

Shiga University's endeavor to promote human resources development for data science in Japan

Takuma Tanaka^{1,2,3*}, Tetsuto Himeno^{1,2,3} and Kaoru Fueda^{1,2,3}

^{1*}Faculty of Data Science, Shiga University, 1-1-1 Banba, Hikone, Shiga, 522-8522, Japan.

^{2*}Graduate School of Data Science, Shiga University, 1-1-1 Banba, Hikone, Shiga, 522-8522, Japan.

^{3*}The Center for Data Science Education and Research, Shiga University, 1-1-1 Banba, Hikone, Shiga, 522-8522, Japan.

*Corresponding author(s). E-mail(s):
takuma-tanaka@biwako.shiga-u.ac.jp;

Abstract

In 2017, Shiga University established the Faculty of Data Science, which was the first faculty in Japan specializing in data science and statistics. This paper reports the Faculty's historical context, curricula, and collaboration with industry and other universities. The career paths of the graduates and the massive open online courses and textbooks provided by the Faculty of Data Science are also summarized.

Keywords: Collaboration, Data science, Data scientist, Project-based learning, Statistical science, Education

1 Introduction

The coronavirus disease 2019 (COVID-19) pandemic has been a stark reminder of the importance of data acquisition, data accumulation, data analysis, and data visualization, i.e., the importance of data science ([Gardner, Ratcliff,](#)

Dong, & Katz, 2021). Although the term “data science” has been used since the 1960s (Irizarry, 2020; Naur, 1966), the presence of data and data science in society has seen explosive growth over the past decade, starting with an article with an eye-catching title, “Data Scientist: The Sexiest Job of the 21st Century” (Davenport & Patil, 2012). Data science has become an indispensable tool in various areas in society including information technology, business, public policy, medicine, and basic science.

Combined with data, artificial intelligence (AI) is projected to deliver 1.2 percent additional GDP growth per year until 2030 (Bughin, Seong, Manyika, Chui, & Joshi, 2018). However, to maximize the value gained from big data, not only technology investment, but also newly trained experts and the adequate policies and management are needed (Byers, 2014). Therefore, governments and political organizations around the world, including the European Union and United Nations, are rushing to fund human development and higher education in data science and AI (European Commission, 2014; UN Global Pulse, 2021). For instance, in 2020, the UK government awarded £24 million of funding to 18 universities that deliver AI and data science conversion courses (Department for Digital, Culture, Media & Sport, Department for Education, Department for Business and Energy & Industrial Strategy, & Office for Artificial Intelligence, 2020)

On the other hand, a severe shortage of data scientists and AI engineers has been reported in Japan compared with other developed countries. A report on big data by McKinsey Global Institute (Manyika et al., 2011) showed that the number of people in Japan with deep analytical training decreased from 2004 to 2008 to about 3,400 in total. Mizuho Research and Technologies (2019) predicted that AI engineers would reach a shortage of 145,000. The Japan DataScientist Society (2019) reported that less than 30% of companies that have at least one data scientist.

The effort to train data scientists in Japan started with the establishment of the Japan DataScientist Society in 2013, and since that time, various policies for the development of data scientists have been implemented. The Ministry of Education, Culture, Sports, Science and Technology (MEXT) sponsored the Data Scientist Training Network in July 2013, the aims of which were the sustainable development and effective utilization of data analytics experts in Japan (Maruyama, Kamiya, Higuchi, & Takemura, 2015). In 2016, the Japan Inter-university Consortium for Mathematics and Data Science Education was established, with the purpose of creating and disseminating a standard curriculum and teaching materials, such as data sets, that could serve as models for schools across the entire country. In 2019, AI Strategy 2019, which was issued by a government council (Integrated Innovation Strategy Promotion Council, 2019), proposed education reform for all high school, college, and university students. With these actions, data science education has made progress in Japan.

One of the most important moments in the effort to raise the standards of data science education in Japan was the foundation of the Faculty of Data Science of Shiga University in 2017. This was the first faculty in Japan to specialize in statistics and data science. The history and context of statistics education before the establishment of the Faculty of Data Science of Shiga University is described in [Takemura \(2018\)](#). As a sequel to [Takemura \(2018\)](#), this paper describes the government's policy, the curricula of the Faculty of Data Science, and its collaboration with industry and other universities. In this paper, we detail how the Faculty of Data Science has established university-level data science education and collaboration with higher education and industry in Japan.

This paper is organized as follows. Section 2 gives a detailed account of the government's policy with an emphasis on higher education. Section 3 describes the curricula of the Faculty of Data Science and its effort to disseminate data science education to Shiga University as a whole. Section 4 is divided into two subsections: the first summarizes the collaboration between the Faculty of Data Science and other universities and high schools, and the second reports its collaboration with industry. The employment status of the graduates is also reported. Section 5 concludes.

2 Policy of human resource development for data science by the Japanese government

The establishment of the Inter-university Consortium for Mathematics and Data Science Education in 2016, the purpose of which is described above, brought about a growing tendency for redesigning data science education in Japan. The standard curriculum created by the curriculum committee of the consortium was based on "The Data Science for Undergraduates: Opportunities and Options" ([National Academies of Sciences, Engineering, and Medicine, 2018](#)).

In 2019, the Japanese government announced a policy called AI Strategy 2019 to strengthen the development of human resources for mathematics, data science, and AI ([Integrated Innovation Strategy Promotion Council, 2019](#)). One of the objectives of the policy was to improve statistics and data science education. Before then, inferential statistics had been taught in the high school curriculum. However, this was not compulsory until 2021, so only about 10% of high school students studied inferential statistics ([Ministry of Education, Culture, Sports, Science and Technology, 2019](#)). MEXT is preparing a new curriculum starting in 2022 to encourage more high school students to study statistics. Moreover, AI Strategy 2019 proposed that all college and university students should acquire data science and AI literacy. To implement this proposal, the Inter-university Consortium for Mathematics and Data Science Education published the skill sets, learning goals, and model curriculum of data science. To promote the model curriculum, the Japanese government enacted a new certification criterion for colleges and universities in which concordance

with the model curriculum was assessed. Moreover, in 2020, MEXT proposed a data scientist training project for master's students for the realization of a Super Smart Society and Doctoral program for Data Related Innovation Expert (D-DRIVE) to enable doctoral students and postdoctoral researchers to improve their ability fully (<https://ddrive.jp/>).

However, the implementation of these projects requires a larger number of experts for teaching data science and statistics. To train these experts, the Institute of Statistical Mathematics was chosen as the core institution of the Statistical Expert Resource Development Consortium to be developed in 2021, and Shiga University served as its satellite center for western Japan. The purpose of this consortium is to provide statistics education to researchers who specialize in various applied fields and to foster statistics experts. It is expected that such researchers retrained in statistics will be able to supervise students and foster data scientists at universities in the consortium.

Therefore, a wide range of policies for the development of human resources in data science is currently being implemented in Japan. However, greater efforts in data science education are needed to develop data science education at a higher level.

3 Data science curriculum in Shiga University

One of the antecedents of Shiga University is Hikone Commercial College, which was established 100 years ago in 1922. The mission of Hikone Commercial College was to train businesspersons in the era of industrialization. In line with this mission, to train businesspersons in the era of big data, Shiga University launched the Center for Data Science Education and Research in 2016 and established the Faculty of Data Science in 2017. The student quota and number of faculty members were 100 and 17, respectively. Graduate courses were established in 2019 (master's course) and 2020 (doctor's course), the quotas for which were 20 and 3, respectively. To date, more than half of the graduate students are employees of companies and public agencies. This fact shows that Shiga University successfully fills the demand for recurrent education. In 2021, following the graduation of the inaugural class of the undergraduate and graduate courses, the university increased the quota of students for the master's course in the Graduate School of Data Science from 20 to 40; the quota for the doctor's course is planned to be increased. These measures are intended to fulfill the social demand for experts in data science. Since the establishment of the Faculty of Data Science, the total number of Faculty and Center members has increased to 44. In response to the rapid progress of AI and its social implementation, researchers in AI and social research were newly hired in the Faculty and Center.

In the same year, curriculum reform of the undergraduate course was carried out. The objectives of the curriculum reform were twofold. The first objective was to encourage the students to obtain more practical skills. The skill sets needed in the graduation project were more rigorously defined and

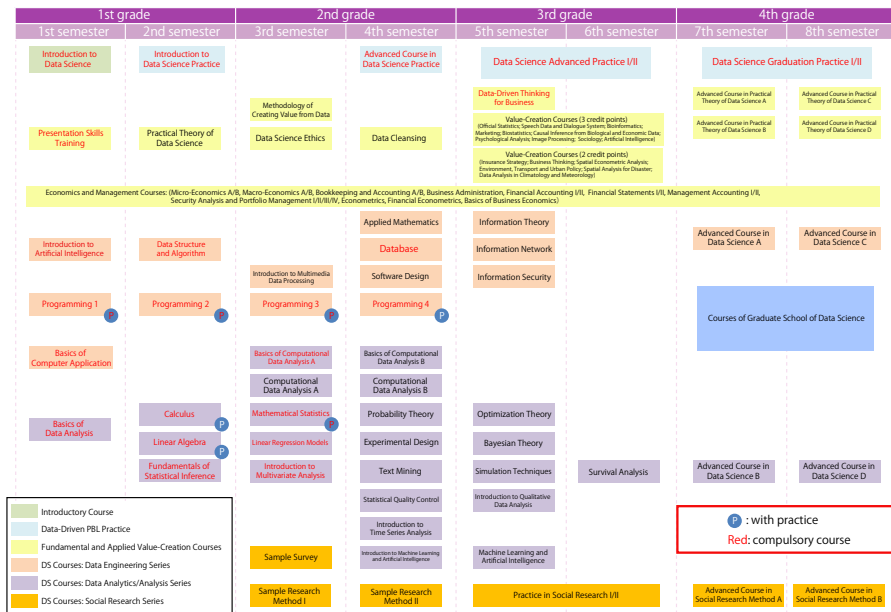


Fig. 1 Curriculum map of the Faculty of Data Science of Shiga University (2021–).

designed to be acquired by students. A course on data cleansing was newly established, and the programming courses were rescheduled to start earlier. The second objective was to deepen the knowledge and skills of the students in the emerging field of AI. To this end, Introduction to Multimedia Data Processing, Speech Data and Dialogue Systems, and Image Processing were newly established.

The reformed curriculum map of the Faculty of Data Science is shown in Fig. 1. Courses are classified into six groups: an introductory course, data engineering courses, data analytics courses, social research courses, fundamental and applied value-creation courses, and data-driven project-based learning (PBL) practices. Data engineering courses include courses on programming and computer science. Data analytics courses include courses in mathematics, statistics, and machine learning. Social research courses include courses on sample surveys and research. Value-creation courses include ethics, data cleansing, and various applied fields of data science. Data-driven PBL practices, of which we give a detailed account in the following, contain practices using data from various fields. Most courses in the first, second, and third semesters and all PBL practices are compulsory.

In the following, we give a detailed account of the curriculum reform using examples from the PBL and data engineering courses. Since its establishment in 2017, the Faculty of Data Science of Shiga University has put PBL front and center. PBL practices are compulsory in the second, fourth, fifth, sixth,

seventh, and eighth semesters. The PBL practice in the second semester (Introduction to Data Science Practice) focuses on the basics of data analysis based on the Problem, Plan, Data, Analysis, and Conclusion (PPDAC) cycle (Wild & Pfannkuch, 1999) and data analysis skills using mainly Excel. The PBL practice in the fourth semester (Advanced Course in Data Science Practice) is aimed at acquiring advanced data analysis skills using Python and R. These two PBL practices lay the foundation of the PBL practices in the fifth through eighth semester, in which the students accomplish their graduation project.

Introduction to Data Science Practice in the second semester includes lectures on the PPDAC cycle and group exercises. After the curriculum reform, we defined the content of each class, such as how to calculate representative values and use logistic regression. Lecture on these topics are given in the first half of the class and practiced in the second. In the second half of the class, the students form groups and tackle problems they have formulated.

Before the curriculum reform, the students practiced either Python or R in Advanced Course in Data Science Practice. After the reform, they pursue the skills of both. Additionally, after the reform, greater emphasis is placed on data preprocessing, especially in regard to missing data.

Data engineering courses are intended to encourage students to acquire practical data processing skills. Although lectures and practices on programming started in the second semester before the curriculum reform, they start in the first semester after the reform. The earlier starting time aims to enable students to acquire essential programming skills before learning PBL practices.

There are four sets of programming lectures and practices in our curriculum. Programming 1 in the first semester is an introduction to Python. Programming 2 in the second semester is an intermediate Python course. The main theme of Programming 2 is learning the standard data-science libraries including NumPy, Pandas, Matplotlib, and scikit-learn. Although Programming 2 was initially designed as an algorithm-oriented programming course, after the curriculum reform, it focused on statistics and AI. Programming 3 in the third semester is an introduction to Java. This is also a course for GUI programming. Programming 4 in the fourth semester is a course on JavaScript, HTML, web programming, web scraping, and 3D visualization. Initially, this course was designed to contain numerical analysis, but this was later omitted to emphasize data acquisition and visualization.

In addition to training specialists, the Faculty of Data Science helps the other faculties of Shiga University integrate data science education into their curricula. There are two faculties at Shiga University other than the Faculty of Data Science: the Faculty of Education and the Faculty of Economy. The Faculty of Education is planning to train teachers capable of promoting evidence-based education. The Faculty of Economy is reforming its curriculum to train businesspersons who can practice data-driven decision-making. The Faculty of Data Science provides these faculties with introductory courses on data science, one of which is planned to be compulsory for all students. Massive open online courses (MOOCs), which are described in the next section, are used in the courses for these faculties.

4 Collaboration with universities and industry

We collaborate with other universities to promote and disseminate data science education in all universities in Japan and with industry to provide data-driven consultancy and train employees to be data scientists. In this section, we review our various efforts, i.e., collaboration with other universities, the MOOCs and textbooks by the Faculty of Data Science, collaboration with high schools, the employment status of graduates, collaboration with industry, and the consortia established to promote collaborations.

4.1 Inter-university collaboration for data science education

In August 2020, the Liaison Committee for Data Science Departments of Universities in Japan (<https://lcdsj.jp/>) was formed with the aim of contributing to the development of data science education and research in Japan through cooperation and collaboration among specialized data science faculties, departments, graduate schools, and majors. At the time of its inception, six universities and departments were involved: Faculty of Data Science of Shiga University, Department of Statistical Science of the Graduate University for Advanced Studies, School of Information and Data Science of Nagasaki University, School of Social Information Sciences of University of Hyogo, Hitotsubashi University, and Rissho University, the latter two of which were planning to open data science faculties in the future. In November 2020, Gunma University, which was planning to establish a data science faculty, also joined. The Liaison Committee promotes education and research in the field of data science at universities and graduate schools through ongoing discussions on education, research, and management, while promoting direct dialog between industry, government, and academia, and issuing joint statement on data science education.

The Faculty of Data Science of Shiga University is developing a textbook series in cooperation with the Inter-university Consortium for Mathematics and Data Science Education and another by itself. The former includes multimedia data analytics and data science mathematics, and the latter includes social research and data visualization. The most introductory textbook of the latter series, “Introduction to Data Science,” has been used in almost 30 Japanese universities and become a new classic in this field. The Korean translation was published in 2020, and a version customized for Nagasaki University is in preparation. This wide acceptance indicates that our textbooks meet the demands of universities in Japan and Asia-Pacific countries.

Data science education at Shiga University is attracting interest from not only universities but also high schools all over the country. Accordingly, Shiga University has entered into cooperative agreements to provide research guidance and teaching materials to high schools in Shiga, Hyogo, Kagawa, and Shimane Prefectures, most of which are Super Science Highschools, i.e., government-chartered high schools. Among these, Kanonji Daiichi High School

in Kagawa Prefecture has been designated as a priority quota school and holds a national statistics meeting every year. We are also collaborating with the private Mukogawa Women's University Junior High School and High School to contribute to a gender-equal society.

In the Faculty of Data Science, we encourage students to participate in internships from early in their education. This is because we believe that participation in internships enables students to consider their career path more concretely and choose their career path smoothly. In fact, the first class of students, who graduated in March 2021, has successfully found jobs. The job offer rate for university graduates in Japan was 96% in 2021, down 2.0 percentage points from the previous year because of COVID-19. The tourism and transportation businesses and aviation and food industries have been particularly affected. However, graduates from the Faculty of Data Science of Shiga University, including those entering into these industries, were hardly affected. The percentages of graduates employed by the telecommunication, manufacturing, marketing consulting, financial, transportation, and other industries were 41%, 26%, 11%, 4%, 4%, and 14%, respectively (students who moved on to graduate school are excluded). These figures indicate the demand for data scientists across a wide range of industries.

4.2 Collaboration with industry

A potentially huge demand is seen from companies to apply the basic research conducted in universities and research institutes to their own business. Because many companies still lack data science personnel and experience, our faculty members are supporting the use of data science in companies in various ways, such as through joint research. In this way, we are expanding the range of applications in companies and the career paths of our graduates. The employees are learning data science while solving problems in the workplace not only through joint research, but also by studying in our graduate school. Conversely, companies give lectures to our students and provide us with data and assignments for exercises and PBL practices.

In 2017, the faculty members of Shiga University founded the General Incorporated Association Omi Data Science Initiative (President: Akimichi Takemura) to promote the advancement of data science technology through industry–academia collaboration. In July 2020, this association and Shiga University jointly established the Shiga University Data Science Collaboration Consortium. The consortium, which consists of students and faculty members of Shiga University, as well as researchers from other universities, provides a venue for mutual exchange among students, faculty members, and companies, as well as information on data science education and research.

5 Summary

In this paper, we have described the Japanese government's policy on data science education, the establishment of the Faculty of Data Science of Shiga

University, its curriculum, and its collaboration with industry and other universities. Our endeavor to promote data science in Japan has been bearing fruit, thereby raising the standard of data science in Japan. Stimulated by the success of the Faculty of Data Science of Shiga University, several universities in Japan have established faculties specializing in data science. However, Shiga University is still a front-runner in both data science education and collaboration with industry. Shiga University will continue its efforts to promote data science education in Japan and communicate it internationally.

Acknowledgments. The authors thank Professor Akimichi Takemura for his careful reading of the manuscript and helpful comments. This work was supported by JSPS KAKENHI Grant Number JP18H04092. On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Bughin, J., Seong, J., Manyika, J., Chui, M., Joshi, R. (2018). *Notes from the AI frontier: Modeling the impact of AI on the world economy*. Report prepared by the McKinsey Global Institute. Retrieved from <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy>
- Byers, A. (2014). Big data, big economic impact? *A Journal of Law and Policy for the Information Society*, 10(3), 757–764.
- Davenport, T.H., & Patil, D. (2012). Data scientist: the sexiest job of the 21st century. *Harvard Business Review*, October 2012 issue, 70–76.
- Department for Digital, Culture, Media & Sport, Department for Education, Department for Business and Energy & Industrial Strategy, Office for Artificial Intelligence. (2020, June 10). *2,500 new places on artificial intelligence and data science conversion courses now open to applicants*. Retrieved from <https://www.gov.uk/government/news/2500-new-places-on-artificial-intelligence-and-data-science-conversion-courses-now-open-to-applicants> (accessed: December 13, 2021)
- European Commission (2014, July 2). *Commission urges governments to embrace potential of big data*. Retrieved from https://ec.europa.eu/commission/presscorner/detail/en/IP_14_769 (accessed: December 13, 2021)
- Gardner, L., Ratcliff, J., Dong, E., Katz, A. (2021). A need for open public data standards and sharing in light of COVID-19. *The Lancet Infectious*

10 *Shiga University's endeavor in data science education in Japan*

Diseases, 21(4), e80.

[https://doi.org/10.1016/S1473-3099\(20\)30635-6](https://doi.org/10.1016/S1473-3099(20)30635-6)

Integrated Innovation Strategy Promotion Council (2019, June 11). *AI Strategy*. Retrieved from <https://www8.cao.go.jp/cstp/ai/aistratagy2019en.pdf> (accessed: December 13, 2021)

Irizarry, R.A. (2020). The role of academia in data science education. *Harvard Data Science Review*, 2(1).

<https://doi.org/10.1162/99608f92.dd363929>

Japan DataScientist Society (2019, November). *Results of a survey on data scientist recruitment*. Retrieved from https://www.datascientist.or.jp/common/docs/c-research_2019.pdf (in Japanese, accessed: December 13, 2021)

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H. (2011). *Big data: the next frontier for innovation, competition, and productivity*. Report prepared by the McKinsey Global Institute.

Maruyama, H., Kamiya, N., Higuchi, T., Takemura, A. (2015). Developing data analytics skills in Japan: Status and challenge. *Journal of Japan Industrial Management Association*, 65(4E), 334–339.

<https://doi.org/10.11221/jima.65.334>

Ministry of Education, Culture, Sports, Science and Technology (2019, December 3). *High school mathematics in new courses of study*. Retrieved from https://www.kantei.go.jp/jp/singi/ai_senryaku/suuri_datascience_ai/dai2/sankou2.pdf (in Japanese, accessed: December 13, 2021)

Mizuho Research and Technologies (2019, March). *The survey report for the supply and demand of IT professional*. Retrieved from https://www.meti.go.jp/policy/it_policy/jinzai/houkokusyo.pdf (in Japanese, accessed: December 13, 2021)

National Academies of Sciences, Engineering, and Medicine (2018). *Data science for undergraduates: opportunities and options*. The National Academies Press. Retrieved from <https://doi.org/10.17226/25104>

Naur, P. (1966). The science of datalogy. *Communications of the ACM*, 9(7), 485.

<https://doi.org/10.1145/365719.366510>

- Takemura, A. (2018). A new era of statistics and data science education in Japanese universities. *Japanese Journal of Statistics and Data Science*, *1*(1), 109–116.
- UN Global Pulse (2021, March 31). *UN Global Pulse annual report 2020*. Retrieved from <https://www.unglobalpulse.org/wp-content/uploads/2021/03/UNGP-2020-Annual-Report-Spread-view.pdf> (accessed: December 13, 2021)
- Wild, C.J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–265.