



ORIGINAL

Thematic Specialization of Institutions with Academic Programs in the Field of Data Science

Especialización temática de las instituciones con programas académicos en el área de las Ciencias de Datos

Denis Gonzalez-Argote¹   

¹Universidad Argentina de la Empresa, Facultad de Ingeniería y Ciencias Exactas, Carrera de Ingeniería Informática. Ciudad Autónoma de Buenos Aires, Argentina.

Cite as: Rojas Gonzalez-Argote D. Thematic Specialization of Institutions with Academic Programs in the Field of Data Science. Data & Metadata. 2023;2:24. <https://doi.org/10.56294/dm202324>

Submitted: 23-01-2023

Revised: 10-02-2023

Accepted: 14-03-2023

Published: 15-03-2023

Editor: Prof. Dr. William Castillo González 

ABSTRACT

Introduction: data science careers are on the rise due to the growing demand for technical skills in this area. Data science careers focus on collecting, organizing, and analyzing data to identify patterns and trends, which allows organizations to make informed decisions and develop effective solutions.

Aim: to analyze the thematic specialization of institutions with academic programs in the area of data science.

Methods: the Scopus database was used to conduct a bibliometric analysis aimed at examining the thematic specialization of institutions with academic programs in the field of data science. SciVal, a bibliometric analysis tool, was employed to extract the relevant data. The study period ranged from 2012 to 2021.

Results: nine higher education institutions were found to offer undergraduate or graduate degrees in the field of data science. There was no correlation found between RSI and Field-Weighted Citation Impact ($r=0,05355$; $P=0,8912$; 95%CI: $-0,6331$ to $0,6930$). Therefore, it cannot be claimed that specialization in the subject area studied influences the greater impact of research. On the other hand, recent accreditation did not influence greater specialization ($r=0,1675$; $P=0,6667$; 95%CI: $-0,5588$ to $0,7484$). Additionally, no differences were found regarding academic level.

Conclusions: the analysis of the thematic specialization of institutions with academic programs in the field of data science shows low scientific production in this field. Moreover, more than half of the analyzed higher education institutions have thematic specialization below the global average. This suggests that there is still a long way to go for these institutions to achieve adequate specialization and compete internationally in the field of data science.

Keywords: Data Science; Bibliometrics; Thematic Specialization; Higher Education; University; University Accreditation.

RESUMEN

Introducción: las carreras en ciencias de datos están en aumento debido a la creciente demanda de habilidades técnicas en esta área. Las carreras en ciencias de datos se enfocan en la recolección, organización y análisis de datos para identificar patrones y tendencias, lo que permite a las organizaciones tomar decisiones informadas y desarrollar soluciones efectivas.

Objetivo: analizar la especialización temática de las instituciones con programas académicos en el área de ciencias de datos.

Métodos: se utilizó la base de datos Scopus para realizar un análisis bibliométrico con el objetivo de examinar la especialización temática de las instituciones con programas académicos en el campo de las ciencias de datos. Se empleó SciVal, una herramienta de análisis bibliométrico, para extraer los datos relevantes. El

período de estudio abarcó desde 2012 hasta 2021.

Resultados: se encontró que nueve instituciones de educación superior ofrecen programas de licenciatura o posgrado en el campo de las ciencias de datos. No se encontró correlación entre el Índice de Especialización en Investigación (RSI) y el Impacto de Citas Ponderado por Campo ($r=0,05355$; $P=0,8912$; IC95%: $-0,6331$ a $0,6930$). Por lo tanto, no se puede afirmar que la especialización en el área estudiada influya en un mayor impacto de la investigación. Por otro lado, la acreditación reciente no influyó en una mayor especialización ($r=0,1675$; $P=0,6667$; IC95%: $-0,5588$ a $0,7484$). Además, no se encontraron diferencias en cuanto al nivel académico.

Conclusiones: el análisis de la especialización temática de las instituciones con programas académicos en el campo de las ciencias de datos muestra una baja producción científica en este campo. Además, más de la mitad de las instituciones de educación superior analizadas tienen una especialización temática por debajo del promedio mundial. Esto sugiere que aún queda un largo camino por recorrer para que estas instituciones logren una especialización adecuada y compitan a nivel internacional en el campo de las ciencias de datos.

Palabras clave: Ciencias de datos; Bibliometría; Especialización temática; Educación superior; Universidad; Acreditación Universitaria.

INTRODUCTION

Data science is a branch of computer science and statistics that focuses on the analysis and interpretation of large datasets. The explosion of data generated by technology has led to an increasing demand for data scientists who can handle large amounts of information and extract valuable insights from it. In an increasingly data-driven world, data science has become an essential discipline for businesses and organizations seeking a competitive edge.^(1,2)

Data science careers are on the rise due to the growing demand for technical skills in this area. Data science careers focus on collecting, organizing, and analyzing data to identify patterns and trends, which allows organizations to make informed decisions and develop effective solutions. Training in data science is essential for those who want to work in the technology and computer industry, as well as in the business and government sectors.⁽³⁾

Regarding undergraduate degrees, many universities offer data science programs that provide students with a solid foundation in mathematics, statistics, and programming. Data science programs typically include courses in computer science, statistics, data analysis, machine learning, and data visualization. Graduates of these programs can work in a wide range of industries, including technology, finance, healthcare, and retail, among others.^(4,5)

For those seeking a more specialized career in data science, graduate programs are a popular option. Graduate programs in data science focus on more advanced skills in areas such as machine learning, data mining, and artificial intelligence. Students can choose from master's programs in data science, applied statistics, business analytics, and other related areas. Graduate programs may also include research projects in collaboration with professors and industry experts.^(6,7)

Training in data science is essential due to the growing need for data scientists in the job market. Companies and organizations worldwide need data science specialists who can analyze large datasets and extract valuable insights. Additionally, data scientists can help identify patterns and trends in data, allowing organizations to make informed decisions and develop effective solutions.⁽⁸⁾

Training in data science can also help students develop valuable skills in areas such as problem-solving, decision-making, communication, and collaboration. Students studying data science learn to work with large amounts of information and find patterns in the data that enable them to develop effective solutions.⁽⁹⁾

Additionally, data scientists must have effective communication skills to be able to convey their findings to colleagues and clients who may not necessarily have a technical background. Collaboration is also an important aspect in the field of data science, as data scientists often work in interdisciplinary teams with experts in different areas to solve complex problems.⁽⁹⁾

This article aims to analyze the thematic specialization of institutions with academic programs in the area of data science.

MÉTODOS

Data Source

The Scopus database was used to conduct a bibliometric analysis aimed at examining the thematic specialization of institutions with academic programs in the field of data science. SciVal, a bibliometric analysis tool, was employed to extract the relevant data. The study period ranged from 2012 to 2021.

Indicators

Three bibliometric indicators were used to analyze the data:

- Number of documents (Ndoc): Refers to the total number of publications produced by each institution during the study period (broken down into overall and in the field of computer science);
- Relative specialization rate: This is the relative effort that a country dedicates to a specific discipline.

The Thematic Specialization Index (TSI) is calculated as follows:

$$TSI = \frac{NdocCS(University)/Ndoc(University)}{NdocCS(World)/Ndoc(World)}$$

The Relative Specialization Index (RSI) is calculated as follows:

$$RSI = \frac{TSI - 1}{TSI + 1}$$

RSI values (between -1 and 1): 0 represented the world position; values above 0 indicate a greater specialization of scientific production with respect to the world; values below 0 indicate a lower specialization with respect to the world.

- Field-Weighted Citation Impact (FWCI): indicates how the number of citations received by an entity's publications compares with the average number of citations received by all other similar publications.

Sample Selection

The list of universities was extracted from the database of accredited careers by CONEAU with a current resolution.⁽¹⁰⁾ Only those institutions offering academic programs in the field of data science were included in the study.

Data Analysis

The data obtained from Scopus and SciVal were analyzed using descriptive and inferential statistics to characterize the thematic specialization of each institution in the field of data science. The results were presented using graphs and tables to facilitate the interpretation of the findings. The bibliometric indicators used in the study provided a comprehensive view of the research performance of each institution and allowed for the comparison of the relative strengths and weaknesses of different institutions in the area under study.

RESULTS

Nine higher education institutions were found to offer undergraduate or graduate degrees in the field of data science. Table 1 displays the main bibliometric indicators of these institutions in the macro area of knowledge related to data science, in this case, the Computer Science category, since there is no specific category for data science yet. Additionally, this table also presents the year of the last accreditation resolution by CONEAU, which was taken as the accreditation year.

Table 1. Main indicators of higher education institutions in the field of data science

Institution	Ndoc in Computer Science	Total Ndoc	TSI	RSI	Field-Weighted Citation Impact	Year of accreditation	Academic Level
Universidad Nacional del Sur	775	5537	1,08	0,04	0,67	2022	Posgrado
Instituto Tecnológico de Buenos Aires	168	509	2,55	0,44	0,78	2022	Posgrado
Universidad Nacional de General San Martín	117	3535	0,26	-0,59	1,12	2021	Grado
Universidad Austral	45	1284	0,27	-0,57	0,81	2022	Grado
Universidad de San Andrés	16	474	0,26	-0,59	1,16	2021	Posgrado
Universidad CAECE	11	42	2,03	0,34	0,23	2021	Grado y Posgrado

Universidad Empresarial Siglo 21	3	17	1,37	0,15	0	2021	Grado
Universidad del Gran Rosario	0	9	0,00	-1,00	0	2023	Grado
Universidad Nacional Guillermo Brown	0	4	0,00	-1,00	0	2019	Grado

There was no correlation found between RSI and Field-Weighted Citation Impact ($r=0,05355$; $P=0,8912$; 95%CI: $-0,6331$ to $0,6930$) (figure 1). Therefore, it cannot be claimed that specialization in the subject area studied influences the greater impact of research.

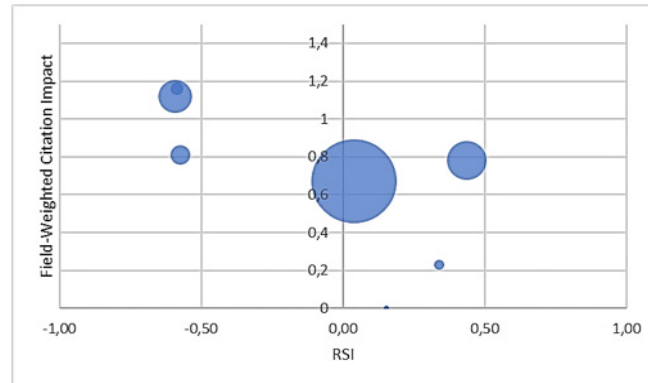


Figure 1. RSI and Field-Weighted Citation Impact in the Computer Science

On the other hand, recent accreditation did not influence greater specialization ($r=0,1675$; $P=0,6667$; 95%CI: $-0,5588$ to $0,7484$). Additionally, no differences were found regarding academic level (table 2).

Table 2. Differences between academic level and RIS and Field-Weighted Citation Impact

Academic level	RSI	ANOVA p value	Field-Weighted Citation Impact	ANOVA p value
Grade	-0,6020		0,3860	
Postgraduate	-0,03667	0,185	0,8700	0,352
Graduate and Postgraduate	0,3400		0,2300	

DISCUSSION

Data science is an essential discipline in a data-driven world. Careers in data science are on the rise due to the growing demand for technical skills in this area. Undergraduate and graduate programs in data science provide students with advanced skills in areas such as machine learning, data mining, and artificial intelligence. (11,12)

Training in data science is important because data scientists can help companies and organizations make informed decisions and develop effective solutions. In addition, students studying data science can develop valuable skills in areas such as problem-solving, decision-making, communication, and collaboration, preparing them for a wide range of careers and opportunities in the job market. (13)

Thematic specialization in university publications allows researchers to delve into a specific area of study and contribute to the understanding and advancement of that particular field; it also helps to develop a reputation and recognition in their field of study, which can be beneficial for their academic and professional career.

Based on the programs they offer, universities often focus on specific fields of study and seek to stand out in those areas. Specialized publications can help universities improve their reputation and position in the field of study, which can attract talented students and funding for research.

In this sense, studies supporting these findings are reported. Yoon and Suh analyzed the relationship between specialization and the productivity of publications in the academic careers of scientists in South Korea. The results indicate that thematic specialization improves productivity of publications and quality of research. (14)

Meanwhile, Leahey et al. (15) demonstrate that thematic specialization improves academic performance and research capacity of researchers. While Caviggioli et al. (16) analyze the impact of thematic specialization

on the development of economics departments in European universities. The results indicate that thematic specialization improves reputation and academic performance of economics departments.

Thematic specialization allows researchers to delve into a specific area of study, contribute to the advancement of the field, and develop a reputation in their field of study. Additionally, thematic specialization can improve the reputation and academic performance of universities. The scientific studies mentioned above support the importance of thematic specialization in publications at universities.^(17,18,19)

On another note, accreditation of a degree program is a process in which an independent organization evaluates and verifies the quality of education offered by an educational institution and its academic program. Accreditation of a degree program is important because it ensures that academic programs meet established quality standards and that graduates are prepared to enter the job market and perform effectively in their field of study.⁽²⁰⁾

The National Commission for University Evaluation and Accreditation (CONEAU) is an Argentine organization that aims to evaluate and accredit the quality of higher education institutions in the country. CONEAU works to improve the quality of higher education in Argentina, including scientific research in undergraduate and graduate degree programs.⁽²¹⁾

CONEAU can have a positive impact on increasing scientific publications in undergraduate and graduate degree programs in Argentina through its evaluation and accreditation processes of higher education institutions. Institutions seeking accreditation by CONEAU must meet certain quality standards and requirements, including scientific research and production of publications in undergraduate and graduate degree programs.^(22,23)

Additionally, CONEAU can offer financial support and resources for scientific research at accredited higher education institutions. This can include funding for research, resources for publication, and access to networks for scientific collaboration and exchange. The availability of these resources and support can be an incentive for students and academics in undergraduate and graduate degree programs to conduct research and publish in scientific journals.

Accreditation can have a positive impact on the quantity and quality of program publications. Accredited academic programs may have more resources and support to research and publish, which can lead to a greater number of publications in academic journals and greater visibility in the academic community.⁽²⁴⁾

Figueroa Escudero *et al.*⁽²⁵⁾ examined the similarities and differences in curricular policy between Mexico and Brazil, and their results showed that accreditation improves the academic quality of UNAM and increases the scientific production of the academic program. Soto *et al.*⁽²⁶⁾ concluded that accreditation improves the quality of higher education and increases scientific production in accredited academic programs. The scientific studies mentioned above support the relationship between accreditation and the increase in program publications.

Limitations of the study and future perspectives

This study has methodological limitations, including the fact that there was no specific thematic category for Data Science, and the category of Computer Science had to be used, which includes topics beyond Data Science. Additionally, there could be publications on Data Science outside the Computer Science field, but it is difficult to search and standardize them. From a coverage perspective, the main limitation is that this study was based only on the Scopus database and did not include other databases such as Web of Science or Compendex.

Despite these limitations, this study provides an approach to an issue in emerging careers, which, given their recent implementation, could develop strategies to increase publications and thematic specialization in the future.

CONCLUSIONS

The analysis of the thematic specialization of institutions with academic programs in the field of data science shows low scientific production in this field. Moreover, more than half of the analyzed higher education institutions have thematic specialization below the global average. This suggests that there is still a long way to go for these institutions to achieve adequate specialization and compete internationally in the field of data science.

The lack of adequate thematic specialization can be a significant factor that limits the quality of scientific production in the analyzed higher education institutions. Therefore, it is essential that these institutions work on continuous improvement of their academic programs and adequate thematic specialization in the field of data science. Additionally, it is important to establish closer links between academia and industry so that academic programs are in line with industry needs, and scientific research has greater practical application.

Regarding the future projection, it is expected that the analyzed higher education institutions will work on continuous improvement of their academic programs in the field of data science and greater thematic specialization. This will improve the quality of scientific production and increase competitiveness internationally. Additionally, more links are expected to be established between academia and industry so that academic programs are more relevant to the job market and innovative solutions are generated in the field of data

science.

The findings of this study suggest that there is a need for higher education institutions to improve their thematic specialization in the field of data science. By doing so, they can improve the quality of their scientific production and compete more effectively at the international level. Furthermore, establishing closer links between academia and industry will be crucial for ensuring that academic programs are relevant to the job market and that scientific research has greater practical application.

REFERENCES

1. Lee I. Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons* 2017;60:293-303. <https://doi.org/10.1016/j.bushor.2017.01.004>.
2. Ferraris A, Mazzoleni A, Devalle A, Couturier J. Big data analytics capabilities and knowledge management: impact on firm performance. *Management Decision* 2018;57:1923-36. <https://doi.org/10.1108/MD-07-2018-0825>.
3. Washington Durr AK. A Text Analysis of Data-Science Career Opportunities and US iSchool Curriculum. *Journal of Education for Library and Information Science* 2020;61:270-93. <https://doi.org/10.3138/jelis.2018-0067>.
4. Huppenkothen D, Arendt A, Hogg DW, Ram K, VanderPlas JT, Rokem A. Hack weeks as a model for data science education and collaboration. *Proceedings of the National Academy of Sciences* 2018;115:8872-7. <https://doi.org/10.1073/pnas.1717196115>.
5. Romero C, Ventura S. Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery* 2020;10:e1355. <https://doi.org/10.1002/widm.1355>.
6. Mbwilo B, Kimaro H, Justo G. Data Science Postgraduate Education at University of Dar es Salaam in Tanzania: Current Demands and Opportunities. En: Nielsen P, Kimaro HC, editores. *Information and Communication Technologies for Development. Strengthening Southern-Driven Cooperation as a Catalyst for ICT4D*, Cham: Springer International Publishing; 2019, p. 349-60. https://doi.org/10.1007/978-3-030-19115-3_29.
7. Irizarry RA. The Role of Academia in Data Science Education. *Harvard Data Science Review* 2020;2. <https://doi.org/10.1162/99608f92.dd363929>.
8. Wang L. Twinning data science with information science in schools of library and information science. *Journal of Documentation* 2018;74:1243-57. <https://doi.org/10.1108/JD-02-2018-0036>.
9. Medicine NA of S Engineering, and, Education D of B and SS and, Education B on S, Sciences D on E and P, Statistics C on A and T, Analytics B on MS and, et al. *Data Science for Undergraduates: Opportunities and Options*. National Academies Press; 2018.
10. Minterio de Educación - Secretaría de Políticas Universitarias. Buscador de títulos universitarios 2022. http://sipes.siu.edu.ar/buscar_titulos_form.php?ah=st640cad288e3fe7.89162175&ai=dngu%7C%7C3731.
11. Volkova NP, Rizun NO, Nehrey MV. Data science: opportunities to transform education. *CTE Workshop Proceedings* 2019;6:48-73. <https://doi.org/10.55056/cte.368>.
12. Daniel BK. Big Data and data science: A critical review of issues for educational research. *British Journal of Educational Technology* 2019;50:101-13. <https://doi.org/10.1111/bjet.12595>.
13. Sarker IH. Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. *SN COMPUT SCI* 2021;2:377. <https://doi.org/10.1007/s42979-021-00765-8>.
14. Yoon J, Suh M-G. Determinants of organizational performance: some implications for top executive leadership in Korean firms. *Asia Pacific Business Review* 2019;25:251-72. <https://doi.org/10.1080/13602381.2018.1548545>.
15. Leahey E, Keith B, Crockett J. Specialization and promotion in an academic discipline. *Research in*

Social Stratification and Mobility 2010;28:135-55. <https://doi.org/10.1016/j.rssm.2009.12.001>.

16. Caviggioli F, Colombelli A, De Marco A, Scellato G, Ughetto E. Co-evolution patterns of university patenting and technological specialization in European regions. *J Technol Transf* 2023;48:216-39. <https://doi.org/10.1007/s10961-021-09910-0>.

17. Lepori B. The heterogeneity of European Higher Education Institutions: a configurational approach. *Studies in Higher Education* 2022;47:1827-43. <https://doi.org/10.1080/03075079.2021.1968368>.

18. Inaba Y. Higher education in a depopulating society: Survival strategies of Japanese universities. *Research in Comparative and International Education* 2020;15:136-57. <https://doi.org/10.1177/1745499920910581>.

19. Hossain MdN, Ahmed SMZ. Use of scholarly communication and citation-based metrics as a basis for university ranking in developing country perspective. *Global Knowledge, Memory and Communication* 2020;69:461-82. <https://doi.org/10.1108/GKMC-09-2019-0112>.

20. Baños Martínez MA, Michelena Fernández E, Baños Martínez MA, Michelena Fernández E. El proceso de acreditación de carreras. *Revista Cubana de Educación Superior* 2021;40.

21. Guaglianone A. Las políticas públicas de evaluación y acreditación de las carreras de grado en Argentina. *Calidad en la educación* 2012:187-217. <https://doi.org/10.4067/S0718-45652012000100006>.

22. Bidiña A. El discurso de evaluación de la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU) en la función investigación. *Estudios del Discurso en Latinoamérica*, Buenos Aires: 2015, p. 24.

23. Bustamante LA. La evaluación de la investigación en las universidades privadas: Presente y horizonte según el discurso de CONEAU. *Education Policy Analysis Archives* 2016;24:96-96. <https://doi.org/10.14507/epaa.24.2325>.

24. Beatriz -Peretto C, Caro NP. El proceso de acreditación de una carrera de grado en Argentina. *Educación y sociedad* 2022;20:42-64.

25. Figueroa Escudero E, García Martínez V, Figueroa Escudero E, García Martínez V. Semejanzas y diferencias de la política curricular entre México y Brasil. *Revista de la educación superior* 2017;46:127-32. <https://doi.org/10.1016/j.resu.2017.08.001>.

26. Soto GYG, López RIG, Rodríguez AL. Calidad en la educación superior en línea: un análisis teórico. *Revista Educación* 2020;44:441-56. <https://doi.org/10.15517/revedu.v44i2.39714>.

FINANCING

None.

CONFLICT OF INTEREST

No conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Denis González Argote.

Original writing-drafting: Denis González Argote.

Writing-revision and editing: Denis González Argote.