓ 対象:大学院生および社会人
- ✓ 費用:50万円+税 (社会人博士は25万円+税)
 - 学生は無料

➤ エキスパートプログラム

- ✓ 対象:「独り立ちデータサイエンティスト養成講座」修了生で、特に許可を得た者
- ✓ 費用:30万円+税