

プログラミング初学者に向けた数理データサイエンス AI 教育

プログラム選択学修項目の実践

二瓶 裕之^{*1}, 西牧 可織^{*1}

^{*1} 北海道医療大学

Practice of option skill sets in MDASH for programming beginners

Hiroyuki Nihei^{*1}, Kaori Nishimaki^{*1}

^{*1} Health Sciences University of Hokkaido

数理データサイエンス AI 教育プログラムリテラシーレベルにおいて選択の学修項目となった「プログラミング基礎、教師あり学習、自然言語処理」などを、プログラミング初学者に対して実施するための工夫について、本学の授業デザインや実践結果の観点から明らかにする。本学は医療系大学であり、初めてプログラミングを学ぶ学生が多いことから、手を動かしながらデータ分析などを体験できるような授業を設計したり、プログラミングを学ぶことの目的や意義をどのように学生へ伝えるのかなども検討を重ねたりしており、本実践は将来的にプログラミングなどの補正教育にもつながると考える。

キーワード: 数理データサイエンス AI 教育プログラム, プログラミング基礎, 医療系大学

1. はじめに

数理データサイエンス AI 教育プログラム（以下、MDASH）リテラシーレベルには、コア学修項目である「導入、基礎、心得」に加えて、「選択」の学修項目が組み込まれている。「選択」については、学生の学修歴や習熟度合いなどに応じて設定するものとされており、「プログラミング基礎、教師あり学習、自然言語処理」などのスキルセットが含まれる。

「選択」のスキルセットで基盤となるプログラミングは、令和4年度新高等学校学習指導要領における「情報Ⅰ・Ⅱ」にも組み込まれたが、「情報Ⅰ・Ⅱ」を履修した学生が大学等へ入学するのは令和7年以降である。現状では、大学入学時点でプログラミングの経験を有する学生は1割から2割程度とされている（1, 2, 3）。そのため、MDASHにおいて「選択」の学修項目を実施するためには、プログラミング初学者に向けた授業設計に加えて、プログラミングなどを導入する目的や意義も学生へ伝える必要があると考える。

このような背景の中、大学におけるプログラミング

教育の実践例は、MDASHにおける実施も含めて、かねてより、理工学系や人文社会学系の学部を中心に報告されている。例えば、プログラミングの初学者向けとしたオンライン授業(1)や授業支援環境の構築(4)、コード補完機能などの効果的な学修方法(5, 6)などが検討されている。また、MDASHの学修項目としての実践も広がり始めており、機械学習・深層学習の修得や問題解決能力の育成を目指したプログラミング教育なども報告されている(7, 8)。

一方、医療系の多くの学部においては、MDASHの実施はまだ過渡期にあるが、北海道医療大学（以下、本学）では令和2年度よりMDASHリテラシーレベルにあたる教育プログラムの実践を始めた。本学が有する学部はすべて医療系であるが、リテラシーレベルの「導入、基礎、心得」については、全学部の必須科目となっている「情報処理演習」等において実施している。さらに、一部の学部で開講している「情報科学」では、「選択」の学修項目も実施し始めた。

本論文においては、MDASHリテラシーレベルの「選

扱」の学修項目となった「プログラミング基礎、教師あり学習、自然言語処理」などのスキルセットを実施するための工夫について、「情報科学」の授業デザインや実践結果の観点から明らかにする。本学は、医療系大学であり、「情報科学」において初めてプログラミングを学ぶ学生も多いことから、手を動かしながらデータ分析などを体験できるような授業を設計したり、プログラミングを学ぶことの目的や意義をどのように学生へ伝えるのかなども含めて検討を重ねたりしており、本実践は将来的なプログラミングの補正教育にもつながると考える。

2. 授業でサインの概要

MDASH の「選択」の学修項目を実施している「情報科学」を開講しているのは看護福祉学部と医療技術学部であり、前者では1年後期の選択科目、後者では1年後期の必須科目となっている。授業はすべて対面形式で実施し、履修者数は、ともに70名程度である。また、「情報科学」の履修条件として、「情報処理演習」等において MDASH リテラシーレベルのコア学修項目である「導入、基礎、心得」を修得していることを定め、スプレッドシートアプリケーションの分析ツール（フィルタやピボットテーブル等）を使うスキルも修得していることを前提としている。

「情報科学」の一般目標は、将来従事する医療職の現場において数理データサイエンス AI を使いこなせる知識やスキルを学ぶこととしている。行動目標としては、医療の現場から取得するデータを活用するうえでの個人情報の扱いができること、分析ツールを使いながら時系列分析や回帰分析ができることを挙げている。また、分析ツールをブラックボックスとせずに医療の現場で目的に沿って正しく使うことを目的として、アルゴリズムやプログラミングの基礎や考え方を身に着けることで、データ分析の計算手順や仕組みを理解することを挙げている。そのうえで、画像診断などでも活用されている機械学習や、診療記録の音声入力などでも活用されている自然言語処理などの数理データサイエンス AI について説明できることとしている。

「情報科学」で扱う授業内容のテーマは表1にまとめた。5つのテーマを、各々、3回の授業で実施して

いる。序盤の2つのテーマは、MDASH リテラシーレベルの「基礎、心得」に該当するテーマであるが、「情報処理演習」で学んだ内容をさらに深めた内容としている。中盤以降の3つのテーマが「選択」に組み込まれているスキルセットに該当するテーマである。

また、本学では、学生全員がパソコン必携となっており、事前学修についても、テキストや資料を事前に読んでおくことをシラバスに記載している。

表 1 「情報科学」で扱う授業内容のテーマ

授業回	テーマ
1～3	個人情報、匿名化処理、情報セキュリティ
4～6	データ処理、データ集計、データ分析
7～9	プログラミング基礎
10～12	教師あり学習（機械学習）
13～15	自然言語処理

3. 授業デザインの工夫

授業デザインの工夫として、まず、初回授業のガイダンスにおいて、「情報科学」の学修内容の一部（時系列分析や回帰分析などのデータ分析、プログラミング基礎など）が、令和4年度新高等学校学習指導要領における「情報I」の学修内容となっていることを伝えている。自身が医療人になった数年後には、新しく現場に加わる医療人の全てが、これらの学修内容を学んでくると実感できるようにしている。

また、各テーマ実施回の冒頭では、扱うテーマが、DX化が進む医療分野において必須の知識や技能となるなど、自身が目指す医療人の観点から各テーマを学ぶ目的や意義を伝えるようにしている。さらに、毎時間、自身の必携パソコンにより演習課題を解かせるなど、手を動かしながらデータ分析やプログラミングなどを体験できるような工夫をしている。

例えば、「個人情報、匿名化処理、情報セキュリティ」を学ぶ意義としては、「医療の分野でも広く用いられている調査・研究において、個人情報を扱ったり匿名化処理を施したりするスキルが欠かせないこと」を伝えている。演習課題では、アンケート結果に対して k-匿名化を施すことで、個人情報の具体的な取扱いを修得

できるようにしている。また、スプレッドシートのフィルターやピボットテーブルを使いながら解くように演習課題を設定することで、「情報処理演習」等で学んだ分析ツールの操作を復習できるように工夫している。

「データ処理、データ集計、データ分析」を学ぶ意義としては、「医療に関わるデータの分析において、時間的な変化に対する時系列分析や目的変数と独立変数に対する回帰分析の手法が欠かせないこと」を伝えている。演習課題では、新型コロナウイルス感染者数に対する時系列分析や人口と医療施設数に対する回帰分析をするなど、医療の観点からのデータ分析のスキルを修得できるようにしている。ここで、時系列分析や回帰分析は「情報科学」で初めて扱う学修内容であることから、まずは、スプレッドシートの分析ツールを使いながら、時系列分析や回帰分析の数学的な概念を視覚的に学べるような工夫をしている。

「プログラミング基礎」を学ぶ意義としては、「データ分析ツールを目的に沿って正しく使えるようになるために、アルゴリズムやプログラミングの基礎や考え方を学ぶことで、データ分析における計算手順や仕組みを理解する」と伝えている。演習課題では、まず、スクラッチ(9)を使って繰り返しや条件などのフローチャートの概念を学び、Edublocks(10)を使ってPythonのライブラリやImportの概念を学んだうえで、Colabatory(11)を使ってPythonによるプログラミングの基礎を学べるようにしている。しかし、演習課題では、自身でアルゴリズムを作ってプログラミングする観点では、グラフを描くなどの簡単な課題にとどめ、むしろ、「データ処理、データ集計、データ分析」で扱ったのと同じ時系列分析と回帰分析をするPythonのソースコードを学生へ提示して、それを読み解くことを課題とするなど、データ分析の計算手順や仕組みを理解することを重視する工夫をしている。

「教師あり学習（機械学習）」を学ぶ意義としては、「画像診断や病気の早期発見など医療の分野においても機械学習やAIによる予測・判断の技術は欠かせない」と伝えている。演習課題では、画像認識モデルを作りながら機械学習の仕組みを学んだり、Scikit-Learnの回帰モデル（LinearRegression）による予測技術を学べるようにしている。ここでも、Pythonのソースコードはすべて学生へ提示し、学生は、ソースコ

ードに従って学習用データを適切に用意したり、適切なパラメータ値を設定したりする演習を行うようにしている。また、予測技術で扱う回帰分析については、「データ処理、データ集計、データ分析」で視覚的な概念を学び、それに続く「プログラミング基礎」で計算手順や仕組みを学んでおり、「情報科学」の一連のテーマの中で段階的に学びを深められる工夫をしている。

最後に、「自然言語処理」を学ぶ意義としては、「医療分野においても、診療記録の音声入力や電子カルテの解析など自然言語処理技術には様々な応用例がある」と伝えている。演習課題では、文章を単語に分ち書きして解析する形態素解析や単語の関係性を解析するn-gramなどのテキストマイニング技術を学んだうえで、マルコフ連鎖などの文書生成技術により、どのように文章が作られていくのかを体感できるようにしている。また、ニューラルネットワークモデルであるDoc2Vecを使って文書類似度なども推論する。テキストマイニングや文書類似度の算出は、他の授業科目においてレポートを学生にフィードバックする際にも利用しており(12)、普段の大学生活で身近に触れている自然言語技術の仕組みを学べるという観点からも興味を持てるような工夫をしている。

なお、本学DX推進計画、ならびに、MDASH推進の観点から、演習課題の一部については本学DX推進計画サイト(13)に公開して、広く普及を図っている。

4. 効果の検証

MDASHの学修項目である「プログラミング基礎、教師あり学習、自然言語処理」などのスキルセットを、プログラミング初学者に対して実施するために行った「情報科学」の工夫について、実践結果の観点から検証する。

検証にあたって、対象とするのはX学部における「情報科学」の履修者70名程度とするが、まず、履修者像を明らかにするために、大学入学前におけるプログラミング学習の経験についてアンケートを実施した(n=68)。その結果、大学入学前に高等学校においてプログラミング学習をした経験のある学生は13名であり、従前の調査結果(1, 2, 3)と同様に2割程度であった。一方、独学でプログラミングを経験していた

学生は1名のみであり、高等学校での学修を除くと大学入学前にプログラミングを経験したことのある学生が非常に少なかったが、これは、本学が医療系大学であることが要因の1つと考えられる。

次に、実践結果の検証として、図1に「情報科学」定期試験の結果を示した。定期試験では、表1のテーマごとに大問題として問1から問5を設定した。

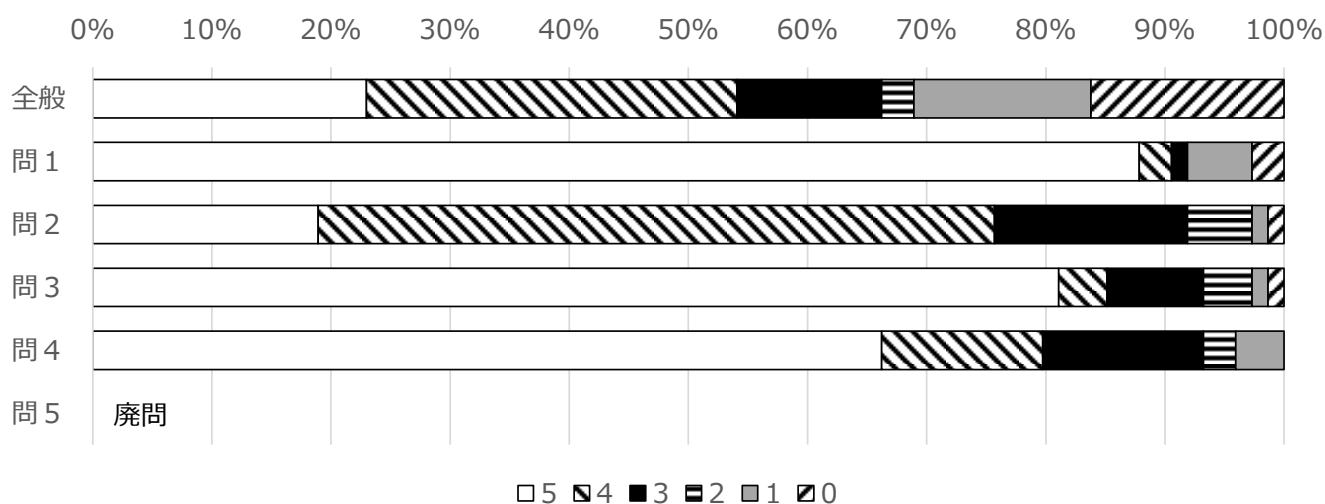


図1 「情報科学」の定期試験の結果

問1「個人情報、匿名化処理、情報セキュリティ」では、分析ツールを使いながら、同一属性のデータがk件以上となるように匿名化ができるかを問うた。

問2「データ処理、データ集計、データ分析」では、統計ダッシュボード(14)から2変数のデータをダウンロードしたうえで、分析ツールを使いながら、2変数間の相関係数と回帰直線を求める回帰分析ができるか、また、回帰分析の結果を正しく読み取れるのかを問うた。

問3「プログラミング基礎」では、プログラミング言語(Python)の書式に従って関数を適切に記述してグラフを描画できるか、適切なデータと時間間隔を指定して時系列データをトレンドとノイズに分離する時系列分析ができるかを問うた。

問4「教師あり学習(機械学習)」では、Scikit-Learnの回帰モデル(LinearRegression)を使った重回帰分析により算出された予測値を正しく読み取れるのかを問うた。

問5「自然言語処理」では、指定した小説のテキストデータを青空文庫(15)からダウンロードしたうえで、文章を単語に分ち書きして解析する形態素解析ができるかを問うた。

試験を実施するにあたっては、学生へ自身の必携パソコンを持ち込むように伝えて、各問の回答をまとめ

たドキュメントファイルを解答用紙として Google フォームから送信するように指示した。また、問1と問2についてはスプレッドシートの分析ツールを、問3から問5については Google Colaboratory を使って解くことを伝えた。なお、Google Colaboratory がオンラインアプリケーションであったり、統計ダッシュボードや青空文庫などの外部サイトも利用したりすることから、インターネットへの接続にも制限を設けないことを伝えた。

そのため、問題全般について、知識を問う問題ではなく、適切な外部サイトから必要なデータを取得するなどしながら、データを正しく分析したり、データの分析結果を正しく読み取ったりするスキルの習熟度を測ることを意図して問題を設定した。

さらに、問3から問5については、問題文中に Python のソースコードを提示しており、プログラミング技術よりも、むしろ、プログラムの計算手順に沿ってデータを適切に用意したり、プログラムの計算結果を正しく読み取ったりできるのかを重視して評価をした。

図1には、さらに、「全般」とした項目もあるが、ここには、解答用紙を作成するにあたって問われる統合型オフィスツール(主に、文書作成)の利用スキルについて評価した結果を示した。

問5の結果には廃問と記載しているが、これは、試験時間の途中より、本学から青空文庫へのアクセスができなくなってしまったためである。問5で外部サイトを利用した意図は、外部サイトから学生自身がデータを取得したうえで、さらに、そのデータを分析させることで、実践的なスキルを測ることができるとの考えであったが、一方で、今回のような事象も発生する可能性もある。

また、図1には、全般、および、問1から問4について、各々、5点満点と換算したときの採点結果の分布を示した。採点の方法は、例えば、問2では、指定したデータが正しく使われているのか、回帰直線や相関係数が正しく求められているのかといった客観的な採点に加えて、グラフの読み取りが正しいのか、事実に基づいた考察ができているのかなどを評価した。

まず、オフィスツールの利用スキルを図った「全般」の項目については、問1から問4と比較すると、各点数に満遍なく学生が分布していることがわかる。

問1については、9割近くの学生が5点となっており、分析ツールを正しく使いながらk匿名化ができていた。問3についても、8割以上の学生が、関数を正しく指定してグラフを描いたり、指定されたデータと時間間隔で時系列分析ができていたりした。

一方、問2と問4については、5点の比率が下がっている。これは、両問については、データを分析したり、分析の結果を読み取ったりすることに加えて、分析の結果についての考察も求めていることに依る。

問4では、適切なデータを使って重回帰分析ができた学生（3点に相当）は9割、結果のグラフを正しく読み取れた学生（4点に相当）は8割近くまでいたが、重回帰分析の予測結果が正しい、もしくは、正しくないと判断される要因を考察できた学生（5点に相当）は6割5分程度であった。

最も得点率が低かった問5については、問題を詳細に分析する。問4では、統計ダッシュボードから1975年と2012年の両年について、都道府県別の「15～64歳の人口」と「病院数」をダウンロードして（ここまでで、1点相当）、各々、散布図を描かせた（2点相当）。次に、相関係数と回帰直線を求めさせて（3点相当）、1975年と2012年度のばらつきの具合、つまり、2012年のばらつきが大きいことを読みとらせた

（4点相当）。最後に、2012年度、つまり、以前と比較して近年では、「15～64歳の人口」と「病院数」のばらつきが大きくなってきていることの社会的な要因を自分なりに考察することを求めた（5点相当）。図1の結果から、相関係数と回帰直線を求めたり、その結果を読み取ったりするところ（4点に相当）までは、7割5分程度の学生ができていた。

これからの結果から、MDASHの学修項目である「プログラミング基礎、教師あり学習、自然言語処理」などのスキルセットを「情報科学」の中に取り入れたことで、正しくデータ分析したり、分析結果を正しく読み取ったりするスキルの習熟度については、一定の効果が見られたものと考ええる。

プログラミング技術については、グラフを描画するなどのスキルを検証したのにとどまっているが、行動目標とした「分析ツールをブラックボックスとせずに医療の現場で目的に沿って正しく使うことを目的として、アルゴリズムやプログラミングの基礎や考え方を身に着けることで、データ分析の計算手順や仕組みを理解する」ことの観点は、達成されたと考える。

特に、回帰分析については、「データ処理、データ集計、データ分析」で視覚的な概念を学び、「プログラミング基礎」で計算手順や仕組みを学び、さらに、「教師あり学習（機械学習）」では、予測技術の1つとしても学ぶなど、「情報科学」の一連のテーマの中で手を動かしながら、段階的に学びを深められる工夫をした効果が表れていると考える。

最後に、「自然言語処理」については、問5「自然言語処理」が廃問となったことから、15回目授業の振り返りコメントを検証する。振り返りコメントは、授業の毎時間取得しており、授業で学んだことや感想を学生に自由に記載してもらっている。

15回目授業では、第3章の「授業デザインの工夫」で述べたように、ニューラルネットワークモデルであるDoc2Vecを構築し、文章類似度を推論した。具体的には、まず、250万文字からなるテキストデータをインターネットからダウンロードして、エポック数などを指定しながらDoc2Vecにより推論モデルを構築させた。次に、推論モデルを利用して、単語や文章をベクトル化させたり、ベクトル化した単語をヒートマップにより可視化させたりした。さらに、ベクトル

の類似性から類似語を抽出したり、対義文章(16)を生成したりする演習課題に取り組むようにさせた。

図2に、振り返りコメントの結果をまとめた(n=68)。振り返りコメントには、「わかった」、「すごい・楽しい」、「難しい」のいずれか、もしくは、複数があったことから、振り返りコメントの概要として、3つのコメントを記載した学生数(複数あり)を示した。「わかった」とした学生は66名であり、その中に「すごい・楽しい」と記載した学生が含まれる。残りの2名は「難しい」と記載した。

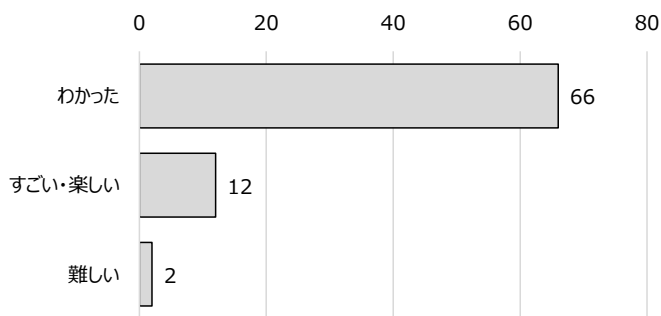


図2には、「わかった」とした学生が、わかったとした項目(複数あり)を記載した。各項目は、AI、プログラミング、自然言語処理などといった授業概要を示すキーワード以外に加えて、ベクトル化や類似語・対義語など授業で扱った詳細なキーワードとも一致した。

15回目授業は、「情報処理演習」から「情報科学」へと至るMDASHの総まとめとして、MDASHの応用基礎レベルにもつながるような高度な学修内容であったが、自由記載コメントの概要からも概ね好評であり、わかったとされた項目も適切であったと考える。

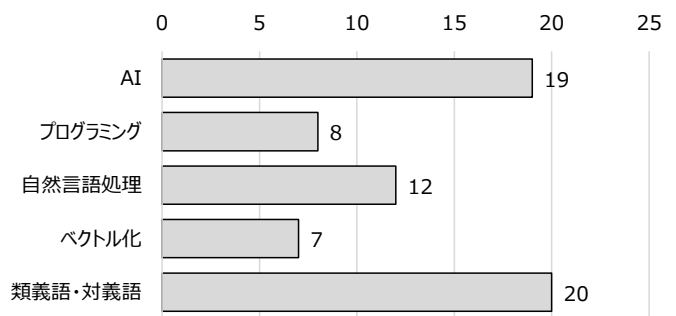


図2 振り返りコメントの結果

5. むすび

MDASHにおいて選択の学修項目となった「プログラミング基礎、教師あり学習、自然言語処理」を本学において実施するための工夫について、授業デザインや実践結果の観点から明らかにした。

本学は医療系大学であることから初めてプログラミングを学ぶ学生が多く、特に、独学でプログラミングを経験していた学生は非常に少なかった。多くの学生がプログラミング初学者である状況において、「プログラミング基礎、教師あり学習、自然言語処理」を実施する工夫として、まず、授業で扱うテーマが、DX化が進む医療分野において必須の知識や技能となるなど、自身が目指す医療人の観点から各テーマを学ぶ目的や意義を伝えた。そのうえで、毎回の授業では、自身の必携パソコンにより演習課題を解かせるなど、手を動かしながらデータ分析やプログラミングなどを体験できるようにした。

特に、プログラミングを学ぶ目的や意義としては、自身でアルゴリズムを作ってプログラミングをする技術よりも、プログラミングの基礎や考え方を身につけることで、データ分析の計算手順や仕組みを理解し、

医療の現場で目的に沿って正しいデータ分析ができることとした。

また、データ分析の手法である時系列分析や回帰分析などは、複数のテーマの間で段階的に学びを深められるようにした。例えば、回帰分析については、「データ処理、データ集計、データ分析」で視覚的な概念を学び、「プログラミング基礎」で計算手順や仕組みを学び、さらに、「教師あり学習(機械学習)」では、予測技術の1つとしても学べるようにした。

これらの工夫について、定期試験の結果から、その効果の検証をした。結果、データを正しく分析したり、分析の結果を正しく読み取ったりするスキルについては、一定の効果が見られるなど、全体的に概ね良好な結果であったと考える。一方で、社会的な要因も加味しながら、データ分析の結果について考察するなどのスキルについては、応用基礎レベル相当など、より高度な内容を扱う授業科目とも連携しながら、更なる実践が必要であると考えられる。

定期試験以外の観点として、学生の振り返りコメントについても、MDASHの総まとめとなった15回授業を対象に検証した。結果、授業内容が、MDASHの

応用基礎レベルにもつながるものとなったが、コメントの結果は概ね好評であったと考える。

「情報科学」で扱った「時系列分析、回帰分析、プログラミング基礎」は、令和4年度新高等学校学習指導要領における「情報Ⅰ」の学修内容となっている。令和7年度以降は、MDASHの実施にあたり、これらの学修内容を高等学校で履修したことを前提として授業を設計することになるが、それと同時に、補正教育の必要性が高まることも予想される。本実践で対象となった学生の多くはプログラミング初学者であり、医療系大学における本実践は将来的にプログラミングなどの補正教育にもつながると考える。

今後の検討課題として、まず、定期試験については、難易度設定の適切さについて検証をしたい。今回の結果では、一部の問題について高得点域に多くの学生が分布したが、これは、難易度の設定が低いと考えることもできる。特に、プログラミング技術を問う問題については、プログラミング言語の書式を問う簡単な内容としたが、他の実践例も参考にしながら、問題設定の適切さなども検討していきたい。

また、自身が目指す医療人の観点から各テーマを学ぶ目的や意義を伝えたことによる効果についても、今回は直接的な検証はしていないので、今後の検討課題としたい。

最後に、現在、MDASHの応用基礎レベル相当の授業科目として、「医療データサイエンス入門Ⅰ・Ⅱ」も開講しており、今後は、リテラシーレベルの次のステップとなる応用基礎レベルの実践についても検証を続けていきたい。

謝辞

本研究はJSPS科研費JP 22H01051の助成を受けたものである。

参 考 文 献

- (1) 布施泉: “Code Commons: Sunaba を用いた大学の一般情報教育としてのプログラミング教育”, 教育システム情報学会誌, Vol.35, No.2, pp.227-232 (2018)
- (2) 布施泉: “初学者を主対象とする大学の一般プログラミング教育のオンライン授業による実施”, 高等教育ジャーナル: 高等教育と生涯学習, No.28, pp.65-72 (2021)

- (3) 深井裕二: “プログラミング教育に関する入学生の学習経験およびこれからの大学教育”, 北海道科学大学研究紀要, No.50, pp.1-5 (2022)
- (4) 布施泉, 中原敬広, 岡部成玄: “プログラムの相互利用と相互評価が可能な初学者用プログラミング授業支援環境の構築”, 教育システム情報学会誌, Vol.35, No.2, pp.221-226 (2018)
- (5) 三浦元喜: “初学者向け Processing プログラミング環境におけるコード補完機能の導入と効果”, 教育システム情報学会誌, Vol.37, No.2, pp.167-172 (2020)
- (6) 時田真美乃, 不破泰: “初学者向けプログラミング基礎教育における可聴化を用いた多重ループの効果的な学習方法”, 教育システム情報学会誌, Vol.38, No.1, pp.49-54 (2021)
- (7) 巳波弘佳: “AI を使いこなす教育プログラムの取組み～AI 活用スキルを身に付けるには～”, 大学教育と情報, No.3, pp.7-12 (2020)
- (8) 渡邊 紀文: “問題解決を重視したプログラミング教育とオンラインでの実践”, Musashino University Smart Intelligence Center 紀要, Vol.2, pp.60-67 (2021)
- (9) SCHRATCH, <https://scratch.mit.edu/> (2023年2月5日確認)
- (10) Edublocks, <https://edublocks.org/> (2023年2月5日確認)
- (11) Colaboratory, <https://colab.research.google.com/> (2023年2月5日確認)
- (12) 二瓶裕之, 浜上尚也, 木村治, 小田雅子: “面接受講と遠隔受講を組み合わせた早期体験学習ワークショップの実施と検証”, 薬学教育, Vol.5, pp.1-9 (2021)
- (13) 北海道医療大学 DX 推進計画サイト, <https://dx.hoku-iryo-u.ac.jp/> (2023年2月5日確認)
- (14) 統計ダッシュボード, <https://dashboard.e-stat.go.jp/> (2023年2月5日確認)
- (15) 青空文庫, <https://www.aozora.gr.jp/> (2023年2月5日確認)
- (16) youwh!: “キテレツおもしろ自然言語処理”, 翔泳社 (2021)