

Synthesis of Data Science Competency for Higher Education Students

Sajeewan Pratsri, Prachyanun Nilsook, Panita Wannapiroon
Division of Information and Communication Technology for Education,
Faculty of Technical Education,
King Mongkut's University of Technology North Bangkok, 10800
Thailand

Received: June 19, 2021. Revised: December 17, 2021. Accepted: January 5, 2021. Published: January 31, 2022.

Abstract—The research aims to Data Science Performance Synthesis for Higher Education Students and Data Science Performance Suitability Assessment for Higher Education Students. The research instruments include 1) data science performance synthesis tables, 2) expert interviews in data science performance assessments, 3) expert questionnaires to assess the consistency of data science performance. Analytical methods include 1) analyzing the frequency obtained from the content analysis table, 2) synthesis of content from interviews, 3) analyzing performance consistency, and components of data science performance, from data science synthesis for higher education students, finding that data performance for higher education students consists of five performances: 1) programming skills, 2) elementary statistics, 3) fundamentals of data science, 4) data preparation, and 5) Big data analytics.

Keywords— Data Science, Data Science Competency, Higher Education Students , Teaching Materials.

I. INTRODUCTION

THIS world is experiencing a digital socioeconomic era, where digital technology is a tool to not only support work but also restructure economic activity patterns, fully assimilating into people's daily lives. Thailand must accelerate the presentation of digital technology as an important tool to drive the country's development, as per the Thailand Digital Economy and Society Development Plan. The Ministry of Information and Communication Technology stated that regarding Strategy 5, on the development of peoplepower that is ready for the digital economy and society, there will be a focus on the development of working-class workforce across all professions. Both public and private sector workers have the ability to create and use digital technology wisely in their

profession, and people directly trained in the field of digital technology can offer knowledge, capability and expertise at international standards to contribute to the creation and employment of high value in an economic and social era that uses digital technology as a driving factor. This strategy comprises 1) developing digital technology skills for people in the labor market, including government personnel and the private sector. 2) promoting the development of specialized technology skills for people in the digital technology profession who work in the public and private sectors and 3) developing information technology executives to implement the digital technology to develop missions as well as create value from the data of the organization. Organizations in modern times need to adapt to keep pace with technological advancements. Ingenious organizations will quickly learn and adapt to change. Ideas to develop organizations into 'modern' organizations will partly drive knowledge management, so there needs to be a clear vision for ways to build organizations, systematically assembling the ideas, skills, and expertise of personnel. This is a process that involves conveying ideas, knowledge, and skills to personnel and one that can lead to practice [1], application, modification, and improvement even before new knowledge is acquired.

As a result of the advancement of digital technology and the Covid-19 pandemic effects , there is competition to develop software and hardware. Both public and private agencies use the Internet as a communication tool, disseminating news, trading products and so on. Owing to the Internet's speed and convenience, there has been an increase in its number of users. As a result, digital technology has played an important role in the education industry. Instructors seek ways to manage learning that involves the instructor and learner being in separate places, also known as distance or remote learning

management. As a result, many administrators, and instructors (satellite or digital TV teleconferencing) and teleconferencing as a substitute for classroom lectures. Starfish Academy develops instructors and designs online learning activities, distributing tablets to children to enable remote learning. Schools may require the cooperation of many parties. In addition to informing people on the approach to teaching, the company ensures the readiness of students and parents regarding equipment that is used for learning. Access to the Internet allows for convenient channels to communicate, for parents to take care of their children and parental responsibilities, for digital reading and writing in the home, and for developing instructors.

II. RESEARCH OBJECTIVES

The research aims to Data Science Competency Synthesis for Higher Education Students and Data Science Competency Suitability Assessment for Higher Education Students.

III. LITERATURE REVIEW

Data Science

Data science is the utilization of data, covering data collection (collect), data management (manage), data analysis and decision-making, with cross-discipline skill sets found at intersections of computer programming statistics and domain expertise [2]. This consists of two distinct but overlapping areas: 1) statistics to model and summarize datasets, 2) computer science to design and use algorithms to store, process and visualize. [3] reply by state that information science is essential for teaching and learning, especially for undergraduates, because society is now driven by a large archive [4]. It involves basic database concepts and the connections between database concepts and functionality. [5] They also believe that data science, data technology and data access play profession-specific roles [6]. The First role is that of data scientists, who identify and interpret complete data sources, manage big data, combine data sources, monitor dataset consistency and create images to help understand data. The second is that of data science researchers, who apply science research and discovery processes, including hypotheses and subsequent hypotheses testing, to obtain practical knowledge related to scientific issues, business processes or hidden relationships between processes Threeth is the data science programme designers, who develop and code big data and perform analytics to support science or organizations, and The Fourth role is that of data and business analysts, who analyse data to retrieve information about service systems and organizational competency, and present them in a functional format. [7] It has been said that data science is the extraction of knowledge that obtained directly from data through the process of discovering or determining hypotheses and hypothesis testing. [8] Data science is therefore of interest to many professionals, regardless of their field of study and their academic past. This interest has been accompanied by a vast increase in the number of job offers in this field, but in recent years offers have been found to provide too-high incentives [9]. An analysis of some of the proposals

for data science positions recently published on the platform revealed that the criteria include understanding of and experience in using high-competency machine learning algorithms and deep learning techniques, experience in programming languages such as Python R or Java, knowledge of relational and non-relational databases, and experience using statistical techniques to analyse data and produce continuous reports. [10] Data science is an all-in-one field that includes data analysis, mining statistics, data mining and models. It aims to turn data into knowledge by searching for patterns and trends in the said data. [11] The term originates from a description of the role of a single person who can act as a business analyst, statistician, engineer, and research scientist. Data scientists are skilled at identifying questions that can be asked about specific data, translating these questions into a series of assumptions, writing computer code that will prepare and question data, and performing analyses to answer questions and test hypotheses [12]. Students in the science domain can be data scientists. Human intelligence in the form of specialized analysis teams is an important resource component for creating a sustainable competitive advantage in business analysis projects. [5] In conclusion, data science is focused on analyzing, collecting, and cleaning data, and involves a wide range of expert approaches (defined as key professional groups), including data science infrastructure managers, data science experts, data science technology experts, data access experts, and providers of information.

Data Science Competency

Data science competency is a framework for data science teams and data ornithologists. The competency framework introduces information science studies and training to enhance the comprehensive knowledge and capabilities of qualified data science ornithologists, including knowledge, skills and experience that are important in information science, basic theory, engineering techniques, ready-to-work status, communication skills and data science competency. [13] It has been said that data science competency comprises management and leadership. The framework of practical data science courses is structured as follows: 1) determining what makes a new generation of qualified data ornithologists, (2) identifying gaps and problems in existing data science courses, (3) exploring how to encourage the next generation of data ornithologists and (4) identifying how to enable students to cope with unknown problems and unpaved knowledge and to gain the necessary new knowledge. [14] The framework of information science qualifications and paves the way from undergraduate studies to master's and PhD degrees, with undergraduate degrees in data science generating knowledge, competence, and experience through three stages of training. The first stage involves learning the basics of information science, which promotes students' informative thinking and covers the basics on both information science and data science [15]. The second stage involves the transference of engineering knowledge and skills to students in terms of data processing, engineering, computation, and programming, as well as the exploration of findings [16]. The third stage

focuses on data science practices related to several major projects, internships and data science applications. [17] that the framework enhances the skills and experience necessary for ensuring results, understanding information, introducing project management, managing social and ethical issues and communicating actionable outcomes to stakeholders [18]. As such, graduates are ready for work in the field of corporate data science as data analysts or business analysts. [19] Is a data engineering analyst with a background in engineering, computer science, mathematics, statistics, physics, and other related fields, who is skilled at using datasets and streaming tools in competency. Data science involves producing big data that requires basic knowledge. [20] mention big data-related software: Spark, AWS and Hadoop, and [21] data science competency, the use of high-competency machine learning algorithms and deep learning techniques and Experience in programming languages such as Python R or Java, knowledge of relational and non-relational databases, and experience using statistical techniques to analyze data and produce reports.

IV. RESEARCH METHOD

The research process is divided into two phases:

Phase 1: Data Science Competency Synthesis for Higher Education Students.

This stage involved synthesizing data science competency for higher education students by collecting data from textbooks, academic documents and research articles from both domestic and international sources At this stage, the researchers analysed, synthesized and demonstrated the data science competency of higher education students, each of which was categorized according to knowledge, skills and

attitude. Capitalize only the first word in a paper title, except for proper nouns and element symbols. For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [8].

Phase 2: Data Science Competency Suitability Assessment for Higher Education

This stage involved assessing the appropriateness of the data science competency for higher education students A questionnaire about data science-synthesis for higher education students included a five-level checklist, in-depth interviews, and information on how to evaluate questionnaires from qualified students. Twelve persons were involved, comprising a higher education data science specialist and experts in higher education technology, who, based on the results of the study, created expert interviews to synthesize data science competency for higher education students. The experts analysed the IOC values to assess the consistency of competency. The interview form included 1) a data science competency table for higher education students, including five competencies related to programming skills, elementary statistics, data science foundation, data preparation and big data analytics and 2) a data science competency synthesis table for higher education students to assess data competency suitability for higher education students.

V. RESULTS

A. Synthesis of data science competency by synthesis of data science competency as shown in Table 1

Table 1 Synthesis of Data Science Competency.

Synthesis of Data Science Competency	[3] (IBM, 2020)	[22] (Fayyad & Hamutcu, 2020)	[23] (Tencies, 2017)	[24] (Dichev & Dicheva, 2017)	[25] (Donoho, 2017)	[26] (Soni et al., 2015)	[5] (Saltz & Grady, 2017)	[27] (Strawn, 2016)	[28] (Cao, 2019)	[2] (Broatch et al., 2019)	[12] (Dill-McFarland et al., 2020)	[29] (Hu et al., 2018)	[30] (Basic et al., n.d.)	Data Science Competency
	1. Programming Skills	√	√	√	√		√	√			√	√	√	√
2. Elementary Statistics	√	√	√	√		√	√			√	√	√	√	√
3. Fundamentals of Data Science	√	√	√	√	√	√		√	√	√	√	√		√
4. Data Preparation	√		√	√	√	√		√	√			√		√
5. Big Data Analytics	√	√				√		√			√			√

From Table 1 concludes that data science competency for higher education students includes 1) programming skills, 2) elementary statistics, 3) fundamentals of data science, 4) data preparation, and 5) big data analytics.

B. Data Science Competency for Higher Education Students

Table 2 Data Science Competency for Higher Education Students

Competency	Programming Skills	Elementary Statistics	Data science foundation	Data preparation	Big Data Analytics
K (knowledge)	1. It can explain the principles, methods and procedures of writing and development of programs. 2. The type of information can be described. 3. Programming basic control commands can be described. 4. The basic functions of programming can be described.	1. Meaning and introduction to statistics 2. Understand probability theory sampling and probability distribution 3. Point and range estimations can be described. 4. The probability distribution of random variables of the type can be discontinued and continuous. 5. It can explain the equation of regression simply.	1. Demonstrate understanding of what information science is and what data science does. 2. It can explain visualization and communication with information.	1. The meaning of data selection can be explained. 2 Demonstrates management capabilities Convert and clean up	1. Understand the concept of big data and how to use big data in the organization. 2. Tell the elements Features and software that can be used to analyze big data 3. Explain the principles of operation of the program to manage big data.
S (Skills)	5. Demonstrate conditional programming skills It can be looped and subprogrammed.	6. Probability Distribution 7. The probability distribution of random variables of the type can be calculated discontinued and continuously. 8. Hypothesis on parameters of one or two populations can be tested.	3. Demonstrate the ability to define problems 4. Demonstrate the ability to collect and explore data.	3. Demonstrate the ability to organize changes and clean data.	4. Programs can be used to manage big data for data management and analysis of information tasks.
A (Attitude)	6. Have a good attitude towards studying, value learning and utilize the work of programming.		4. Value data sciencers to the organization		5. Have a good attitude and be aware The importance of big data management

From Table 2 is a table of which is divided by knowledge, skills, and attitudes according to each of the main competencies, and consists of detailed measurement and competency evaluation based on lessons and basic abilities, as shown in Table 3.

Table 3 Details of measuring and evaluating competency based on lessons and ground capabilities

Data Science Competency	Basic competency criteria and capabilities	Lesson
1. Programming Foundation	1. It can explain the principles, methods and procedures of writing and development of programs. 2. Describe the artificial coding and flowcharting process. 3. The type of information can be described. 4. Demonstrate conditional programming skills It can be looped and subprogrammed. 5. Have a good attitude towards studying, value learning and utilize the work of programming.	Competency 1 programming foundation 1. Python Programming Basics 1.1 Types of data and variables 1.2 Expressions and operators 1.3 Display function 1.4 Receive functions 1.5 Mathematical functions 1.6 Text handling function 2. Select make and loop commands 2.1 Selective command (if statement) 2.2 Loop Statement 2.3 Break command 2.4 Continue command 3. Function programming 3.1 Functions 3.2 Activating functions 3.3 Parameters of function 3.4 Restoring from functions 3.5 Variable scope
2. Statistics foundation	1. Meaning and introduction to statistics 2. Understand probability theory sampling and probability distribution 3. Point and range estimations can be described. 4. The probability distribution of random variables of the type can be discontinued and continuous. 5. Probability Distribution 6. The probability distribution of random variables of the type can be calculated discontinued and continuously. 7. Hypothesis on parameters of one or two populations can be tested.	Competency 2 Statistics foundation 1. Basic Statistics 1.1 Basic theory 1.2 Why statistics?) 1.3 Population and Sample 1.4 Probability Distribution 1.5 Sampling Distributions 2. Data & Data presentation 2.1 Data Type 2.2 Measure of location 2.3 Measure of Dispersion 2.4 Normal distribution test 3. Statistical Inference for a Single Sample 3.1 Hypothesis Testing 3.2 Testing hypotheses when knowing Variance (1- Sample Z test) 3.3 Testing hypotheses when variance is not known (1-Sample t-test) 3.4 Variance test
3. Data science foundation	1. Demonstrate understanding of what information science is and what data science does. 2. It can explain visualization and communication with information. 3. Demonstrate the ability to define problems 4. Demonstrate the ability to collect and explore data. 4. Value data sciences to the organization	Competency 3 Data science foundation 1. Meaning of Data Science 1.1 The era of information and information 1.2 Data Science 1.3 Data Science Process 1.4 Design thinking for data science 2. Data Collection and Survey 2.1 Data collection 2.2 Data preparation 2.3 Data Exploration 2.4 Personal information 3. Data Analysis 3.1 Descriptive analysis 3.2 Predictive analysis 4. Visualization and Communication with Information 4.1 Data communication 4.2 Data Visualization 4.3 Proper visualization 4.4 Data-based storytelling 4.5 Precautions for presenting information

Data Science Competency	Basic competency criteria and capabilities	Lesson
4. Data preparation	1. The meaning of data selection can be explained. 2. Demonstrates management capabilities Convert and clean up 3. Demonstrate the ability to organize changes and clean data.	Competency 4 Data preparation 1. Data Selection 1.1 Parsing 1.2 Correcting 1.3 Standardizing 1.4 Duplicate Elimination 2. Data transformation 2.1 Data collection 2.2 Data Exploration 2.3 Check data 3. Data link (combining data) 3.1 Descriptive analysis 3.2 Predictive analysis
5. Big Data Analytics	1. Understand the concept of big data and how to use big data in the organization. 2. Tell the elements Features and software that can be used to analyze big data 3. Explain the principles of operation of the program to manage big data. 4. Programs can be used to manage big data for data management and analysis of information tasks. 5. Have a good attitude and are aware of the importance of big data management.	Competency 5 Big Data Analytics 1. The meaning of big data 1.1 Key elements of data 1.2 Key characteristics of big data 1.3 Evolution of big data 1.4 Large data format 2. Exploit various aspects of big data 2.1 The process of from big data to data relationships 2.2 How to make big data

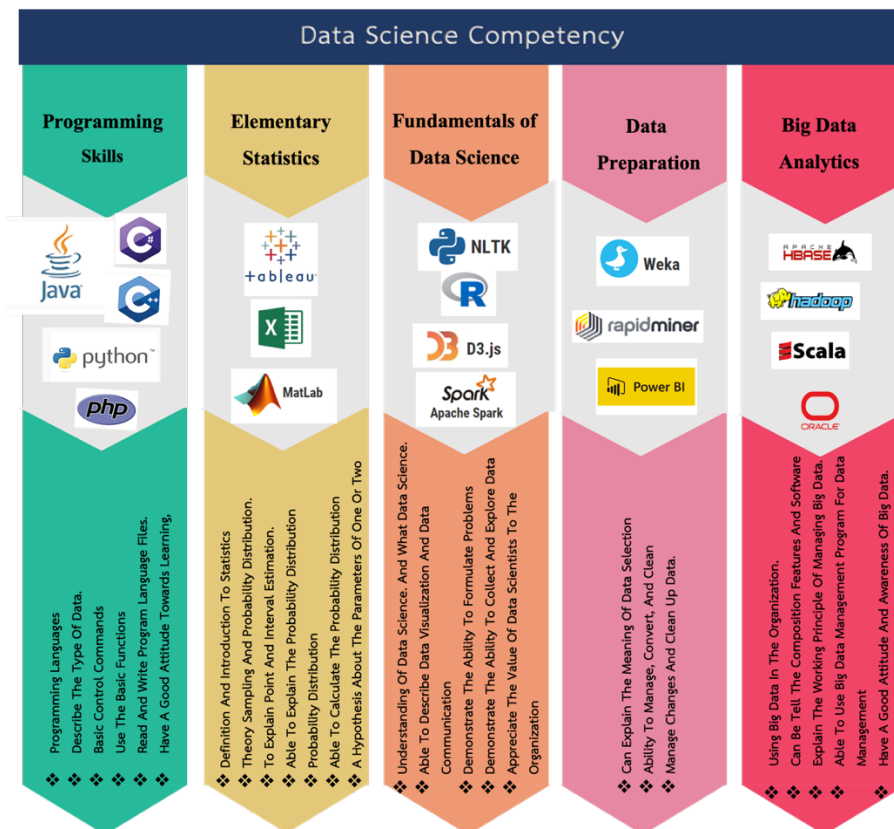


Figure 1 Data Science Competency for Higher Education Students and Teaching Materials

Figure 1 details the data science competency for higher education students, consisting of five competencies. The first refers to programming skills, which addresses the principles, methods and procedures for writing and developing

programmes, and describes the artificial coding and flowcharting process. The specific type of information is described here, which can be looped and sub programmed, and it demonstrates conditional programming skills. Students can

give meanings to the information and gain an introduction to programming, relating to Python, C Plus and Java. The second competency is elementary statistics. Understanding probability theory sampling and probability distribution can lead to explaining point and interval estimations. The probability distribution of random variables of this type is discontinuous and continuous and can be calculated discontinuously and continuously. Hypotheses about the parameters of one or two populations, such as Microsoft Office Excel, MATLAB or Tableau. The third competency relates to the fundamentals of data science, demonstrating an understanding of what data science is and what data science does, as well as an ability to define problems, collect and explore data and provide high-value data science to organizations, including Apache Spark,

R language, Natural Language Toolkit and d3.js. Data preparation is the fourth competency and involves explaining the meaning of data selection, demonstrating management capabilities Convert and clean up data and the ability to organize changes and clean data using the programmes Vega and RapidMiner. The fifth competency is big data analytics – understanding the concept of big data and how big data is used in an organization, including the features and software that can be used to analyse big data, and describing the principles of how a programme works in big data management. The programmes used include Apache HBASE, Hadoop, Oracle, and Scala.

Table 4 Evaluation of The Suitability of Data Science for Higher Education Students

Data Science for Higher Education Students	IOC	Consideration
Competency 1 programming foundation		
1. Have an introduction to language programs	0.83	Appropriate
2. The type of information can be described.	0.75	Appropriate
3. Use the basic control command of the language program.	0.83	Appropriate
4. Use the basic functions of the language program.	0.92	Appropriate
5. Can read and write files of language programs.	0.92	Appropriate
6. Have a good attitude towards studying, value learning and utilize the work of programming.	0.83	Appropriate
	0.85	Appropriate
Competency 2 Elementary Statistics		
1. Meaning and introduction to statistics	0.83	Appropriate
2. Understand probability theory sampling and probability distribution	0.83	Appropriate
3. Point and range estimations can be described.	0.92	Appropriate
4. The probability distribution of random variables of the type can be discontinued and continuous.	0.83	Appropriate
5. Probability Distribution	0.83	Appropriate
6. The probability distribution of random variables of the type can be calculated discontinued and continuously.	0.92	Appropriate
7. Hypothesis on parameters of one or two populations can be tested.	0.75	Appropriate
	0.84	Appropriate
Competency 3 Fundamentals of Data Science		
1. Demonstrate understanding of what information science is and what data science does.	0.92	Appropriate
2. It can explain visualization and communication with information.	0.75	Appropriate
3. Demonstrate the ability to define problems	0.92	Appropriate
4. Demonstrate the ability to collect and explore data.	0.75	Appropriate
5. Value data sciencers to the organization	0.92	Appropriate
	0.85	Appropriate
Competency 4 Data Preparation		
1. The meaning of data selection can be explained.	0.75	Appropriate
2 Demonstrates management capabilities Convert and clean up	0.75	Appropriate
3. Demonstrate the ability to organize changes and clean data	0.92	Appropriate
	0.81	Appropriate
Competency 5 Big Data Analytics		
1. Understand the concept of big data and how to use big data in the organization.	0.92	Appropriate
2. Tell the elements Features and software that can be used to analyze big data	0.83	Appropriate
3. Explain the principles of operation of the program to manage big data.	1.00	Appropriate
4. Programs can be used to manage big data for data management and analysis of information tasks.	0.92	Appropriate
5. Have a good attitude and are aware of the importance of big data management.	0.83	Appropriate
	0.90	Appropriate
Total	0.85	Appropriate

I. DISCUSSION

This research presents elements of data science competency for higher education students, consisting of five competencies. The first refers to programming skills, which addresses the principles, methods and procedures for writing and developing programmed, and describes the artificial coding and flowcharting process. The specific type of information is described here, which can be looped, and sub programmed, and it demonstrates conditional programming skills. Students can give meanings to the information and gain an introduction to programming, relating to Python, C Plus and Java. The second competency is elementary statistics. Understanding probability theory sampling and probability distribution can lead to explaining point and interval estimations. The probability distribution of random variables of this type is discontinuous and continuous and can be calculated discontinuously and continuously. Hypotheses about the parameters of one or two populations, such as Microsoft Office Excel, MATLAB or Tableau. The third competency relates to the fundamentals of data science, demonstrating an understanding of what data science is and what data science does, as well as an ability to define problems, collect and explore data and provide high-value data science to organizations, including Apache Spark, R language, Natural Language Toolkit and d3.js. Data preparation is the fourth competency and involves explaining the meaning of data selection, demonstrating management capabilities Convert and clean up data and the ability to organize changes and clean data using the programmed Vega and RapidMiner. The fifth competency is big data analytics – understanding the concept of big data and how big data is used in an organization, including the features and software that can be used to analyze big data, and describing the principles of how a programmed works in big data management. The programmed used include Apache HBASE, Hadoop, Oracle and Scala. consistent with [13] He said that undergraduate students in information science may build knowledge and experience through three stages of training: 1) learning the basics of information science, which promotes students through information science thinking, basic information science and the basics of data science, 2) transferring engineering knowledge and skills to students in data processing, engineering, computational and programming, as well as survey discovery, and 3) focusing on data science practices related to several important projects, internships and the application of data science. And consistent with [2] that information science is an essential skill for teaching and learning, especially for undergraduates, to be known and used because society is now driven by a large archive. Contains basic database concepts and connections between database concepts and functionality.

ACKNOWLEDGMENT

The researcher would like to thank Thepsatri Rajabhat University and King Mongkut's University of Technology North Bangkok that supported this research.

REFERENCES

- [1] S. Pratsri, P. Nilsook, and P. Wannapiroon, "Developing a Conceptual Framework for Remote Practice Learning," pp. 140–144, 2021, doi: 10.1145/3481056.3481104.
- [2] J. E. Broatch, S. Dietrich, and D. Goelman, "Introducing Data Science Techniques by Connecting Database Concepts and dplyr," *J. Stat. Educ.*, vol. 27, no. 3, pp. 147–153, 2019, doi: 10.1080/10691898.2019.1647768.
- [3] IBM, "The Data Science Skills Competency Model," 2020.
- [4] U. Ramnarain and M. Hlatswayo, "Teacher beliefs and attitudes about inquiry-based learning in a rural school district in South Africa," *South African Journal of Education*, vol. 38, no. 1, 2018, doi: 10.15700/saje.v38n1a1431.
- [5] J. S. Saltz and N. W. Grady, "The ambiguity of data science team roles and the need for a data science workforce framework," *Proc. - 2017 IEEE Int. Conf. Big Data, Big Data 2017*, vol. 2018-Janua, pp. 2355–2361, 2017, doi: 10.1109/BigData.2017.8258190.
- [6] T. Meepung, S. Pratsri, and P. Nilsook, "Interactive Tool in Digital Learning Ecosystem for Adaptive Online Learning Performance," *High. Educ. Stud.*, vol. 11, no. 3, p. 70, 2021, doi: 10.5539/hes.v11n3p70.
- [7] Y. Demchenko *et al.*, "EDISON Data Science Framework (EDSF): Customising Education and Training for Career Development and Capacity Building," vol. 675419, no. 675419, p. 675419, 2015.
- [8] P. Pereira, "Towards Helping Data Scientists," *Proc. IEEE Symp. Vis. Lang. Human-Centric Comput. VL/HCC*, vol. 2020-Augus, pp. 16–17, 2020, doi: 10.1109/VL/HCC50065.2020.9127198.
- [9] S. R. Johnson, E. T. Pas, and C. P. Bradshaw, "Understanding and measuring coach-teacher alliance: A glimpse inside the 'black box,'" *Prev. Sci.*, vol. 17, no. 4, pp. 439–449, 2016, doi: 10.1007/s11121-016-0633-8.

- [10] G. J. Miller, "The influence of big data competencies, team structures, and data scientists on project success," *2019 IEEE Technol. Eng. Manag. Conf. TEMSCON 2019*, 2019, doi: 10.1109/TEMSCON.2019.8813604.
- [11] G. U. Interface, "Data Science for," no. July, pp. 1–10, 2018.
- [12] K. A. Dill-McFarland *et al.*, "An integrated, modular approach to data science education in the life sciences," *bioRxiv*, pp. 1–20, 2020, doi: 10.1101/2020.07.25.218453.
- [13] L. Cao, "Data Science: Profession and Education," *IEEE Intell. Syst.*, vol. 34, no. 5, pp. 35–44, 2019, doi: 10.1109/MIS.2019.2936705.
- [14] L. Cao, "Data science: A comprehensive overview," *ACM Comput. Surv.*, vol. 50, no. 3, 2017, doi: 10.1145/3076253.
- [15] K. T. Rodolfa, A. de Unanue, M. Gee, and R. Ghani, "A Clinical Approach to Training Effective Data Scientists," *arXiv*, 2019.
- [16] M. Al-Emran, V. Mezhuyev, and A. Kamaludin, "Technology Acceptance Model in M-learning context: A systematic review," *Comput. Educ.*, vol. 125, pp. 389–412, Oct. 2018, doi: 10.1016/j.compedu.2018.06.008.
- [17] A. Farahi and J. C. Stroud, "The Michigan Data Science Team: A Data Science Education Program with Significant Social Impact," *2018 IEEE Data Sci. Work. DSW 2018 - Proc.*, pp. 120–124, 2018, doi: 10.1109/DSW.2018.8439915.
- [18] N. Rafique, "Education, Political Awareness, and Political Participation: a Case of Rahim Yar Khan District of Pakistan," *Int. J. Educ. Inf. Technol.*, vol. 15, pp. 372–384, 2021, doi: 10.46300/9109.2021.15.39.
- [19] P. Pereira, J. Cunha, and J. P. Fernandes, "On Understanding Data Scientists," *Proc. IEEE Symp. Vis. Lang. Human-Centric Comput. VL/HCC*, vol. 2020-Augus, 2020, doi: 10.1109/VL/HCC50065.2020.9127269.
- [20] S. Pratsri and P. Nilsook, "Design on Big data Platform-based in Higher Education Institute," *High. Educ. Stud.*, vol. 10, no. 4, p. 36, 2020, doi: 10.5539/hes.v10n4p36.
- [21] K. S. Praharshita, S. S. Aravabhumi, S. Attaluri, S. Mandava, S. Raghavendran, and S. K. Hasane Ahammad, "Bigdata and machine learning models for dimensionality reduction platform," *Journal of Critical Reviews*, vol. 7, no. 1, pp. 449–452, 2020, doi: 10.31838/jcr.07.01.88.
- [22] U. Fayyad and H. Hamutcu, "Analytics and Data Science Standardization and Assessment Framework," *Harvard Data Sci. Rev.*, pp. 1–33, 2020, doi: 10.1162/99608f92.1a99e67a.
- [23] L. C. Tencies, "Data Science Competency Framework," *Australia*, 2017.
- [24] C. Dichev and D. Dicheva, "Towards Data Science Literacy," *Procedia Comput. Sci.*, vol. 108, no. December, pp. 2151–2160, 2017, doi: 10.1016/j.procs.2017.05.240.
- [25] D. Donoho, "50 Years of Data Science," *J. Comput. Graph. Stat.*, vol. 26, no. 4, pp. 745–766, 2017, doi: 10.1080/10618600.2017.1384734.
- [26] M. Soni, H. Singh, and N. Sethi, "A state of the art survey of data mining techniques for software engineering data," *Int. J. Appl. Eng. Res.*, vol. 10, no. 55, pp. 1512–1522, 2015.
- [27] G. Strawn, "Data Scientist," *IT Prof.*, vol. 18, no. 3, pp. 55–57, 2016, doi: 10.1109/MITP.2016.41.
- [28] L. Cao, "Data Science: Profession and Education," *IEEE Intell. Syst.*, vol. 34, no. 5, pp. 35–44, 2019, doi: 10.1109/MIS.2019.2936705.
- [29] H. Hu, Y. Luo, Y. Wen, Y. S. Ong, and X. Zhang, "How to Find a Perfect Data Scientist: A Distance-Metric Learning Approach," *IEEE Access*, vol. 6, pp. 60380–60395, 2018, doi: 10.1109/ACCESS.2018.2870535.
- [30] D. Basic, I. Advanced, and E. Participates, "Learning and Information Technology Standard Competency Matrix."

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US