

## 予測力の育成に向けた「データサイエンス活用」の 取組

メタデータ	言語: Japanese 出版者: Musashino University Smart Intelligence Center 公開日: 2024-03-25 キーワード: 予測力, 選択力, AIツール, オープンデータ, オンデマンド教材 作成者: 中村, 太戯留, 岡田, 龍太郎 メールアドレス: 所属:
URL	<a href="https://mu.repo.nii.ac.jp/records/2000196">https://mu.repo.nii.ac.jp/records/2000196</a>

特集論文

## 予測力の育成に向けた「データサイエンス活用」の取組

### A Case Study of Developing Prediction Skills in “Utilization of Data Science” Classrooms

中村太戯留<sup>\*1,2,3</sup>, 岡田龍太郎<sup>\*1,2</sup>

武蔵野大学 <sup>\*1</sup>スマートインテリジェンスセンター /

<sup>\*2</sup>データサイエンス学部 / <sup>\*3</sup>教養教育リサーチセンター

#### 概要

本稿では、「データサイエンス活用」という武蔵野大学の AI 副専攻科目における、予測力や選択力の育成に向けた取組みについて報告する。過去や現在のデータを用いて未来を見抜く予測力、複数の事柄の比較検討を通じて最適な事例を選ぶ選択力の基礎を、AI ツールを活用して反復的に学ぶ活動を実践した。特に、仮説を立てて検証するプロセスの学修時間を手厚くするために、オンデマンド教材を活用して知識修得は受講生のペースで進められるように、またオープンデータを活用して題材に多様性をもたせるように工夫した。学期末に授業方法に関するアンケートを実施したところ、各学修項目の 5 段階評価は受講前より受講後が高くなっており、この取組みの有用性が示唆された。

**キーワード：** 予測力, 選択力, AI ツール, オープンデータ, オンデマンド教材

#### 1. はじめに

文部科学省では、データサイエンスや AI を活用して課題を解決する実践的な能力の育成を推進している<sup>[1]</sup>。武蔵野大学では、学科における主専攻の学びに加えて、AI 副専攻（副専攻：AI 活用エキスパートコース）の学びを設定し、AI 活用人材の育成に取り組んでいる<sup>[2]</sup>。本稿で報告する「データサイエンス活用 1」と「データサイエンス活用 2」は、AI 副専攻のなかの専修科目に位置づけられている連続履修が想定された 2 科目である（以下、「データサイエンス活用」という表記はこの 2 科目の総称とする）。本稿では、「データサイエンス活用」における予測力と選択力を育成する取組みについての実践結果を報告する。

#### 2. 教育実践の方法

##### 2.1. 科目概要

「データサイエンス活用」では、AI ツールを活用して、過去や現在のデータを用いて未来を見抜く予測力、複数の事柄の比較検討を通じて最適な事柄を選ぶ選択力の基礎を学ぶ。

大量のデータに基づいて意思決定するために、これまで専門的な知識やプログラミングのスキルを長期間かけて学修した専門家が必要であった。それに対して、AI ツールの発展により AI の民主化が進み、非専門家が AI ツールを活用してその意思決定を実施できるようになってきている。そこで「データサイエンス活用」では、データサイエンスのクラウドサービス(DataRobot<sup>[3]</sup>)や表計算ソフト(Excel<sup>[4]</sup>)を活用したデータドリブンな意思決定のための基礎的な知識とスキルを、多様なオープンデータも利用して実践的に学ぶ。

表 1 各週の学修項目

Table 1 Study items for each week.

週	「データサイエンス活用 1」	「データサイエンス活用 2」
1	DataRobot の使い方	DS プロジェクトの仮説と検証
2	目標設定, ターゲットと特徴量, 可視化(1)	機械学習の性能評価のための用語解説
3	データ収集, データクレンジング, 可視化(2)	データの尺度水準とデータクレンジング
4	特徴量エンジニアリング, 可視化(3)	時系列データの分析
5	ミニプロジェクト(1)	ミニプロジェクト(1)
6	ミニプロジェクト(2)	ミニプロジェクト(2)
7	ミニプロジェクト(3) - 発表と相互評価	ミニプロジェクト(3) - 発表と相互評価

## 2.2. 実践方法

AI 副専攻の専修科目である「データサイエンス活用 1」はオンライン授業として 14 回分が開講されており (100 分/回), 半数の 7 回は同時双方向型として時間割で決められた時間に, 残りの 7 回はオンデマンド型として任意の時間にそれぞれ受講することが想定されている。学修内容としては, ターゲットと特徴量の関係性の可視化, 守りの前処理としてのデータクレンジング, 攻めの前処理としての特徴量エンジニアリング, そしてデータサイエンスの知識やスキルを活用したミニプロジェクトの遂行について, データサイエンスのクラウドサービスや表計算ソフトを活用して学んでもらった(表 1)。また, 知識として, データサイエンスや機械学習の基礎は, オンデマンド教材も併用して受講生のペースで学んでもらった。2023 年度は 3 学期に 2 クラスを開講しており, 合計 76 名が受講した。

データサイエンスの知識やスキルの定着を図るためには, AI ツールを活用して反復的に学ぶ活動が必要となる。そこで, オープンデータを活用して題材に多様性をもたせるように工夫した。具体的には, 「データサイエンス活用 1」の第 1 週において「貸し倒れ予測」<sup>[5]</sup>のデータを分析対象としてデータサイエンスのクラウドサービス(DataRobot<sup>[3]</sup>)の使い方を学んでもらった。第 2 週は「赤ワインの品質予測」<sup>[6]</sup>のデータを分析対象としてターゲットと特徴量の関係性の可視化について学んでもらった。第 3 週は「データサイエンティストの転職予測」<sup>[7]</sup>のデータを分析対象として守りの前処理としてのデータクレンジングについて学んでもらった。第 4 週は「航空会社の乗客の満足度予測」<sup>[8]</sup>のデータを分析対象とし

で攻めの前処理としての特徴量エンジニアリングについて学んでもらった。そして、第5週から第7週は「お弁当の需要予測」<sup>[9]</sup>のデータを分析対象としてデータサイエンスの知識やスキルを活用したミニプロジェクトの遂行について学んでもらった。すなわち、「データサイエンス活用1」においては、具体的に手を動かしてもらうことを通してデータサイエンスの知識やスキルの定着を狙いとした学びを実践した。

一方、「データサイエンス活用2」においては、仮説を立てて検証するプロセスの学修時間を手厚くすることを念頭に、反復的に学ぶ活動を継続しつつ、「データサイエンス活用1」で具体的に手を動かしてもらった学修項目と関連した理論の理解も狙った学びを実践した。学修内容としては、データの特徴や種類を確認し、ドメイン知識を活用して仮説を立て、そして結果から仮説を検証するというデータサイエンスの一連のプロセスについて、データサイエンスのクラウドサービスや表計算ソフトを活用して学んでもらった(表1)。問題の種類としては、二値や多値のカテゴリを予測する分類問題や、連続的な数値を予測する回帰問題を扱い、時系列データの分析も扱った。また、機械学習モデルの評価指標について学び、適切なモデルを選択したり特徴量を再設計したりする方法を学んでもらった。そして、データサイエンスの知識やスキルを活用したミニプロジェクトに挑戦してもらった。

反復的な学びにおいて活用したオープンデータとして、第1週と第2週は「タイタニックの乗客の生存予測」<sup>[10]</sup>のデータ、第3週は「レストランの新店の売上予測」<sup>[11]</sup>のデータ、そして第4週から第7週は「雑誌に掲載された漫画のアニメ化予測」<sup>[12]</sup>のデータを分析対象とした。なお、2023年度は4学期に2クラスを開講しており、合計75名が受講した。

### 3. 教育実践の結果

本稿では、学期末に授業方法に関するアンケートを実施した結果を検討した。回答にはMicrosoft Formsを用い、各項目に対して「そう思わない」の1と「そう思う」の5を両端とする5件法で回答してもらった。「データサイエンス活用1」(DS1)では、2クラスの合計66名(回答率:86.8%)から得たデータを検討対象とした。

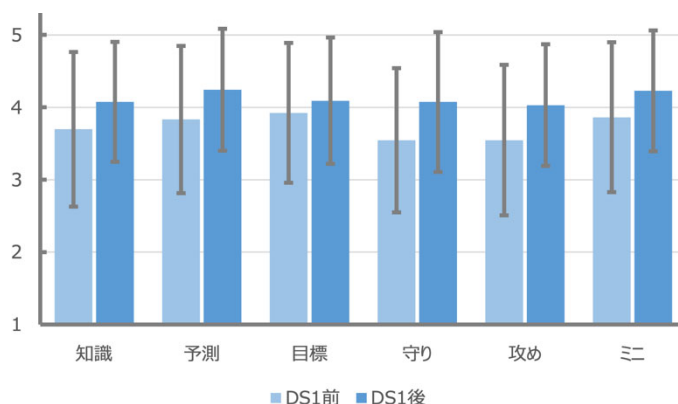


図1 DS1の学修状況の受講前後の推移

Figure 1 Changes in academic progress before and after DS1 classes.

表 2 DS1 の授業方法に関するアンケートの結果

Table 2 Results of questionnaire on teaching methods for DS1 classes.

項目 (1~6 は受講前に関する設問, 7以降は受講後に関する設問)	平均値	SD
1. オンデマンド教材を用いてデータサイエンスに関する事例や知識を学ぶことに関心があった	3.70	1.067
2. データサイエンスツールを活用して, モデルを生成し, 予測結果を取得する方法を学ぶことに関心があった	3.83	1.017
3. データサイエンスで実施するための目標設定(テーマや達成レベル)を学ぶことに関心があった	3.92	0.966
4. データクレンジング(守りの前処理)の方法を学ぶことに関心があった	3.55	0.995
5. 特微量エンジニアリング(攻めの前処理)の方法を学ぶことに関心があった	3.55	1.040
6. データサイエンスの一連の流れについてミニプロジェクトを通して学ぶことに関心があった	3.86	1.036
7. オンデマンド教材で学んだデータサイエンスに関する事例や知識をミニプロジェクトなどで活用することができる	4.08	0.829
8. データサイエンスツールを活用して, モデルを生成し, 予測結果を取得することができる	4.24	0.842
9. データサイエンスの目標設定においてインパクトと実現可能性の検討にもとづいてテーマ等を検討することができる	4.09	0.872
10. データサイエンスツールや表計算ソフトを活用して, データクレンジング(守りの前処理)を実施することができる	4.08	0.966
11. データサイエンスツールや表計算ソフトを活用して, 特微量エンジニアリング(攻めの前処理)を実施することができる	4.03	0.841
12. データサイエンスの一連の流れを理解してミニプロジェクトを推進することができる	4.23	0.837
13. 最終課題でこの科目で学んだことを活用することができた	4.29	0.873
14. 最終課題で満足のいく結果を出すことができた	3.98	0.903
15. グループメンバーと協力して最終課題を進めることができた	4.44	0.862
16. 「貸し倒れの予測」というテーマに関心をもって取り組むことができた	4.17	0.887
17. 「赤ワインの品質予測」というテーマに関心をもって取り組むことができた	4.18	0.875
18. 「データサイエンティストの転職予測」というテーマに関心をもって取り組むことができた	4.15	0.808
19. 「航空会社の乗客の満足度予測」というテーマに関心をもって取り組むことができた	4.35	0.774
20. 「お弁当の需要予測」という(ミニプロジェクトの)テーマに関心をもって取り組むことができた	4.30	0.877
21. 講師のほかに担任講師がいる方式はサポートやアドバイスを受けるのに役立つと思う	4.35	0.936
22. 講師や担任講師のほかにSA(Student Assistant)がいる方式はサポートやアドバイスを受けるのに役立つと思う	4.35	0.886
23. SAの授業時間外サポートは講師や担任講師には聞きづらい相談をしたりアドバイスを受けたりするのに役立つと思う	3.80	1.268

具体的な項目および5件法の平均値と標準偏差(SD)は表2に示す通りで, 中央値は4であった(項目4と5の中央値は3, 項目15と19~22の中央値は5). 受講前(項目1~6)より受講後(項目7~12)が高いため( $p < 0.05$ ), 学修目標を概ね達成できたと考えられる(図1). 具体的なテーマ(項目16~20)も関心をもって取り組んだことが確認できた. すなわち, 第1週から第3週で分析対象とした「貸し倒れ予測」<sup>[5]</sup>のデータ, 「赤ワインの品質予測」<sup>[6]</sup>のデータ, そして「データサイエンティストの転職予測」<sup>[7]</sup>のデータの関心の平均値は4.1程度であった. それに対して, 第4週から第7週で分析対象とした「航空会社の乗客の満足度予測」<sup>[8]</sup>のデータや「お弁当の需要予測」<sup>[9]</sup>のデータの関心の平均値は4.3程度とやや高い結果となっており, 受講生が強い関心をもって学修に臨んだことを確認できた.

表 3 DS2 の授業方法に関するアンケートの結果

Table 3 Results of questionnaire on teaching methods for DS2 classes.

項目 (1~7は受講前に関する設問, 8以降は受講後に関する設問)	平均値	SD
1. データサイエンスツールを使って, モデル生成, 特徴量の検討, 予測までの一連のプロセスを <b>実行</b> することができる	3.49	1.134
2. 表計算ツールを使って, 特徴量エンジニアリング( <b>攻め</b> の前処理)や精度の計算をすることができる	3.62	1.211
3. データサイエンスを行う際に <b>仮説</b> を立てて, その仮説を表現する特徴量について説明することができる	3.56	1.175
4. 機械学習モデルによって <b>予測</b> を行うということの概要について理解している	3.40	1.264
5. 分類と回帰の <b>違い</b> について理解している	2.89	1.345
6. 混同行列と <b>精度</b> の指標について理解している	3.05	1.337
7. データサイエンスの一連の流れについて <b>理解</b> している	3.43	1.241
8. データサイエンスツールを使って, モデル生成, 特徴量の検討, 予測までの一連のプロセスを <b>実行</b> することができる	4.11	0.764
9. 表計算ツールを使って, 特徴量エンジニアリング( <b>攻め</b> の前処理)や精度の計算をすることができる	4.21	0.722
10. データサイエンスを行う際に <b>仮説</b> を立てて, その仮説を表現する特徴量について説明することができる	4.05	0.750
11. 機械学習モデルによって <b>予測</b> を行うということの概要について理解している	4.05	0.705
12. 分類と回帰の <b>違い</b> について理解している	3.51	0.965
13. 混同行列と <b>精度</b> の指標について理解している	3.75	0.879
14. データサイエンスの一連の流れについて <b>理解</b> している	4.08	0.768
15. ミニプロジェクトで本科目で学んだことを <b>活用</b> することができた	4.25	0.915
16. ミニプロジェクトで満足のいく <b>結果</b> を出すことができた	3.75	0.933
17. グループメンバーと <b>協力</b> して最終課題を進めることができた	4.10	1.011
18. ミニプロジェクトのテーマは興味( <b>関心</b> )を持って取り組めるものであった	4.05	0.869
19. 講師のほかに <b>担任</b> 講師がいる方式はサポートやアドバイスを受けるのに役立ったと思う	4.40	0.685
20. 講師や担任講師のほかに <b>SA</b> (Student Assistant)がいる方式はサポートやアドバイスを受けるのに役立ったと思う	4.35	0.722
21. SAの授業 <b>時間外</b> サポートは講師や担任講師には聞きづらい相談をしたりアドバイスを受けてたりするのに役立ったと思う	3.57	1.058

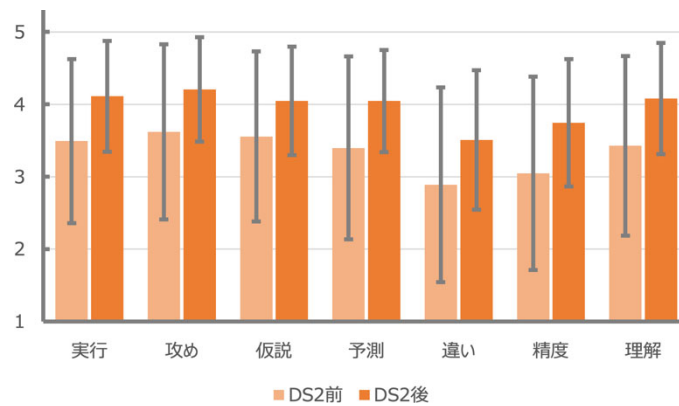


図 2 DS2 の学修状況の受講前後の推移

Figure 2 Changes in academic progress before and after DS2 classes.

また、「データサイエンス活用 2」(DS2)においても同様にしてアンケートを実施し、2クラスの合計 63 名(回答率: 84.0%) から得たデータを検討対象とした。具体的な項目および 5 件法の平均値と標準偏差(SD)は表 3 に示す通りで、中央値は 4 であった(項目 4~6 の中央値は 3, 項目 19 の中央値は 5)。受講前(項目 1~7)より受講後(項目 8~14)が高いため( $p < 0.05$ ), 学修目標を概ね達成できたと考えられる(図 2)。ミニプロジェクトのテーマ(メディア芸術データベースを利用した「雑誌に掲載された漫画のアニメ化予測」<sup>[12]</sup>のデータ)も関心をもって取り組んだことが確認できた。

次に、ミニプロジェクトの状況(表 2 の項目 13~15 と 20, 表 3 の項目 15~18)および授業方法(表 2 の項目 21~23, 表 3 の項目 19~21)に関して両科目間で比較したところ、大きな差は認められなかった(図 3)。ただ、ミニプロジェクトで満足のいく結果を出せたかと、SA の授業時間外サポートは両科目で他の項目よりもやや低いため対策を講じる必要がある。

ミニプロジェクトについては、ミニプロジェクトに費やすことができる期間が短かった可能性が考えられるため、もう少し早めの週から準備を始められるようにするなどの対策を講じる必要がある。また、AI 副専攻では多様な学科の受講生が参加しており、授業時間外のグループワークにおいて、グループメンバーと時間を合わせるのが大変だった可能性が考えられるため、授業時間内のグループワークの時間をもう少し増やすなどの対策を講じる必要がある。さらに、オンライン授業でのグループワークに不慣れであった可能性も考えられるため、グループワークの実施方法に関する指導を手厚くするとともに、毎回の授業においてアイスブレイクの時間を明示的に指示するなどの対策を講じる必要がある。

授業時間外の SA サポートについては、受講生が利用したい時間と SA が活動できる時間にズレがあった可能性が考えられるため、受講生のニーズに合わせて時間調整する仕組みを考案するとともに、授業時間内のサポートを手厚くするなどの対策を講じる必要がある。

これらを合わせて考えると、同時双方向型のオンライン授業の時間をもう少し増やして、授業時間内のグループワークや SA サポートを手厚くしていく必要があると考えられる。

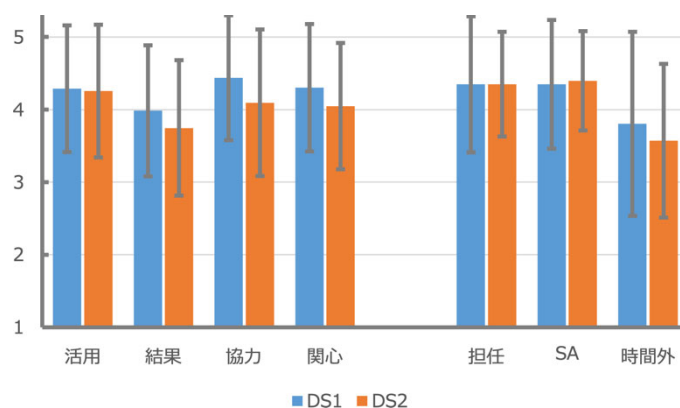


図 3 DS1 と DS2 のミニプロジェクトの状況比較(左)および授業方法の比較(右)  
Figure 3 Comparison of DS1 and DS2 mini-project situations and teaching methods.

#### 4. おわりに

本稿では、「データサイエンス活用」という科目において実践した予測力および選択力の育成に向けた取り組みについて報告した。すなわち、過去や現在のデータを用いて未来を見抜く予測力、複数の事柄の比較検討を通じて最適な事例を選ぶ選択力の基礎を、AI ツールを活用して反復的に学ぶ活動を実践した。特に、仮説を立てて検証するプロセスの学修時間を手厚くするために、オンデマンド教材を活用して知識修得は受講生のペースで進められるように、またオープンデータを活用して題材に多様性をもたせるように工夫した。

受講生を対象とした学期末のアンケート調査の結果によると、各学修項目の5段階評価は受講前より受講後が高くなっており、この取り組みの有用性が示唆された。また、各科目で取り扱った具体的なテーマ(表2の項目16~20, 表3の項目18)も関心をもって取り組んだことが確認できた。講師のほか担任講師やSAがいる方式も評価されていた。ただ、ミニプロジェクトで満足のいく結果を出せたかの評価がやや低いことについては、ミニプロジェクトに費やすことができる期間が短かった可能性などを考慮し対策を講じる必要がある。SAの授業時間外サポートの評価がやや低いことについては、受講生が利用したい時間とSAが活動できる時間にズレがあった可能性などを考慮し対策を講じる必要がある。

2023年度は、同時双方向型のオンライン授業を7回とオンデマンド型のオンライン授業を7回という組み合わせで「データサイエンス活用1」の授業を構成し、続く「データサイエンス活用2」も同様の組み合わせで構成した。本稿での検討結果の同時双方向型をもう少し増やす必要性を考慮し、2024年度以降は、「データサイエンス活用1」の授業を同時双方向型のオンライン授業を10回とオンデマンド型のオンライン授業を4回という組み合わせに改良することを基軸として調整し、続く「データサイエンス活用2」も同様の組み合わせに調整することを試みていく。このような調整により、授業時間内のグループワーク活動の時間が増加してミニプロジェクトを進めやすくなることや、教員やSAからのサポートの時間が増加することで授業全体を通して質問や相談がよりしやすくなることが期待される。

**謝辞** 本科目の実施に際して、毎回の授業参加と学期末のアンケートの回答をした受講生、授業を実施した講師、並びに関係者の皆さまに、心より感謝の意を表する。

#### 参考文献

- [1] 数理・データサイエンス・AI教育プログラム認定制度(応用基礎レベル) : [https://www.mext.go.jp/a\\_menu/koutou/suuri\\_datascience\\_ai/00002\\_00003.htm](https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002_00003.htm) (参照 2024-2-9)
- [2] 副専攻(AI活用エキスパートコース) : <https://risyuyouran.musashino-u.ac.jp/faculty/curriculum-faculty/ai/> (参照 2024-2-9)
- [3] DataRobot : <https://www.datarobot.com/jp/> (参照 2024-2-9)
- [4] Microsoft Excel : <https://www.microsoft.com/ja-jp/microsoft-365/excel> (参照 2024-2-9)

## 付録

「データサイエンス活用 1」において分析対象として用いたデータ :

- [5] 貸し倒れ予測 : [https://community.datarobot.com/iuvcf86456/attachments/iuvcf86456/jp07c/74/1/P2PLendData\\_JP.zip](https://community.datarobot.com/iuvcf86456/attachments/iuvcf86456/jp07c/74/1/P2PLendData_JP.zip) (参照 2024-2-9)
- [6] 赤ワインの品質予測 : <https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009/> (参照 2024-2-9)
- [7] データサイエンティストの転職予測 : <https://www.kaggle.com/datasets/arashnic/hr-analytics-job-change-of-data-scientists> (参照 2024-2-9)
- [8] 航空会社の乗客の満足度予測 : <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction> (参照 2024-2-9)
- [9] お弁当の需要予測 : <https://signate.jp/competitions/24> (参照 2024-2-9)

「データサイエンス活用 2」において分析対象として用いたデータ :

- [10] タイタニックの乗客の生存予測 : <https://www.kaggle.com/competitions/titanic> (参照 2024-2-9)
- [11] レストランの新店の売上予測 : [https://community.datarobot.com/iuvcf86456/attachments/iuvcf86456/jp07c/74/3/New\\_Restaurants.zip](https://community.datarobot.com/iuvcf86456/attachments/iuvcf86456/jp07c/74/3/New_Restaurants.zip) (参照 2024-2-9)
- [12] 雑誌に掲載された漫画のアニメ化予測 : [https://mediag.bunka.go.jp/madb\\_lab/lod/ds/](https://mediag.bunka.go.jp/madb_lab/lod/ds/) (参照 2024-2-9)