

論文

文系向けデータサイエンス教育の在り方に関する一考察 — 山口県立大学における全学基盤教育での実践事例を通して —

阿部 真育¹
ABE Maiku

山口県立大学 国際文化学部 国際文化学科¹
Department of Intercultural Studies, Yamaguchi Prefectural University

要 旨

現在、日常生活において、ビッグデータが容易に取得できる時代となっており、データサイエンスや機械学習といったキーワードが既に目新しいものではない。大学教育機関並びに民間企業や官公庁においても、データサイエンスに関する学部や部署が設立されており、様々な教育プログラムが展開されている。既に標準レベルのデータサイエンス教育の基盤が構築されつつある今般において、今まで主に大都市圏の理工系教員中心に進められてきたデータサイエンス教育を、地方や文系にまで展開する動きが生じ始めている。地方公立大学における文系向けデータサイエンス教育の実践は、国家的にも課題の抽出や解決策の提案が急務とされている。山口県立大学では、2022年度よりデータサイエンス概論というデータサイエンスの素養を身につけることを目的とした講義を全学1年生向けに開始した。本論文では、2022年度の本講義において学生に課した課題の提出状況のデータ分析を通して、文系向けのデータサイエンス教育の在り方について検討を行った。その結果、データサイエンス教育をほぼ初めて受講する文系学生が、データ分析の演習時に感じる困難度と課題に対する積極度合の可視化可能性が示された。

Abstract:

Today, big data can be easily obtained in our daily lives. As a result, keywords such as data science and machine learning are no longer new. Universities, private companies, and government agencies have established departments and faculties related to data science, and various educational programs are being developed. The implementation of data science education for liberal arts at regional public universities is an urgent national issue to be identified and solutions to be proposed. Yamaguchi Prefectural University started a lecture course called Introduction to Data Science for all first-year students in FY2022 with the aim of acquiring data science backgrounds. In this paper, we examine the state of data science education for students in the humanities by using data on the submission status of assignments given to students in this course in FY 2022. The results showed the possibility of visualizing the degree of difficulty that students in the humanities, who are taking data science education for the first time, experience during data analysis exercises and the degree of their positive attitude toward the assignments.

はじめに

1974年にPeter Naurが提唱したデータサイエンス（以下、DS）という学問分野¹は、人工知能の冬の時代を打開する大量のビッグデータの蓄積により、企業等の実務の分野に急速に拡大した。国内においても、内閣府の統合イノベーション戦略推進会議が

2019年に公開した『AI戦略2019』において、データサイエンティストの2025年度までの年間育成人数の提言がなされた。その実現に向けて、主要国立大学や企業により、教育機関側が提供すべきモデルカリキュラムや学習者が修得すべきスキルセットの構築がなされている。整理されたモデルカリキュラム

やスキルセットに基づき、文部科学省が主幹となって『数理・データサイエンス・AI教育プログラム認定制度²（以下、DS認定制度）』が2020年度から開始されており、現在リテラシーレベルと応用基礎レベルのDS認定制度の申請が受け付けられている。今後さらに、エキスパートレベルのDS認定制度も追加されると考えられる。2021年度は、数理・DS教育に関する文部科学省補助事業の5年間の検討の区切り目であり、2022年度からは新たなフェーズに移行し³、AI戦略2022の公開もなされた⁴。DSのリテラシー教育に関しては、以前より文理を問わないことが記載されていたが、先の5年間においては、DS教育における応用基礎レベルやエキスパートレベルの方向性を示すことが第一義的な目的となっており、主に国立大学理系向けの対応が行われていた。しかしながら、新たなフェーズの一つの方向性に、文系向けや地方国公立大学でのDS教育の推進が挙げられており、これまでのノウハウを踏襲した形で、文系学生や企業の約7割を占めると言われる文系学部出身の社会人に対する、DS教育の取組が推進されつつある。実際、その機運は地方大学等でのDS関連学部・学科の設立等により明示的になっている⁵。学部・学科が設立される分野は、社会科学や地域振興に関わる学部を中心に進められているものがほとんどであるが、既存の学部の定員減少や廃止による学部・学科新設であり、これまで各学部で行われていた教育内容の転用による対応が予想され、国の求めるDSのリテラシーレベルに合致したものを、大学組織として提供しようとしているか否かについては、不透明な部分も多い。

そのような状況に陥っている理由としては、国や地方自治体からの補助金獲得を狙った性急な成果の見える化といった、大学の経営戦略的な要素も大きい。未だにDSが理工系の分野であると考えている教育機関関係者が多く、そのような環境が文系教員や文系学生のDSやAIに対する理解不足を加速し、DS教育の展開への消極性を助長していると考えられる。その兆候は、文部科学省や大学等が主催で開催されるDS関連のシンポジウムに招聘される企業が極めて理工学系のITベンチャーであることが多いことから垣間見える。筆者が参加した、数理・データサイエンス・AI教育強化拠点コンソーシアム主催のシンポジウムに講演者として招聘されたベンチャー企業の代表取締役が、文系出身の学生より

も理工系の学生の方を、自身の企業は求めているという発言もしており、DS先進企業として教育機関側が認知している企業の偏向性が如実になってきている事例と言える。一方で、一部上場企業等でのリカレント教育は、2022年度は約490億円を見込まれるほど加速しており、その中でもDSを含むIT・デジタル系人材育成に関する教育プログラムへの投資が顕著という報告もある⁶。リテラシーレベルのDSスキルを社会人が保有していることは、もはや常識になりつつある。海外においては、大学の基盤教育とも言えるリベラルアーツの一部に、文系理系問わず統計学等のデータサイエンスに関連した教育が組み込まれており、日本の教育機関の文系理系の垣根の存在自体が、DSスキルの早期修得の阻害要因の一つと考えられる。

研究背景

山口県では、2020年12月に閣議決定された国の『デジタル社会の実現に向けた改革の基本方針⁷』等に基づき、2021年3月に『やまぐちデジタル改革基本方針⁸』を公開している。その中において、デジタル技術を活用した施策の企画・立案・実施等を行う専門人材である、デジタル人材の確保や育成の必要性についても記載されている。しかしながら、大学教育との連携の方向性等の記述はほとんどなく、自治体職員の教育の場を設けるのみの記述に概ね留まっており、若年層の育成についての課題抽出自体も未だ進められていないように見受けられる。

山口県立大学は、山口大学と山口学芸大学と連携し、文科省の地域活性化人材育成事業（SPARC：Supereminent Program for Activating Regional Collaboration）に2022年度から採択され、文系DX人材の育成を行うことを提言し⁹、高大接続からリカレント教育に至るまでの教育改革の推進について検討を始めている。

本論文では、前述の情勢下において、2022年度から山口県立大学にて開始され、筆者が担当したDS概論の講義内容並びに、その講義に対する学生の課題対応状況の分析等を通して、文系向けのDS教育の在り方について考察を行う。なお、本講義は、文系向けDS教育として、山口県立大学にて2022年度から全学部一年生に対して新たに開始された講義であり、今後の継続的なデータ収集により、これまであまり蓄積がなされていない文系向けDS教育の効

果の可視化が期待される。しかしながら、本論文にて対象とする科目は、DS概論の2022年度実施分の1講義のみであるため、本論文の目的は、本講義にて蓄積されたデータの一次分析の結果の提示及び考察を通して、文系向けDS教育の在り方について一提言をするに留める。

DS概論の特徴

DS概論は、2022年度山口県立大学入学の全学部学生必修科目（2022年度受講者数329名）であり、前期に開講された。授業の各回のシラバス概要とDS認定制度との対応関係について表1に示す。モデルカリキュラムの対応箇所に記載してある数字の具体的内容については、紙面の関係上記載は省略するが、DS概論の講義内容は、DS認定制度リテラシーレベルに求められている学習内容を網羅的に含有しており、更に応用基礎レベルの一部にも対応している構成とした。従って、全国的なDS教育標準を鑑みた形で、網羅的なDSに関する知識の修得か

ら、自身の専門分野に対するDSの適応可能性の想像にも繋げる体験を、講義や複数回の演習を通して、全学部学生が1年生の内に経験できる構成になっている。

しかしながら、DS認定制度リテラシーレベルで求められている内容全てを完全に修得することは、例え理工系分野の学生であっても極めて困難であり、DS認定制度との対応が良いからと言って、必ずしも優れたDS教育が実施できているとは限らない。そこで本論文では、第10回を除き講義内で毎回実施された課題への取組状況の分析により、提供したDS教育の難易度の妥当性検討、並びにある学科に着目して、DS教育に対する学生の反応の傾向把握を行う。

用いるデータと分析手法

本論文に用いるデータは、2022年度のDS概論の第10回を除いた全ての講義において実施した、課題のデータである。特に課題提出期限と実際の提出日

表1 DS概論の講義内容とDS認定制度モデルカリキュラムとの対応

講義回数	講義概要	学習目標	【モデルカリキュラム対応箇所】 リテラシーレベル	【モデルカリキュラム対応箇所】 応用基礎レベル
1回目	ガイダンスとDS分野を巡る政策的な現状	DS分野を巡る国内外の流れを理解する	1-1,1-3,1-5,1-6	1-1
2回目	人工知能（Artificial Intelligence：AI）とは		1-1,1-3,1-5,1-6	1-1
3回目	社会実装を目指したDSの最前線		1-1,1-3,1-5,1-6	1-1
4回目	DSプロセス①	課題の抽出から社会実装までのプロセスについて理解する		1-1,1-2
5回目	DSプロセス②			1-1,1-2
6回目	DSに関わるプライバシー保護の基礎		3-1,3-2	
7回目	Business Intelligenceツール（BIツール）	データの可視化や経営判断のツールとして用いられるBIツールやエクセルを用いたデータ分析手法について理解する	2-1,2-2,2-3	1-3,1-5
8回目	BIツールによるオープンデータの可視化【演習回】		2-1,2-2,2-3	1-3,1-5
9回目	空間情報の取扱い【演習回】		2-1,2-2,2-3	1-3,1-5
10回目	エクセルでDS体験①【演習回】		2-1,2-2,2-3	1-3,1-5
11回目	エクセルでDS体験②【演習回】		2-1,2-2,2-3	1-3,1-5
12回目	エクセルでDS体験③（Pythonプログラミング体験含む）【演習回】		2-1,2-2,2-3	1-3,1-5
13回目	エクセルでDS体験④（Pythonプログラミング体験含む）【演習回】		2-1,2-2,2-3	1-3,1-5
14回目	DS利活用の現場（ゲストスピーカー1）	社会でDSがどのように活用されているのかについて知り、DSに関するキャリアパスについて具体的なイメージを持つ	1-5	1-1
15回目	・DS利活用の現場（ゲストスピーカー2） ・これまでの振り返り		1-5	1-1

の差分に着目し、課題への取組の積極性の可視化を試みる。課題は、各講義にて疑問に感じたことを記述する箇所とその疑問点に関して自身で調査して記述する箇所の大きく2部構成となっており、事後学習の促進に寄与する仕組みとなっている一方で、インターネットに記載してある内容の転記のみでも対応可能な内容となり、学生の課題への積極性は、如実に提出日に反映される。なお、DSの最新の情報は、書籍よりも明らかにインターネットの方が豊富であることから鑑みて、引用部分の明記を条件として、インターネット上の情報の転記については、特に制約を設けない旨を講義内で学生に伝えている。

上記の状況を鑑みて、本論文では課題への積極性を定義する上で、課題に対してどの程度の時間をかけたかに着目し、その程度を課題への積極性とした。その定義を原則とした上で、本論文におけるデータ分析結果に基づいて、課題への積極性の再定義を行う。これは、学生への授業アンケート等を介して把握を試みられている習熟度やコンピテンシーとはアプローチが異なっており、客観的な学生の学習意欲の測定可能性と多様な授業への応用可能性を秘めている。

なお、課題内容の研究利用に関しては、学生に対

する講義内での許諾状況を鑑みて、本論文においては、飽くまで課題の提出日のみに着目して積極性を測ることとし、課題の具体的な記述内容等に関しては取り扱わないこととする。また、当初提出締切日に誤った課題を提出していた学生においても、ほとんどの学生が再提出依頼に対して迅速に対応していたことを鑑みて、当初提出日のデータを分析には用いることとした。

結果と考察

本論文に用いるデータのサンプルサイズは、3学部5学科の1年生で、329サンプルである。5学科の具体名称の記載は控えるが、本論文ではA学科、B学科、C学科、D学科、E学科とし、それぞれ69サンプル、59サンプル、104サンプル、56サンプル、42サンプルとなっている。課題提出日からの差分を学生毎に算出し、各講義回別に整理したものを図1に示す。図1において、課題提示時に設定した課題提出日より前に課題の提出が成された場合、差分が負の値となるようにしている。1回目から3回目に関しては、LMS (Learning Management System) 上での課題提出に関して学生が不慣れな部分もあり、提出締切日に関して柔軟な対応を行っ

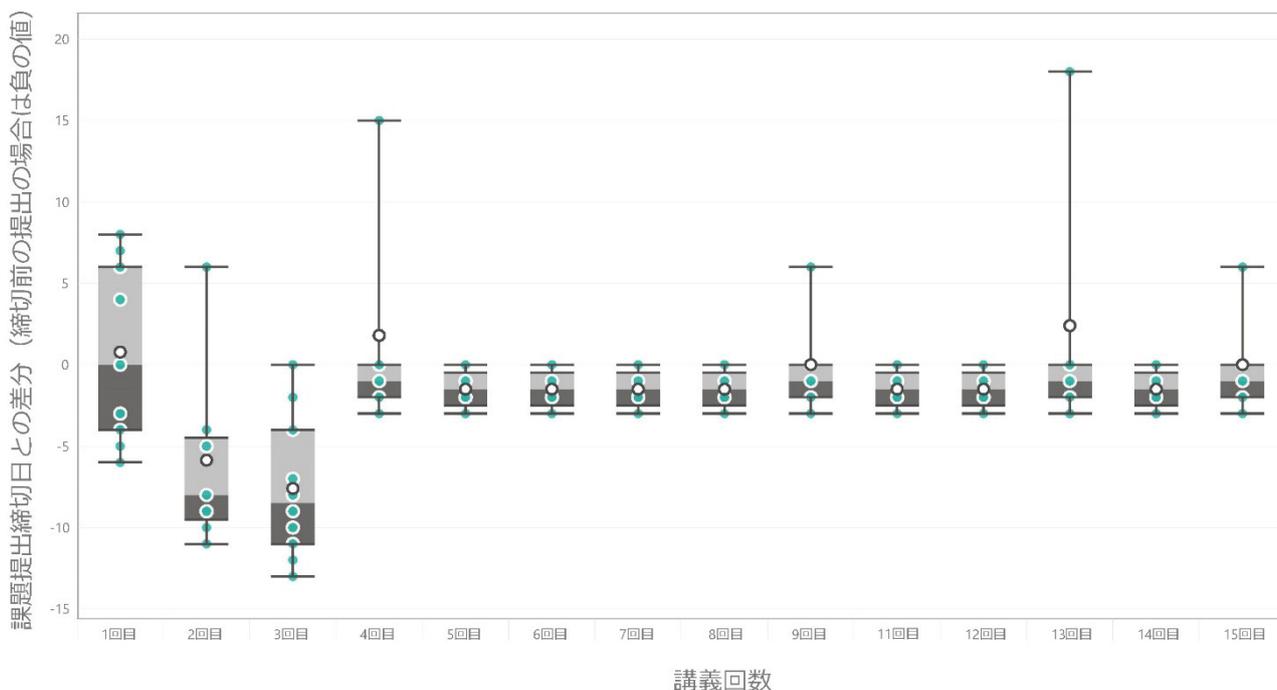


図1 講義毎の課題提出締切日と学生が課題提出した日の差分

※課題提示時に設定した課題提出日より前に課題の提出が成された場合、差分が負の値となる。4回目の講義以降は提出締切日(講義実施日の週の金曜)を厳格化したため、箱ひげ図の最小値は、講義実施日当日(火曜)提出となるため、-3となる。

た。そのため、提出日の分散が大きくなっている。一方、4回目の講義以降は提出締切日（原則、講義実施日の週の金曜）を厳格化したため、箱ひげ図の最小値は、講義実施日当日（火曜）提出となるため、-3となる。なお、正の値を示している箇所に関しては、LMSを介さない課題の提出等に対して、筆者自身がLMSに登録した日のデータが反映されている。

従って、本論文では、学生がLMS等のシステムにある程度慣れ、筆者自身が課題登録を行う等の例外的な対応の場合を除いた、4回目以降の課題提出が当日提出から締切日提出（-3～0）の学生のデータのみに着目する。因みに、10回目は課題なしとしたため除いている。

対象とするデータ数を整理したものを表2に示す。この表の各回平均より、15回目のゲストスピーカーの講演回における課題の当日提出が最大値を示していることを考慮したとしても、課題提出日は講義当日と締切当日が高い値を示していることが分かる。一方で、8回目から13回目は演習回であるが、エクセルを用いた演習回で最初の課題が出された11回目の課題提出なしの数値が他の回に比べて高くなっており、さらに締切日提出の数値が、11回目から13回目にかけて純増している。演習回に関しては、表計算ソフトウェアの処理等に慣れていない学生も多くいるため、講義内での細かな作業指示等に掛かる時間が、講義のみの場合と比べて極めて多くなるが、前述した締切日提出の数値の純増と併せて、課題提出なしの数値の純減が見られることから鑑みて、学生の課題への積極性が顕著に表れる講義形式とも考

えられる。

学生個人の課題への積極性を把握する上では、時系列で個人の動向を観察することが重要となる。そこで、本論文では、サンキーダイアグラム（平行プロット図や沖積図と呼ぶ場合もある）によるデータの可視化を通して、学生の課題への積極性の把握を試みた。なお、図の作成にあたり、Google Colaboratory Python Ver. 3.8.16にてプログラミングを行い、ライブラリのPlotly Ver.5.5.0を用いて可視化を行った。

図2に各学科で色分けをした4回目以降の課題提出日のサンキーダイアグラムを示す。この図において、講義回数間を繋いでいるラインは各学生の動向を示しており、講義回毎にどのタイミングで課題を提出したかの推移を把握することができる。また、前述した『課題提出なし』に関しては、赤枠にて示している。赤枠の部分に着目すると、課題提出がなかった学生は、特定の学科に属しているわけではなく、複数の学科により構成されていることが分かる。また、課題提出日に着目すると、常に『講義当日提出（-3）』であるといったように、常に同じタイミングで提出している学生が複数名いることが把握できる。一方で、講義回毎に課題の提出タイミングが異なる学生も存在し、このような学生は、課題に対して柔軟に対応していると考えられる。常に講義当日提出の学生の中には、講義時間中、或いは講義終了直後に課題を提出している学生も見られたため、そのような学生に比べて、提出タイミングが講義回毎に異なる学生は、課題への取組の積極性が高い可能性が考えられる。一方、学科単位でサン

表2 全学部学生の各講義回の課題提出状況

講義回数	講義当日提出(-3)	講義翌日提出(-2)	締切前日提出(-1)	締切当日提出(0)	課題提出なし	合計
4回目	91	64	61	97	15	328
5回目	97	65	67	78	22	329
6回目	110	61	64	73	19	327
7回目	109	63	65	70	22	329
8回目	78	59	91	79	22	329
9回目	92	51	86	72	27	328
11回目	59	50	91	80	49	329
12回目	66	63	63	98	39	329
13回目	52	38	101	113	24	328
14回目	97	40	75	76	41	329
15回目	135	44	56	64	29	328
各回平均	89.6	54.4	74.5	81.8	28.1	

キーダイアグラムを見ることで、より学生単位の推移が確認できるようになる。可読性を考慮して、サンプルサイズが最も小さいE学科のみを抽出したサンキーダイアグラムを図3に示す。この図において、線が密であるか疎であるかが、その学科の課題取組に対する積極度合いの多様性を示していると考えられ、6回目から13回目まで、継続的に『締切当日提出(0)』である学生が密に見られる。筆者が課題評価を行うに当たり、E学科の学生は締切当日に提出する学生が多い印象は強かったが、実際にその感覚と図3の結果は合致しており、実際に提出された課題の内容も、深く調べられているものが多かった。

今後の課題

本論文により得られた結果より、どのような講義においても取得可能な課題の提出日の推移データに着目することで、試行的ではあるが、学生の講義に対する積極性の把握や講義自体の難易度の検討材料の可能性まで示すことができた。しかしながら、表2及び図2、図3の結果は飽くまで1年度のみの数値であるため、課題への積極性を向上させる講義形

態の議論や積極性と課題提出日との関連性を示すためには、同精度の継続的なデータ蓄積が必要と考えられる。

但し、同精度の継続的なデータの蓄積を行うためには、課題内容の精査に加えて、課題データの収集とデータ前処理の省力化が必要となる。その際、LMSの改良が必要となるが、改良を行う上で、分析方針等を事前に検討していないと、単に市販のシステム或いは分析ツールを分散的に導入して対応することとなり、結果的に作業量が増えて、逆にデータの精度が落ちる事案も教育機関においては散見される。システムの改良を行う上では、分析手法までの詳細検討も含めたシステム活用の方針を先に定める必要がある。

また、学生の習熟度や積極性を測ることに注力しすぎて課題内容自体を充足させすぎると、課題提出のタイミングが必然的に遅くなり、本論文で提案した分析手法の利点はなくなる。多くの教育機関で、学生向けアンケートの乱立や200を超えるような設問を有するアンケートの実施等が見受けられ、それらのアンケート結果の分析による学生像の把握が行

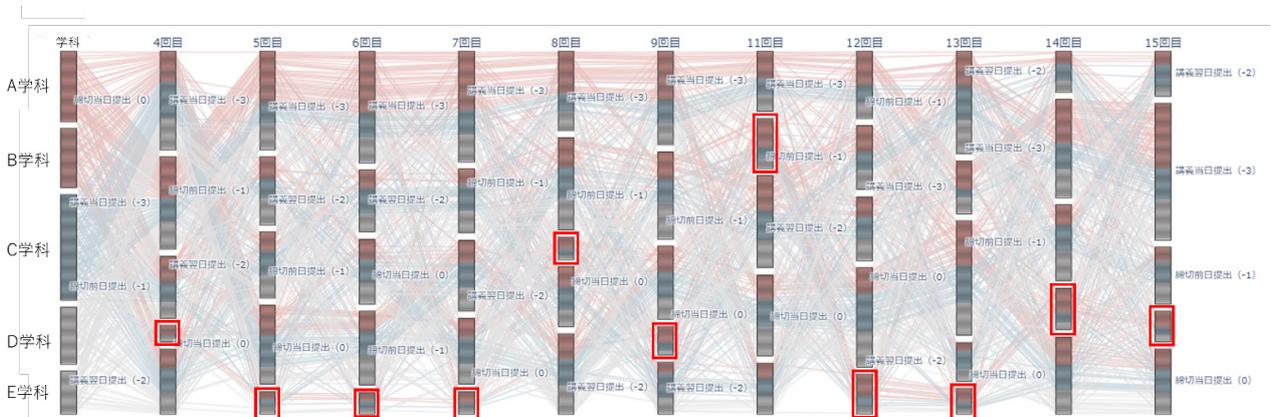


図2 全学科の各講義回の課題提出状況の推移

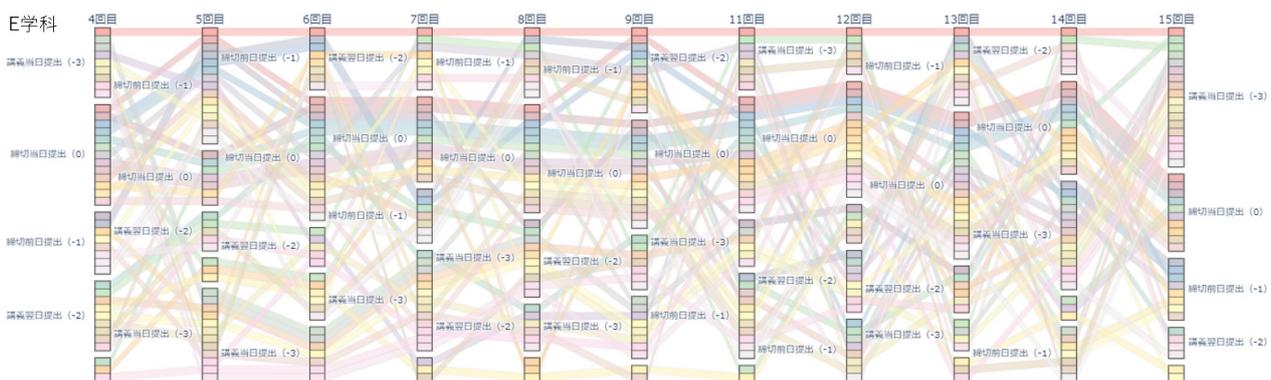


図3 E学科の各講義回の課題提出状況の推移

われているが、用いなければならない分析手法も複雑化するため、本来の目的である個々の学生に対する提言にまで落とし込まれていないのが現状である。

そのような状況が生じている根本的な原因は、最終的な分析を理工系職員に委ねるといった、教育機関組織内に未だに存在する、文系理系の垣根の存在と考えられる。海外、特に欧州においては、教員が学長になる際は、自身の専門分野とは全く異なる分野の博士号を新たに取るのが必須になっていることが多く、そのためにリベラルアーツが重要視され、どのような分野にも関わるものとして、DS教育がリベラルアーツに含まれている。筆者が教育アドバイザーとして関わった、北海道札幌市の市立札幌旭丘高校においては、2022年から数理DS学科が新たに設置されたが、その設置の際には数学や情報の教員以外に、国語や英語の教員も主要メンバーとして参画しており、2022年の数理DS学科1年生の副担任は英語教員が担っている¹⁰。DSに対する得手不得手の分岐タイミングが中学生や高校生のタイミングであることが多く¹¹、初等中等教育段階での今後のDS教育の重要性は増していくと考えられる。

その他、地方公立大学においては、入学試験科目に数学が含まれていない所も多く、入学志望の段階でDS教育への積極性に関する選択バイアスが存在する。この問題に対しては、中村（2020）¹²が述べるようなEdTechの活用のみでは解決しないことが多分にある。また、葛西ら（2022）¹³が述べるように、プログラミング等の技術的なスキルの部分に不得手を感じている学生は多く存在するが、教員側にも未だに多数存在している。そのため、DS教育担当教員による講義のデジタル化が重要ではなく、むしろ講義後サポートといった、ソフト面での重要性が今後増すと考えられ、そのようなサポート要員に対する応募者数を地方においても一定数確保するためには、サポート要員に対して十分な雇用待遇を維持するための方策検討もDS教育を進める上で、重要になると考えられる。

参考文献

1. Peter Naur; Concise Survey of Computer Methods, *Studentlitteratur*, Lund, Sweden, 1974
2. 文部科学省：https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00001.htm

- (2022年12月9日情報取得)
3. 数理・データサイエンス・AI教育強化拠点コンソーシアム、<http://www.mi.u-tokyo.ac.jp/consortium/overview.html> (2022年12月9日情報取得)
4. 内閣府；AI戦略2022、<https://www8.cao.go.jp/cstp/ai/index.html> (2022年12月9日情報取得)
5. 旺文社 教育情報センター；2023年新設予定の学部・学科情報、https://eic.obunsha.co.jp/pdf/educational_info/2022/0502_1.pdf (2022年12月9日情報取得)
6. (株)矢野経済研究所；2022 リカレント教育ビジネスレポート、2022
7. デジタル・ガバメント閣僚会議：デジタル社会の実現に向けた改革の基本方針、<https://www.city.shijonawate.lg.jp/uploaded/attachment/15914.pdf> (2022年12月9日情報取得)
8. 山口県：やまぐちデジタル改革基本方針、<https://www.pref.yamaguchi.lg.jp/uploaded/attachment/108362.pdf> (2022年12月9日情報取得)
9. 山口県立大学：<https://www.yamaguchi-pu.ac.jp/au/ap/sparc/> (2023年1月6日情報取得)
10. 市立札幌旭丘高校：<https://www23.sapporo-c.ed.jp/asahigaoka/index.cfm/1,0,41,html> (2023年1月8日情報取得)
11. 野津田雄太ら：大学生アンケートからの文系理系学生の特徴に関する分析、情報処理学会論文誌教育とコンピュータ、Vol.1、No.4、pp.83-92、2015
12. 中村眞一：Society5.0, IoT, AI, ICT そして「未来の教室」神奈川大学心理・教育研究論集、47号、pp. 169 - 196、2020
13. 葛西正裕ら：文系AI人材教育に対する調査研究、経済研究所所報、第2号、pp.50-78、2022