

2024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

2024年2月14日

電気通信大学

自己紹介

- 原田 慧(はらだ けい)

- D×2プログラム所属の教授

- 実務家教員

- ✓ 京都大学→名古屋大学(数理学博士)

- データサイエンティスト(12年)

- ✓ 23年3月まで普通に企業で働いていました

- Kaggle Master

- ✓ Kaggle: データサイエンスの競技会場

- ✓ 最適化関連が得意

- ✓ AtCoder青色(max黄)



アカデミアを
離れてみたら

博士、道なき道をゆく

岩波書店編集部 編

岩波書店

<https://www.amazon.co.jp/dp/4000614835>

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。

▶ 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。
 - Kaggle GrandmasterやKaggle Master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D-DRIVE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」
- ✓ 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：卒業生の目指す姿

➤ 口先じゃなくて「手が動く」ようになる

- ✓ 卒業生の方や、受講生を送り込んだ会社の方によく言われること
- ✓ 手が動くようになるまで反復する
- ✓ わかった気分では先に進めないレポート

➤ 多くの道具ではなくて、簡単な道具を練度高く使いこなせるようになる

- ✓ 知識を詰め込むのではなく、知識の「使い方」を学ぶ
 - ✓ 深層学習とかtransformerとかの講義はオプション
 - ✓ それよりも線形回帰や決定木を使いこなすことが大事
- ゲームで例えるなら・・・

「伝説の剣は持ってなくても、目の前の困っている村人のために、その辺に落ちている木の棒で戦える戦士」

➤ 2020年度受講生の感想

- ✓ <https://de.uec.ac.jp/entry2024/#voice>
- ✓ 創薬を博士後期課程で研究
 - 創薬に強化学習を適用するという新しい研究
 - 創薬ベンチャーに就職

- ✓ 機械学習の理論研究のために実践を学ぶ
 - 現在は本講座の講師

- ✓ その他、近日追加で公開予定
 - ◆ <https://dx2.inf.uec.ac.jp/index.html>

3. 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。

➤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	応用レベル	エキスパートレベル	
	コンピュータサイエンス	統計検定 2級自習	確率・統計	DE実践論
	Pythonプログラミング	回帰モデル作成	デザイン思考	Kaggle講義
	ITリテラシー	判別モデル作成	モック作成	DS特論
	Python100本ノック	深層学習の基礎	自然言語処理	因果推論
	EDAとモデル作成の基礎	教師なし学習	レコメンド	(自由選択)

文科省	リテラシーレベル	応用基礎レベル	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の監督と指示のもと業務を遂行できる。

DSプロジェクトにおいて、リーダーや専門家と適宜相談しながら、主担当としてデータ分析の中心部分を遂行できる。

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

応用レベル

エキスパートレベル

DS基礎

統計検定2級自習

回帰モデル作成

判別モデル作成

確率・統計

DE実践論

デザイン思考

Kaggle講義

モック作成

DS特論

独り立ち

自然言語処理

因果推論

レコメンド

(自由選択)

文科省

リテラシーレベル

DS協会

見習いレベル

DEFP

到達目標

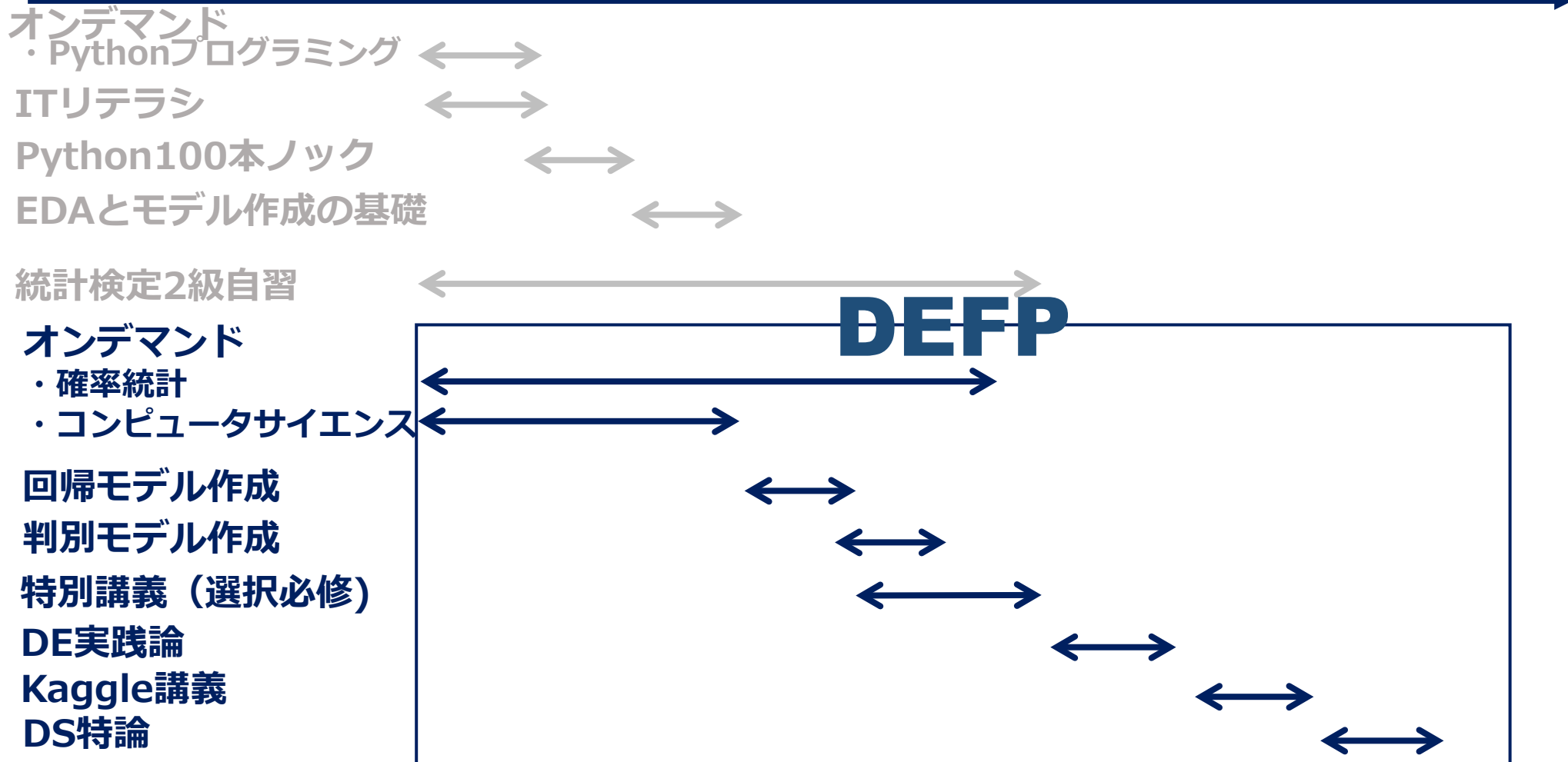
現在、データを活用して業務を行う上で必須とされるスキルを身につけた上で、データサイエンスのプロジェクトでリーダーの監督と指示のもと業務を遂行できる。

DSプロジェクトにおいて、リーダーや専門家と適宜相談しながら、主担当としてデータ分析の中心部分を遂行できる。

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2024・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

➤ 講義実施曜日：原則は土曜日午後（13-18時）

- ✓ オンデマンド講義 + 復習の時間は別途確保（週10時間程度）

➤ 必修講義

- ✓ オンデマンド講義
 - 確率・統計
- ✓ 対面学習（Zoom） 2024年度は7月～
 - 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
 - モック作成
 - 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

- **DEFPの受講生は、提供しているUEC DS Programs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。

4. 講座の内容

➤ITリテラシ

- ✓ 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

- ✓ 「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

- ✓ 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）（選択）**

- ✓ 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」とはどのようなことなのか、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）（必修）**

- ✓ 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)

- ✓ 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)

- ✓ 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

- ✓ 要求通りにものを作るのではなく、「新しい」モノやコトを提案するという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、考え方を、多くの事例や演習で学びます。

➤モック作成

- ✓ データサイエンスを活用の提案を実施する上で「こういうものを作ります」という試作品を作ることは、関係者のイメージ形成のために非常に重要です。
- ✓ クラウドサービスを活用して、実際に動くものを作ります。

➤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の話聞き、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事例を学びます。
- 【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

- ✓ 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Kagglers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、講師にコメントしてもらいます。
 - 講師の解法も解説してもらいます。
- ✓ ノートブックの共有
 - 発表した上位者と講師のコードを共有します

➤データサイエンティスト特論

- ✓ 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、ご指導くださる。
- **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くださる。
- **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くださる。
- **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

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を推奨します

- ✓ **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、自信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）・春

- ✓ 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

▶ インターン受け入れ企業

- ✓ DEFPコンソーシアム参加企業のうち、インターン受け入れの予定がある企業が受け入れを行います。
- ✓ 電通大の学生でない＝他大所属でも参加可能です。

▶ インターン内容

- ✓ 期間は比較的長期になります。：数ヶ月単位
- ✓ 業務内容：受け入れ企業の実際の業務においてデータ分析等のデータサイエンス業務に就いていただきます。
- ✓ アルバイト料・交通費等は受け入れ企業の条件によります。

▶ マッチング会

- ✓ 日時：2024年6月予定
- ✓ 方式：Zoomによるオンライン・ミーティング