

2023年度

データアントレプレナーフェロープログラム
について

2023年1月30日

電気通信大学 特任教授

齊藤史朗

1. 講座の枠組み

▶ データ関連人材育成プログラムの要点

発掘

ポテンシャルのある学生を取り込む

教育機関

- ・ 学生の募集のための大学間ネットワーク

育成

データサイエンティストとしての素養を育成

教育機関

- ・ **教育内容**
- ・ 教育コンテンツ開発のための大学間ネットワーク

活躍促進

活躍の場を提供：
インターン先から就職先まで

企業・組織

- ・ 出口（インターン・就職先）としての企業・組織とのネットワーク

■ 産学の連携プロジェクト

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

- ✓講師としての参画
 - 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なっていただいています。

2. 教育の目標： どのような人材を育てたいのか

➤ 文科省のDDRIVE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電通大グループの教育内容の基軸

- ✓「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（ビジネス力）を鍛えることがプログラムの重要な目的となります。

➤ データサイエンスは「使う」ためのものです

- ✓ 「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」
- ✓ 使われない「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（アルゴリズム）を教えるのではなく、どういう場面で、どのように「使う」のかを教えています。

➤ 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。

DEFPの特長：90点取れるデータサイエンティスト

➤今さらPBLやら実データの使用だけではアドヴァンテージでない

- ✓ 実データを使ったPBL(Project Based Learning)は8年間の実績あり
 - でも、最近では多くの大学でやっている
- ✓ 80点のデータサイエンティストではなく、90点を取れる人を育てる

➤データをよく見る

- ✓ 「データサイエンス」は「データ」の科学

➤アルゴリズムの性質をよく知る

- ✓ 単に手法の使い方を覚えて、形式的に精度を上げるに止まらずに、「アルゴリズムの気持ちができる」ようになる
- ✓ 何のためにその手法を使うのかの吟味ができるようになる

➤実績の一つとして

- ✓ KaggleのJigsawコンペで6位入賞チーム4名のうち2名がDEFP生
- ✓ <https://www.kaggle.com/c/jigsaw-toxic-severity-rating/leaderboard?tab=private>

3. 講座の概要

講義の進め方

➤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。

➤大学院生と社会人学生の共同のプログラムです

✓ 多様性です。データサイエンスは視点の種類が多いほど、良い解を見つける可能性が高まります。

➤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。

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

教師なし学習

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plusレベル

エキスパートレベル

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。

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

多様な

深層学習の基礎

自然言語処理

レコメンド

因果

DEFPP

文科省

リテラシーレベル

応用基礎plusレベル

エキスパートレベル

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

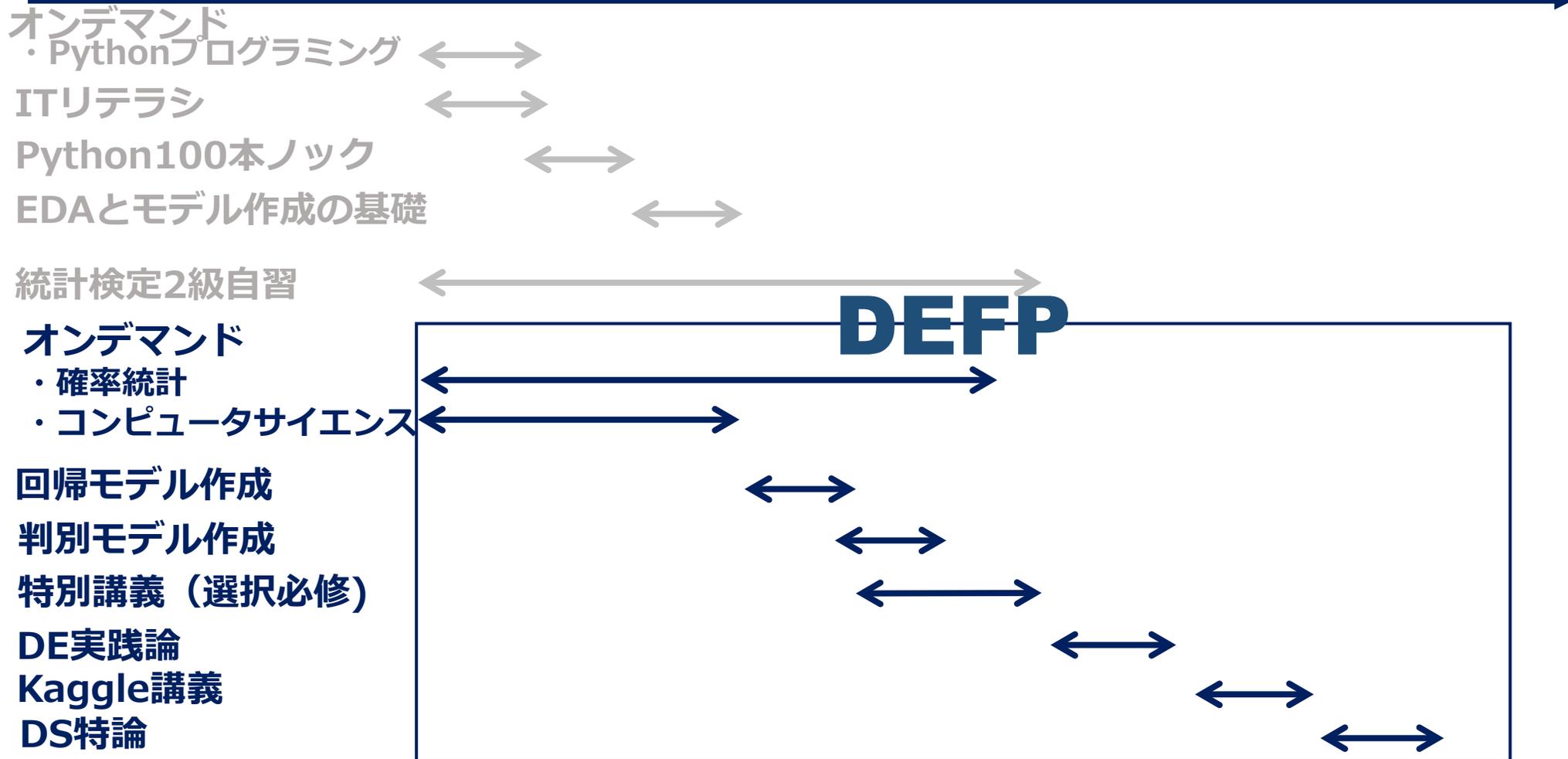
現在、データを活用して業務を行う上で必須とされるスキルを身につけた上で、データサイエンスのプロジェクトでリーダーに言われたことをきちんと実行できる。

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。

UEC DSプログラム2023・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



必修講義など（具体的な日にちは変更する場合があります。）

➤ 講義実施曜日：原則は土曜日午後。そうでない日は特記。

➤ 必修講義

✓ オンデマンド講義

- この2科目を終了しないと後半の対面講義を受講することができません。

✓ 対面学習

- 回帰モデル作成（不動産データ分析）
- 判別モデル作成（マーケティングデータ分析）
- デザイン思考(8月6日(日)、13日(日))
- データアントレプレナー実践論
- Kaggle講義
- データサイエンティスト特論

➤ 選択必修（以下のどちらかの1科目を必修。両方履修も可。）

- ✓ 医療データ分析の実践（8月15日(火)～18日(金)）データサイエンティストによるモック（MVP）作成

- **DEFPの受講生は、提供しているUEC DS Programmsの講義を全て追加料金なしで受講できます。**
- **DS基礎講座や「統計検定2級合格支援」講義**
 - ✓ 得意・不得意はありますので、自信がない分野については、学期の最初の方に開講されている講義を受講して、復習をしておいてください。
 - 不安な方はご相談ください。
 - ✓ 統計検定2級レベルの確率・統計の知識
 - このレベルが受講条件ですが、心配な方は、統計検定2級を早い時期に取るようにしてください。必修の「確率・統計」は統計検定準一級のレベルなので、統計検定2級程度のことはしっかりとできておいてください。
- **そのほかの発展的なオンデマンド講義**
 - ✓ 必修ではありませんが、できるだけ多くの講義を受講して、講義内の練習問題等をやって、内容を身につけておいてください。
 - ✓ 選択必修の講義によっては、オンデマンド講義のどれかを事前学習しておくことを要請する場合があります。

4. 講座の内容

➤ ITリテラシ（講師：MNTSQ・平田拓水）

- ✓ データや情報を使って仕事をするに際しての必要な基礎スキルを身につける。
 - Googleクラウドを利用した、ネットワーク越しの作業（Linuxコマンド等）
 - HTMLの基本やネットワーク・セキュリティに関する問題(httpsなど)
 - ドキュメント（プログラム）のバージョン管理（GitHub）

➤ Python100本ノック（講師：電通大・齊藤史朗）

- ✓ 「データの集計と可視化」は目を瞑ってでもできるようにします。
- ✓ データサイエンティスト協会が公開している「データサイエンス100本ノック(構造化データ加工編)」のPython編を100問全部やります。

➤ EDAとモデル作成の基礎（講師：helixdata・赤間悟）

- ✓ 計量経済学や計量政治学で著名なデータと研究を利用して、実際の具体的な課題を解決するために「集計」と「可視化」をどう利用するのかを学びます。
- ✓ コンペの練習問題を使って回帰モデル(Jリーグ)と判別モデル(titanic)の作り方からsubmitまでを学びます。

➤ コンピュータサイエンス特論（講師：電通大・西野哲朗）

- ✓ そもそも、コンピュータが何をしているのかの、一番重要な部分を皆さんが考える、考え方を教えます。
- ✓ 数学とプログラミングを使って何らかの「問題を解く」とはということなのか、問題解決の深い部分に取り組む基礎を作ります。

➤ 確率・統計（講師：電通大・原田慧）

- ✓ 通常のデータサイエンス業務に要求されるよりも、一段高度な確率と統計の考え方を見つけていただきます。
 - レベル的には統計検定の準一級くらいになります。
- ✓ 優れたデータサイエンティストになりたいなら、ここら辺にこだわってほしいというところに重点を置いてあります。

➤ 回帰モデル作成(不動産価格分析)(講師：G.A.Tech・橋本武彦、他)

- ✓ 住宅の価格を予測するモデルを作成することを通じて、実践的な回帰モデル作成の技術を身につけます。
- ✓ 単に精度を上げるという以上に、モデルにとって重要なことが何であることを学びます。
- ✓ データ分析をして価値を出すという業務について実践的に学びます。

➤ 判別モデル作成(マーケティング分析)(講師：コニカミノルタ・清水隆史、他)

- ✓ 小売現場の実際の販売データを使って、判別モデルを作成し、実践的な判別モデル作成の技術を身につけます。
- ✓ 業務におけるモデルの意義や、各ステークホルダーとの関係など、データ分析の結果の社会実装の仕方についても学びます。

➤デザイン思考(講師：セイコーエプソン・岡山俊明)

- ✓ 既存のものの単なる「改善」ではない、「新しい」モノやコトを作ることとはどういうことなのかを学びます。
- ✓ 「新しい」モノやコトを作り出すためのマインドセットを、多くの事例や演習で学びます。

➤データアントレプレナー実践論（講師：ソニー・島田啓一郎、他）

✓データサイエンス社会実装のやり方を学びます。

- 【講義】データサイエンスの最前線で活躍されている講師（9人）の話を聞き、データサイエンスを社会実装する際の課題や障害の克服の仕方等を学びます。
- 【演習】ある課題を解決するデータサイエンスの実装のイメージを、チームで話し合って作ります。

✓最終課題（ピッチコンテスト）

- 自分自身が考える「データサイエンスの社会実装」の例をプレゼンします。
- スライド4枚程度・持ち時間＝3分。
- 審査員はコンソーシアム会員企業所属員および電通大教員
- 優秀者には表彰・賞品あり

➤ Kaggle講義（講師：DeNA・加納龍一、Panasonic・阪田隆司、他）

- ✓ 過去のKaggleの良問をテーマに応じて精選（InClassを作っています）。
- ✓ テーマに関する講義を90分実施
 - Top Kagglersが何を重視しているのかを学びます。
- ✓ 150分、そのテーマの課題について、モデルを作って投稿します。
 - リアルタイムにJackさん、Cashさんと競うことができます。
 - 150分経ったら、締め切って、private leaderboardを開けます。
- ✓ 発表とコメント
 - Leaderboard上位の3、4名に解法を簡単に説明してもらいます。
 - それらの解法について、加納さん、Jackさん、Cashさんにコメントしてもらいます。
 - JackさんやCashさんの解法も解説してもらいます。
- ✓ ノートブックの共有
 - 発表した上位者とJackさんCashさんのノートブックを配布します。

➤データサイエンティスト特論（講師：電通大・原田慧、他）

- ✓ 実社会の実データを素材としたデータ分析演習
- ✓ 分析テーマの設定自体をデータに基づいて参加者が決定する
 - テーマ自体が設定されている「コンペ」とは異なる
 - データに基づかない「アイデアソン」とも異なる
- ✓ 参加者がテーマを設定する段階で、講師・T.A.のほか、データスポンサーの業務ご担当者が指導（過去の経験や、ビジネス的な筋の良し悪しなど）を行う
 - データスポンサーより大きめの問題（ex.ユーザエクスペリエンスの向上により収支を向上させる）をいただき、それを具体的な課題に落とし込む
 - 具体的課題の解決のためにどのようなデータ分析・モデル開発を行うかを決定
- ✓ そこで決定したデータ分析&モデル開発を行い、その出来栄えを課題落とし込みやソリューションの妥当性ととともに評価する

- **2017年**：アスクル様の売り上げデータ（中分類まで）
 - ✓ よりLTVの高いお客様（アスクル様社内の基準による）となつていただくための施策を考える。
 - ✓ 毎回、事業部長様が出席してご指導下さる。
- **2018年**：Kaggleの不動産データ分析のデータ
 - ✓ 不動産AIの先端企業であるG.A.テクノロジーズ様が指導担当
 - ✓ 毎回ご出席の上で、ご指導下さる。
- **2019年**：DeNA様の携帯ゲームデータ
 - ✓ ユーザーエクスペリエンス向上のための施策を考える
 - ✓ 当該ゲームのマネージャー様が毎回出席してご指導下さる。
- **2020年**：小田急電鉄様の乗降客数データ、小田急カードのデータ
 - ✓ コロナ後の鉄道事業の戦略を考える
 - ✓ 小田急カードのご担当者が毎回出席してご指導下さる。
- **2021年**：Jリーグのクラブチームのマーケティング・データ
 - ✓ 各クラブの収支向上の戦略を考える

➤ 修了要件として

- ✓ 以下の二つの講義のどちらかを履修して、所定の修了要件を満たしていること

➤ 医療データ分析の実践（講師：横浜市大・清水沙友里）

- ✓ 本講義では、レセプトデータ等の実際の医療データを用いて、テーマ立案からデータ分析までを徹頭徹尾実施し、分析ニーズが高まる医療データ分析について、実践形式で行う。

➤ データサイエンティストによるモック（MVP）作成（講師：ヘリクスデータ・赤間悟）

- ✓ 本講義ではエンドユーザの迅速なフィードバックを得ることを目的として、Streamlitなどを使うことでデータサイエンティスト単独で簡単に動くモックを作成する方法を紹介します。

➤ 修了要件として

- ✓ 以下の講義は、履修しなくても修了できますが、できるだけ多くの講義を試聴して知見を広げておくことが望ましいです。
- ✓ 他の講義で参照される場合もありますので、そうした場合には、必ず試聴しておくようにしてください。

➤ ブロック崩しを作ってPythonのオブジェクト指向を学ぶ

➤ 因果推論

➤ 深層学習の基礎

➤ 自然言語処理_基礎編

➤ 自然言語処理_応用編

➤ レコメンドモデル作成

➤ 教師なし学習

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）

- ✓ 統計検定2級（大学教養課程の「確率・統計」履修）レベルの知識がある。
- ✓ なんらかのプログラミング言語を使ったことがある。
- ✓ なんらかのデータ分析を行ったことがある。

➤ 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。
- ✓ 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。
- ✓ 4月からオンデマンド講義は受講可能なので、自身のない科目は基礎講座を受講して学んでください。（追加費用は不要です。）
 - オンデマンドの確率・統計講義はレベルが高いため、早めに着手してください。

➤ ディプロマ・ポリシー（修了条件）

- ✓ 9月末までに、必修のオンデマンド講義を修了していること。
- ✓ リアルタイムの講義に出席してチームに貢献or相応の成績を残していること。

6. 費用

➤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

➤DS基礎講座（3ヶ月コース）

- ✓ 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（6ヶ月コース）・春夏

- ✓ 対象：社会人
- ✓ 費用：50万円+税

➤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」の修了生で、特に許可を得たもの
- ✓ 費用：30万円+税