Emerging Trends in Retail Analytics: A Bibliometric Analysis of the Last Decade

Tendencias Emergentes en la Analítica del Retail: Un Análisis Bibliométrico de la Última Década

Juan David Velásquez-Henao

*Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia. jdvelasq@unal.edu.co*

Received: June 16th, 2023. Received in revised form: November 10th, 2023. Accepted: November 10th, 2023.

**Abstract**

.

*Keywords*:.

**Resumen**

.

*Palabras clave*:.

# Introduction

.

.

The rest of this paper is organized as follows: Section 2 discusses the methodology used. Section 3 presents the results. Section 4 discusses the findings. Finally, Section 5 presents the conclusions.

# Literature Review

# Materials and Methods

This section outlines and examines the standard workflow commonly employed in literature analysis and tech-mining studies, which also serves as the methodological foundation of this paper. This process has been widely discussed in the research literature, including by Aria and Cuccurullo (2017), Donthu et al. (2021), and Page et al. (2021). The adopted methodology comprises four key stages:

1. Design of the study.

2. Collection and preparation of data.

3. Analysis of the data.

4. Interpretation of the findings.

## Design of the Study

The study's parameters are presented below, following the literature review criteria established by Kitchenham (2004). Scopus was used as the bibliographic database to retrieve information on engineering journals indexed in SciELO, aligning with the study's objective of analyzing SciELO-indexed engineering journals. The search strategy consisted of listing the ISSNs of the relevant journals from SciELO and using them to perform targeted searches within Scopus. The study parameters are:

* Database: Scopus.
* Years of analysis: From Jan 2015 to Dec 2024.
* Data retrieval: Feb 3, 2025.
* Search string: The operator ISSN() was used to retrieve the information of each engineering journal indexed in SciELO from Scopus.
* Inclusion criteria: None.
* Exclusion criteria: Documents whose abstracts are not written in English.

The exclusion criterion was established based on two primary considerations. First, most abstracts in the dataset are written in English, making it the dominant language of scholarly communication in the selected journals. Second, the text processing tools employed in this study are configured to analyze English-language content, ensuring consistency and accuracy in the analysis. As a result, documents with abstracts written in other languages were excluded from the dataset.

**Fig. 1** presents the PRISMA flow chart. During the identification phase, the search string returned 29,499 documents. A time restriction was applied in the screening phase, excluding 12,849 papers published before 2015 or in 2025. The eligibility phase involved reviewing the titles and abstracts of the remaining 16,603 documents published between 2015 and 2024, excluding 1,307 papers. Consequently, the final dataset consists of 15,286 documents.

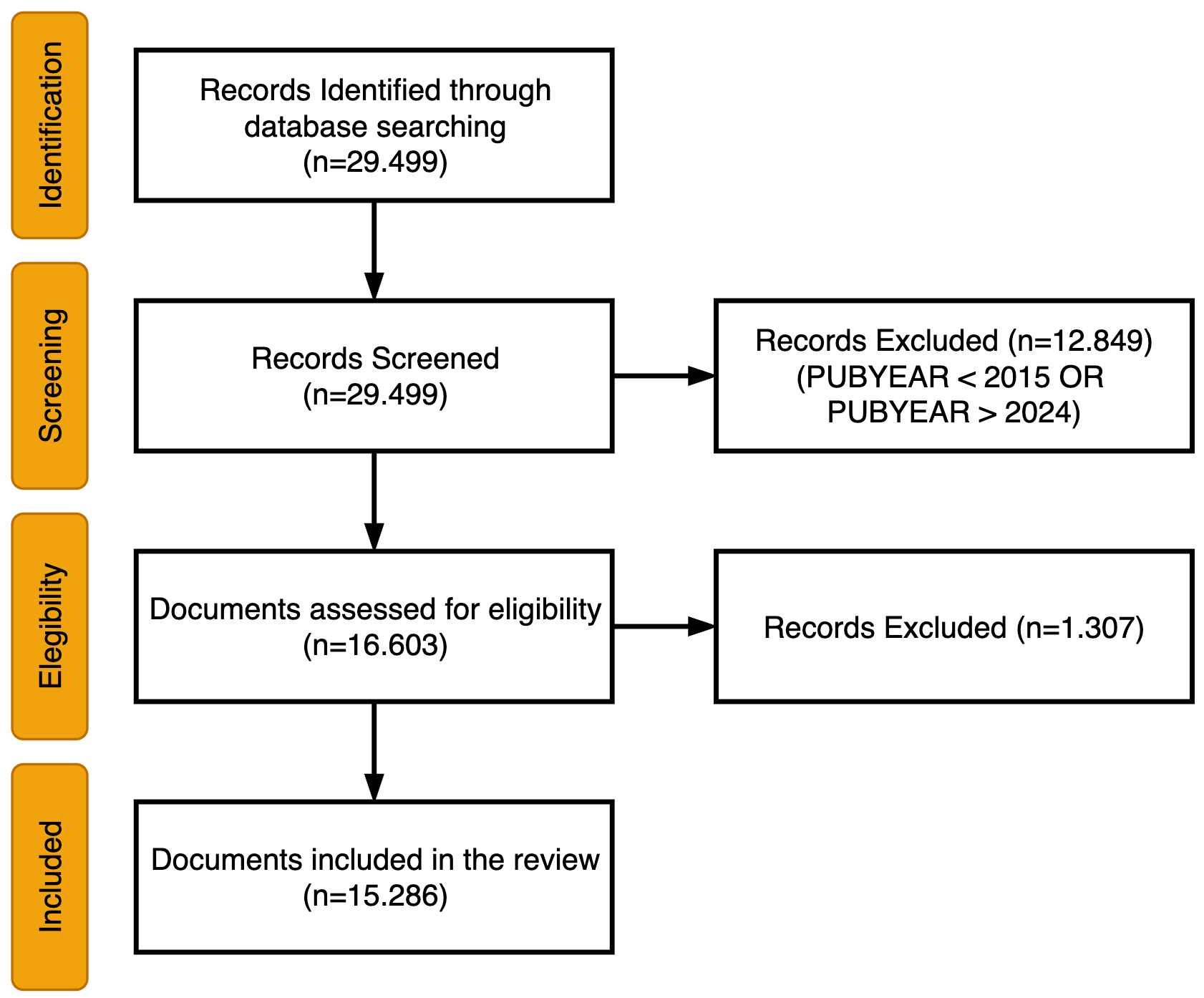


Figure 1. The PRISMA flow chart

Source: The authors.

## Collection and preparation of data

Following widely accepted practices in the literature, all bibliographic data were downloaded from Scopus in CSV format for analysis (Donthu et al., 2021; Page et al., 2021). The downloaded dataset includes document titles, abstracts, author keywords, index keywords, authors, affiliations, source titles, and bibliographic references.

Data processing combined automated procedures with manual refinements to ensure consistency and quality. The cleaning process applied to the raw author keywords included:

* Converting all text to uppercase.
* Translating British spelling to American variants.
* Removing multiple consecutive spaces within strings.
* Standardizing the formatting of hyphenated words.

This preprocessing step was essential to ensure accurate grouping and analysis of keywords in the subsequent stages. Author keywords were used directly for thematic exploration, reflecting the terminology and concepts prioritized by the authors themselves. Identifying dominant themes based on cleaned author keywords supports a multifaceted analysis of the engineering papers published in SciELO, as detailed in the following section.

## Analysis of the data

Various analyses were conducted to understand better the papers published in engineering journals indexed in SciELO. The following sections detail each analysis's specific procedures and findings.

## Interpretation of the findings

The interpretation phase focused on reviewing and synthesizing the results obtained from the various analyses. Patterns, trends, and thematic clusters identified during the data analysis phase were examined to derive meaningful conclusions aligned with the study's objectives. This step involved contextualizing the findings within the broader research landscape and ensuring consistency between the results and the study’s research questions.

# Results and Discussion

This section presents the fundamental bibliometric indicators of the analyzed dataset.

## General Metrics

The dataset includes scientific publications from January 2015 to December 2024, totaling 15,286 documents and reflecting an annual growth rate of 28.05%. The average document age is 5.77 years, with each work receiving approximately 4.11 citations overall, or 0.41 citations per year. Publications originate from 37 different sources, averaging 413.14 documents per source. The dataset comprises 14,822 journal articles, 35 conference papers, 21 editorials, two errata, one letter to the editor, and 402 review papers. A total of 40,773 authors (1,609 unique) contributed to the publications, with a strong tendency toward collaboration—averaging 3.61 authors and 3.92 co-authors per document. International collaborations account for 18.84% of the dataset. The contributing authors are affiliated with 10,687 organizations across 127 countries.

## Leading Scopus Subject Areas

This section analyzes the subject areas assigned by Scopus, which are attributed at the journal level rather than to individual articles. These classifications help identify the disciplinary focus of the research published in SciELO-indexed engineering journals. The dataset used in this study comprises 15,286 documents, which are distributed across 18 of Scopus's 27 subject areas. Notably, eight subject areas are associated with 1,000 or more documents. The most prominent areas are:

* Engineering – 7,257 documents (as anticipated).
* Materials Science – 2,664 documents.
* Computer Science – 1,884 documents.
* Chemical Engineering – 1,629 documents.
* General – 1,609 documents.
* Agricultural and Biological Sciences – 1,401 documents.
* Environmental Science – 1,263 documents.
* Business, Management and Accounting – 1,017 documents.

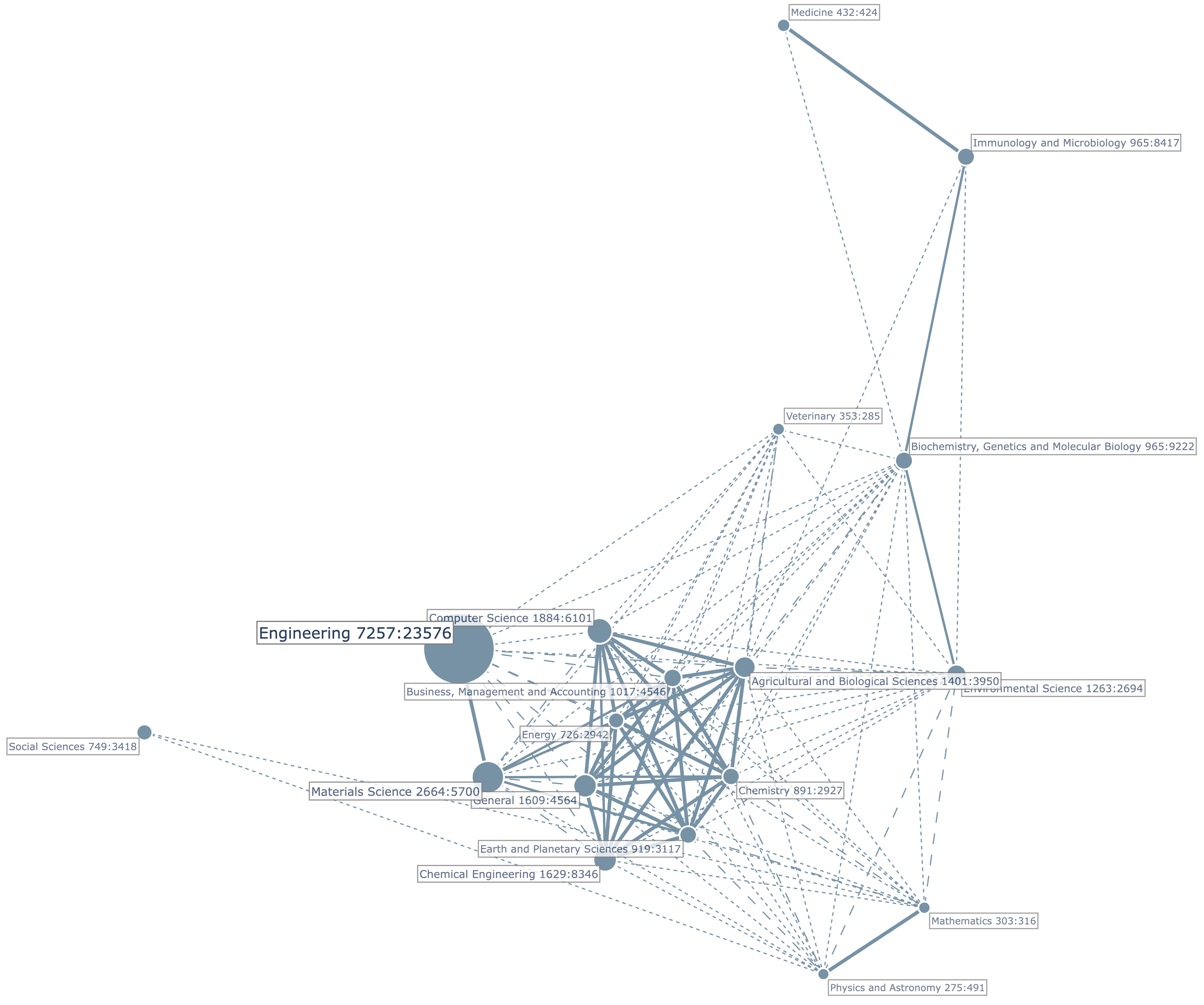


Figure 2. Correlation map of Scopus Subject Areas crossed with engineering SciELO journals.

Source: The authors.

## Basic Metrics

### Authors

### Organizations

### Countries

### Sources

## Correlation Analysis

### Countries

### Keywords

### Descriptors

## Correlation Analysis

### Citation Network

### Co-citation Network

### Coupling

## Dominant Themes

### Dominant Clusters

### Trending Terms per Year

## Publication Trend

.

## Leading Scopus Subject Areas

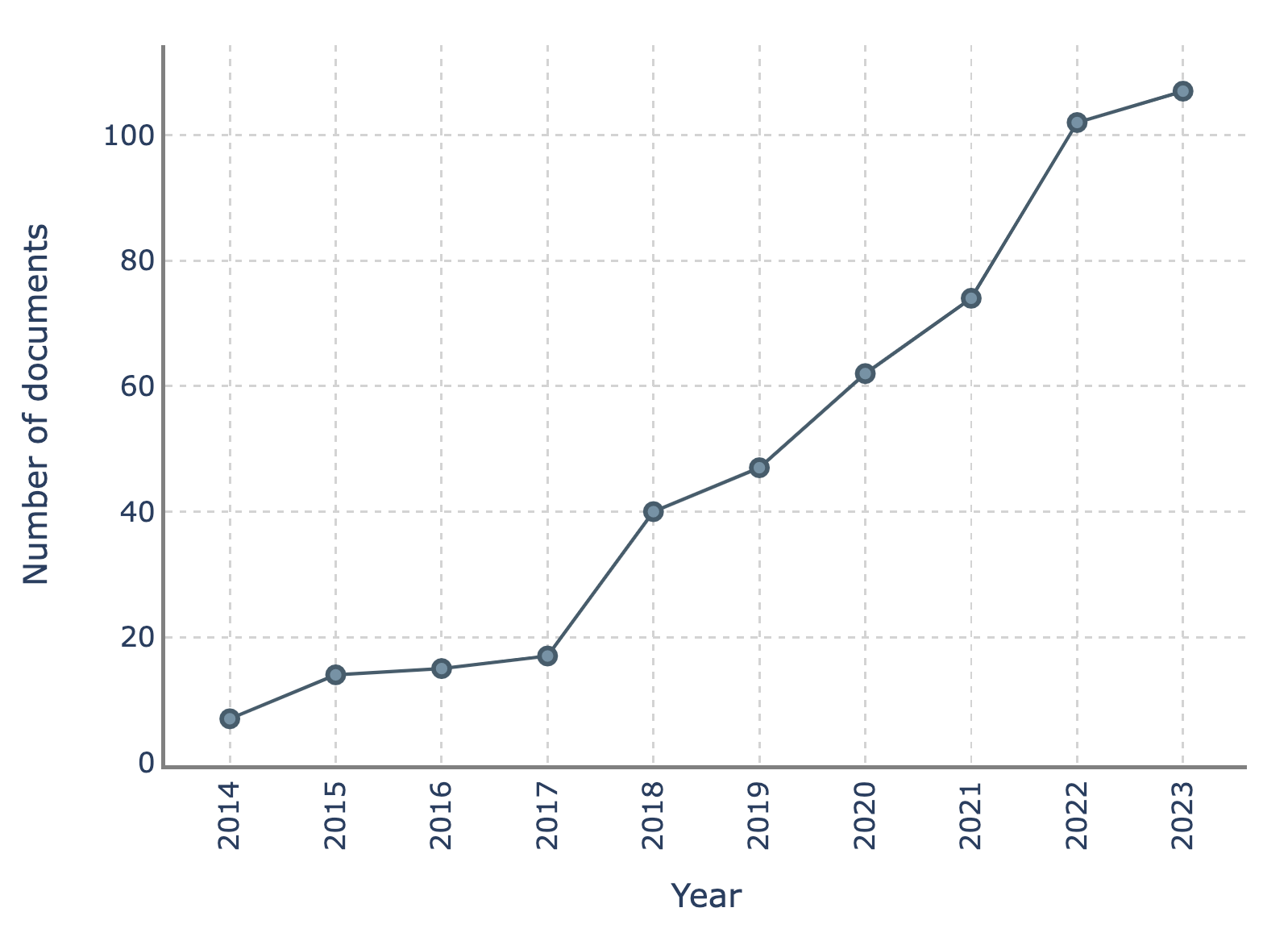


Figure 2. Number of documents published by year.

Source: The authors.

## Cited References

.

.

## Similarity among Scopus subject areas

.

## Leading Countries

.

## Leading Institutions

.

## Leading Publication Sources

.

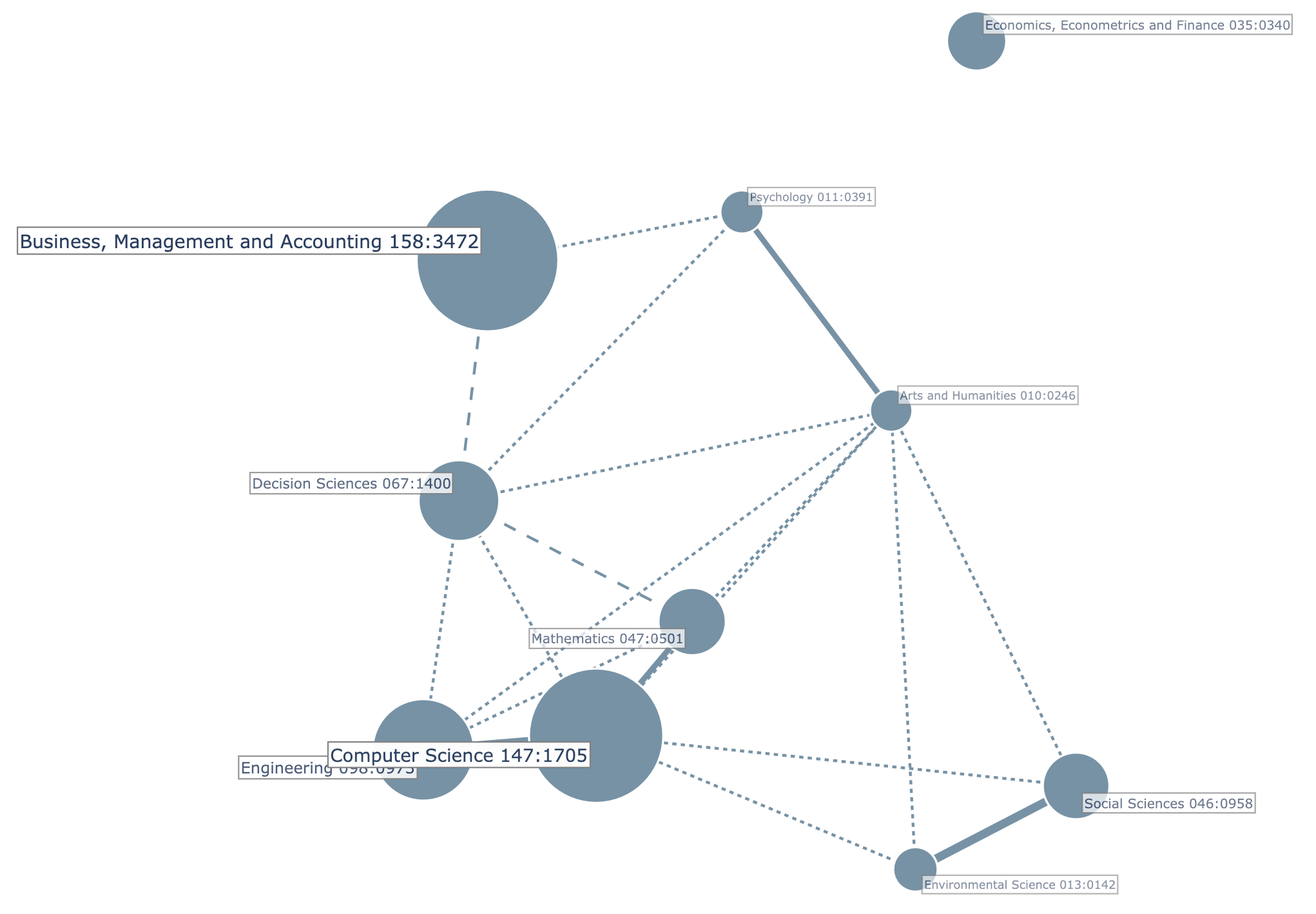


Figure 3. Correlation map of Scopus Subject Areas crossed with cited journals.

Source: The authors.

## Most Cited Documents

.

Table 1.

Most Cited Documents in Retail Analytics.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Title | Authors | Document  Type | Rank  Global  Citations | Global  Citations | Rank  Local  Citations | Local  Citations |
| Shopping intention at AI-powered automated retail stores (AIPARS) | Pillai et al. [17] | Article | 1 | 201 | 2 | 13 |
| A study on investments in the big data-driven supply chain, performance measures Stuckenschmidt [\*142\*] and organizational performance in Indian retail 4.0 context | Gawankar et al. [19] | Article | 2 | 130 | 7 | 6 |
| Shopping with a robotic companion | Bertacchini et al. [20] | Article | 3 | 127 | 4 | 9 |
| Big data analytics and demand forecasting in supply chains: a conceptual analysis | Hofmann and Rutschmann [22] | Article | 4 | 111 | 16 | 4 |
| Consumer engagement via interactive artificial intelligence and mixed reality | Sung et al. [23] | Article | 5 | 103 | 52 | 1 |
| Drivers and impact of big data analytic adoption in the retail industry: A quantitative investigation applying structural equation modeling | Lutfi et al. [24] | Article | 6 | 95 | 27 | 3 |
| Retail business analytics: Customer visit segmentation using market basket data | Griva et al. [25] | Article | 7 | 91 | 11 | 5 |
| Agent-Based Modeling of Retail Electrical Energy Markets with Demand Response | Dehghanpour et al. [26] | Article | 8 | 91 | 117 | 0 |
| Indian shopper motivation to use artificial intelligence: Generating Vroom’s expectancy theory of motivation using grounded theory approach | Chopra [21] | Article | 9 | 87 | 5 | 8 |
| Retail sales forecasting with meta-learning | Ma and Fildes [27] | Article | 10 | 86 | 12 | 5 |
| Daily retail demand forecasting using machine learning with emphasis on calendric special days | Huber and Stuckenschmidt [16] | Article | 11 | 80 | 1 | 16 |
| State-of-the-art and adoption of artificial intelligence in retailing | Weber and Schütte [18] | Article | 18 | 70 | 3 | 12 |
| Revolution of Retail Industry: From Perspective of Retail 1.0 to 4.0 | Har et al. [28] | Article | 32 | 49 | 8 | 6 |
| Artificial intelligence in retail: applications and value creation logics | Cao [29] |  | 40 | 43 | 6 | 8 |
| Low cost embedded system for increasing retail environment intelligence | Pierdicca et al. [30] |  | 49 | 37 | 9 | 6 |
| Incorporating big data within retail organizations: A case study approach | Aversa et al. [31] | Article | 58 | 31 | 10 | 6 |

Source: The authors.

.

.

Table 3.

Dimensions for analysis of emerging topics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Rank  Occurrences | Occurrences | Rank  Citations | Citations |
| JOURNALS |  |  |  |  |
| Journal of Retailing and Consumer Services | 1 | 18 | 1 | 632 |
| Lecture Notes in Networks and Systems | 2 | 14 | 56 | 25 |
| Lecture Notes in Computer Science | 3 | 10 | 16 | 79 |
| ACM International Conference Proceeding Series | 4 | 9 | 42 | 37 |
| International Journal of Retail and Distribution Management | 5 | 8 | 3 | 232 |
| Advances in Intelligent Systems and Computing | 6 | 8 | 74 | 18 |
| Sustainability (Switzerland) | 7 | 5 | 12 | 95 |
| Lecture Notes in Electrical Engineering | 8 | 5 | 92 | 12 |
| International Journal of Production Research | 9 | 4 | 2 | 246 |
| Procedia Computer Science | 10 | 4 | 13 | 93 |
| Annals of Operations Research | 19 | 3 | 9 | 116 |
| International Journal of Information Management | 29 | 2 | 4 | 182 |
| European Journal of Operational Research | 30 | 2 | 5 | 162 |
| Computers in Human Behavior | 31 | 2 | 6 | 148 |
| International Journal of Logistics Management | 32 | 2 | 7 | 131 |
| International Journal of Physical Distribution and Logistics Management | 33 | 2 | 8 | 118 |
| Expert Systems with Applications | 34 | 2 | 10 | 99 |
| COUNTRIES |  |  |  |  |
| India | 1 | 124 | 2 | 1131 |
| United States | 2 | 69 | 1 | 1190 |
| China | 3 | 56 | 5 | 429 |
| United Kingdom | 4 | 51 | 3 | 1020 |
| Germany | 5 | 35 | 4 | 497 |
| Italy | 6 | 20 | 6 | 412 |
| Canada | 7 | 20 | 8 | 188 |
| Russia | 8 | 18 | 25 | 63 |
| Brazil | 9 | 14 | 7 | 215 |
| Ireland | 10 | 13 | 9 | 188 |
| Hong Kong | 21 | 7 | 10 | 177 |
| ORGANIZATIONS |  |  |  |  |
| Maynooth Univ. (IRL) | 1 | 8 | 9 | 129 |
| Univ. of Appl. Sciences Upper Austria (AUT) | 2 | 6 | 5 | 158 |
| Amity Univ. (IND) | 3 | 6 | 36 | 76 |
| Univ. of Bristol (GBR) | 4 | 5 | 14 | 107 |
| Dublin City Univ. (IRL) | 5 | 5 | 37 | 76 |
| Massachusetts Inst. of Technol. (USA) | 6 | 5 | 58 | 55 |
| Univ. of Moratuwa (LKA) | 7 | 5 | 204 | 14 |
| Univ. of Duisburg-Essen (DEU) | 8 | 4 | 20 | 100 |
| Univ. of Tennessee (USA) | 9 | 4 | 33 | 81 |
| Univ. of Bologna (ITA) | 10 | 4 | 59 | 55 |
| Swansea Univ. (GBR) | 37 | 2 | 1 | 242 |
| Montana State Univ. (USA) | 38 | 2 | 4 | 194 |
| Univ. of Mannheim (DEU) | 39 | 2 | 6 | 156 |
| Pune Inst. of Bus. Manag. (IND) | 116 | 1 | 2 | 201 |
| Sri Balaji Univ. (IND) | 117 | 1 | 3 | 201 |
| California State Univ. (USA) | 118 | 1 | 7 | 130 |
| Nac. Inst. of Ind. Eng. (NITIE) (IND) | 119 | 1 | 8 | 130 |
| Università della Calabria (ITA) | 120 | 1 | 10 | 127 |
| AUTHORS |  |  |  |  |
| Razmochaeva N.V. | 1 | 7 | 117 | 46 |
| Bezbradica M. | 2 | 5 | 57 | 76 |
| Cirqueira D. | 3 | 5 | 58 | 76 |
| Helfert M. | 4 | 5 | 59 | 76 |
| Klionskiy D.M. | 5 | 5 | 240 | 21 |
| Griva A. | 6 | 4 | 12 | 122 |
| Frontoni E. | 7 | 4 | 17 | 103 |
| Pantano E. | 8 | 4 | 20 | 103 |
| Frazzon E.M. | 9 | 3 | 15 | 106 |
| Pereira M.M. | 10 | 3 | 16 | 106 |
| Huber J. | 33 | 2 | 4 | 156 |
| Stuckenschmidt H. | 34 | 2 | 5 | 156 |
| Dwivedi Y.K. | 137 | 1 | 1 | 201 |
| Pillai R. | 138 | 1 | 2 | 201 |
| Sivathanu B. | 139 | 1 | 3 | 201 |
| Gawankar S.A. | 140 | 1 | 6 | 130 |
| Gunasekaran A. | 141 | 1 | 7 | 130 |
| Kamble S. | 142 | 1 | 8 | 130 |
| Bertacchini F. | 143 | 1 | 9 | 127 |
| Bilotta E. | 144 | 1 | 10 | 127 |

Source: The authors.

Table 4.

Emergent topics clusters.

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster Name | Num Terms | Percentage | Main Terms |
| Retail Sales prediction | 9 | 12.5 | Prediction; Retail Sales; Sales Forecasts; Sales Data; Time Series; Retail Trade; Strategic Decisions; Sales Promotions; Sales Prediction |
| AI-Driven Customer Insights | 9 | 12.5 | Artificial Intelligence; Retail Industry; Retail Organizations; Customer Satisfaction; Artificial Intelligence Technology; Computer Vision; Experience; Information Technology; Business Performance |
| Consumer Behavior and Price Dynamics | 8 | 11.1 | Consumer Behavior; Insights; Retail Operators; Data Sets; Consumption Behaviors; Customer Engagement; Price Dynamics; Pricing |
| ML for Predictive Modeling | 7 | 9.7 | Random Forest; Decision Trees; Retail Location; Logistic Regression; Predictive Models; Support Vector Machine; Boosting |
| AI-driven Retail Performance | 7 | 9.7 | Learning Systems; Neural Network; Deep Learning; Convolutional Neural Networks; Radio Frequency Identification; Performance Metrics; Supervised Learning |
| Data-Driven Social and Consumer Dynamics | 6 | 8.3 | Social Media; Consumer; Retail Banks; Finance; Robots; Retail Data |
| Customer-Centric Data-Driven Strategies | 6 | 8.3 | Customers; Data Mining; Customer Relationship Management; Business Analytics; Electronic Commerce Websites; Mobile Devices |
| Consumer-Centric Experience | 5 | 6.9 | Customer Experience; Management; Customer Service; Customer Data; Consumer Data |
| Predictive Customer Behavior Systems | 5 | 6.9 | Customer Behavior; Information System; Recommender Systems; Transaction Data; Customer Demands |
| Predictive Customer Behavior Systems | 4 | 5.6 | Decision Making; Decision Support Systems; Decisions; Efficiency |
| Human-Centered Business Process Strategy | 3 | 4.2 | Strategy; Business Processes; Human Resource Managers |
| Fashion Analytics | 3 | 4.2 | Data Science; Data Analysis; Fashion |

Source: The authors.

.

# Conclusions

.

# References

[1]

**J.D. Velásquez-Henao** earned his BS in Civil Engineering in 1994, an MS in Systems Engineering in 1997, and a PhD in Energy Systems in 2009, all from the National University of Colombia in Medellin, Colombia. From 1994 to 1999, he worked in electricity utilities and consulting companies in the power sector. In 2000, he joined the National University of Colombia in Medellin and was appointed a full professor of computer science in 2012. Between 2004 and 2006, he served as an Associate Dean for Research, and from 2009 to 2018, he led the Computing and Decision Science Department at the Facultad de Minas, National University of Colombia, Medellin. His research and publications span simulation, modeling, optimization, and forecasting in energy markets. He specializes in nonlinear time-series analysis and forecasting using statistical and computational intelligence techniques, numerical optimization with metaheuristics, and analytics and data science. He currently instructs postgraduate courses in data science, machine learning, and big data in the Analytics program, emphasizing Python programming.

ORCID: 0000-0003-3043-3037