doi: 10.31674/ijrtbt.2021.v06i02.004

IJRTBT

A Review of the Publication Trend of Data Analytics

Robin Au1, Kelvin Leong[1]*, Anna Sung1, Ching Lee[2]

[1]University of Chester CH1 4BJ, U.K.

[2]The Hong Kong Polytechnic University 122001, Hong Kong

**Corresponding Author's Email: k.leong@chester.ac.uk*

Abstract

Data Analytics has been considered as a promising topic. This paper aims to provide a comprehensive review on the publication trends of Data Analytics. More specifically, in this study we systematically identified and analysed 18-years real-world publication data obtained from Web of Science database for the research purpose. These data include the first related publication available in the database. In all, 18610 related publications have been identified during 2004 to 2021. By analysing the publication trends from the database, we suggest that Data Analytics is an emerging global research topic that draws attention from affiliations and funding sponsors over the world. On top of the industrial voice saying Data Analytics is an emerging topic, the findings from this paper can provide a real-world reference for industrial participators, policy makers, and academia, to conduct, promote and support the Data Analytics related research. As per our best knowledge, this is the first time that a comprehensive study has been conducted to systematically review the publication trends of Data Analytics. We hope that this study can provide new insights on related research.

Keywords: Data Analytics; Publication Trend; Big Data

Introduction

Nowadays, data may be the most valuable resource in the business world. In the age of big data, data is everywhere and a very practical challenge in business is having too much data to analyze. On this, Data Analytics become a valuable solution to the challenge.

Big Data & Analytics (BD&A) has been a phenomenon as it creates new model of decision support that enable organizations to extract and store data not only from internal systems but also with external data sources such as social media platform sites, online news, blogs, web contents, data generated from interconnected devices known as the internet of things (IoT), and other external traditional and modern databases (Baharuden, Isaac & Ameen, 2019).

Data Analytics is a systematic process aims to discover interesting, meaningful, and useful knowledge. The process involves different key steps, including extracting, cleaning, transforming, modeling, and visualization of data. Recent years, Data Analytics has been considered as an important topic in business world. Data Analytics is important because it helps businesses to achieve performance optimization. Previous study Columbus (2018) indicated that with 59% of enterprises using analytics in some capacity. In terms of market size, Borasi, Khan, & Kumar (2021) predicted that the global business analytics and big data market size will reach to $684.12 billion by 2030. Leong et al. (2021) suggested Data Analytics is a key component behind the global development of FinTech (Financial Technology). In fact, Data Analytics can help a business with everything from personalizing a service for an individual customer to identifying potential opportunities in the industry.

This pioneer study reviews the publication trends of Data Analytics. As per our best knowledge, we are not aware of similar comprehensive studies had been conducted for the purpose. Therefore, this study can fill this gap by analysing real-world publication data obtained from the historical publication record of Web of Science - including the first related publication available in the database.

In brief, Data Analytics is an emerging research topic. This paper contributes a systematic review on the publication trends of Data Analytics. On top of the industrial voice demanding more research to be conducted on Data Analytics related topics, the findings provide a real-world reference for industrial participators, policy makers, and academia, to conduct, promote and support the Data Analytics related research. The policy making bodies must improve security laws and tighten our cyber security system so that the online customers can use debit card, credit card or online payments in a secured way (Ghosal, 2018). Data management and processing within an acceptable time are pre-requisites for handling big data given the vast volumes of data that are moving continuously, which often makes it difficult to search and engage with the data (Alkatheeri et al. 2020).

The remaining of this paper is organised as follows. Section two explains the research designs of this study. Section three reports and summarises the analysis results. Finally, in section four, further discussion and conclusion are provided.

Research Methodology

This study covers the period from 2004 to 2021. Given the first available publication related to Data Analytics in the Web of Science database was found since 2004, that means we analysed the entire population of all the Data Analytics related literature in the whole period.

In fact, it is worth mentioning that this study is not intended to make statistical generalisations based on the sample selected, instead, by exploring the entire data source, this study can provide analytical generalisations about the publication trends of Data Analytics. In other words, this study can provide a comprehensive portrait about the trends of Data Analytics in terms of related research publications in the field during the study period (i.e., 2004 to 2021). Following parts will provide a more detail explanations about the analysis approaches.

For the purpose to understand the publication trends, data were collected from Web of Science. Web of Science is a website service that provides subscription-based access to multiple databases, and it covers comprehensive citation data for many different academic disciplines. As a powerful database service, building on over 171 million records and almost 1.9 billon cited references, Web of Science allows users to track trends across disciplines and time.

Based on the collected data from the database, we identified and reviewed the publication trend from seven different perspectives, which are as follows:

- The number of related publications over time
- Distribution by countries
- Distribution of affiliations
- Funding Agencies
- Types of documents
- Languages
- Research Areas

The reason why we selected above seven perspectives is because this study aims to include as many as possible perspectives in order to deliver more complete and manifold views on the related publication trends. On this point, the seven perspectives were selected because relevant types of information are the most accessible in the database (i.e., Web of Science) that can be generated for the purpose. Moreover, given the selected approach being used in this study is directly repeatable, we therefore suggest the findings in this study are reproducible and transparent. These two features are important because reproducible and transparent are the two key features that should be taken into consideration in the systematic literature review on business and management related research as per Fisch & Block (2018). In practice, similar approach has been applied in other studies on other topics, such as Leong et al. (2021); Liao, Kickul & Ma (2017); Wang & Chen (2010); White & McCain (1998).

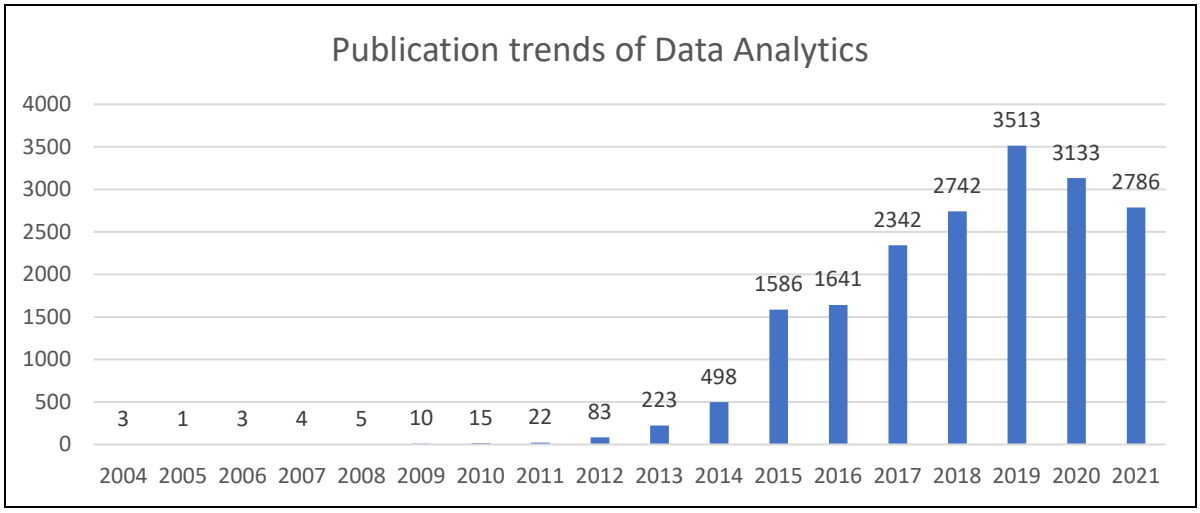
Result

This section aims to report our findings based on our publication analyses. These findings deliver a comprehensive understanding from different perspectives on the international trends of Data Analytics. Further discussions will be provided in the discussion and conclusion section.

The number of related publications over time

According to the results of searching publications containing the term "Data Analytics" in all selected fields in Web of Science database, we identified and reviewed in total 18610 related publications which includes the first publication found in 2004. Furthermore, an obvious increasing pattern from 2015 to 2019 is shown as per figure 1.

Figure 1: Publication trends of Data Analytics from 2004 to 2021



Distribution by countries

In terms of distribution by countries, the identified publications containing "Data Analytics" were contributed by researchers from 138 countries. Table 1 summarises the top 25 countries during the period. In conclusion, as per table 1, USA has the highest participation rate among all the countries. In brief, 5571 (i.e., 29.9%) publications involved scholars were from the USA.

Table 1: Distribution by countries

Rank	Countries/Regions	Total Count	% Of all identified publication
1	USA	5571	29.936
2	People's Republic of China	2547	13.686
3	India	2041	10.967
4	England	1480	7.953
5	Australia	1026	5.513
6	Germany	970	5.212
7	Canada	965	5.185
8	Italy	902	4.847
9	France	616	3.31
10	South Korea	589	3.165
11	Spain	586	3.149
12	Taiwan	506	2.719
13	Saudi Arabia	423	2.273
14	Ireland	406	2.182
15	Greece	385	2.069
16	Japan	382	2.053
17	Pakistan	372	1.999
18	Sweden	358	1.924
19	Netherlands	350	1.881
20	Singapore	341	1.832
21	Malaysia	318	1.709
22	Brazil	298	1.601
23	Switzerland	276	1.483
24	Portugal	220	1.182
25	Norway	213	1.145

Distribution of affiliations

The top 25 countries during the period are shown in table 2. The U.S. Department of Energy (DOE) involves most of the identified publications, followed by University of California System and University System of Georgia. In brief, the top five affiliations in the list are also from U.S.

Table 2: Distribution by affiliations

Rank	Affiliations	Total Count	
1	United States Department of Energy (DOE)	365	1.961

2	University of California (UC) System	345	1.854
3	University System of Georgia (USG)	244	1.311
4	State University System of Florida (SUSF or SUS)	233	1.252
5	University of Texas System (UT System)	219	1.177
6	Chinese Academy of Sciences (CAS)	213	1.145
7	International Business Machines (IBM)	204	1.096
8	Pennsylvania Commonwealth System of Higher Education (PCSHE)	196	1.053
9	University of North Carolina (UNC)	174	0.935
10	Indian Institute of Technology System (IIT System)	170	0.913
11	National Institutes of Technology (NIT) system	157	0.844
12	University of London (UoL)	145	0.779
13	Tsinghua University (THU)	144	0.774
14	University of Technology Sydney (UTS)	131	0.704
15	Georgia Institute of Technology (Georgia Tech)	127	0.682
16	The Hong Kong Polytechnic University (PolyU)	127	0.682
17	Massachusetts Institute of Technology (MIT)	126	0.677
18	Centre national de la recherche scientifique (CNRS)	124	0.666
19	University College Dublin (UCD)	123	0.661
20	University of Illinois System (UIUC)	123	0.661
21	University of New South Wales (UNSW)	123	0.661
22	Nanyang Technological University (NTU)	122	0.656
23	National Institute of Education (NIE), Singapore	122	0.656
24	California State University (CSU)	115	0.618
25	Oak Ridge National Laboratory (ORNL)	115	0.618

Funding Agencies

National Natural Science Foundation of China is the largest funder for Data Analytics in terms of identified publication, the foundation involves 5.18% of all identified publication, followed by National Science Foundation and European Commission

Table 3: Funding Agencies

Rank	Funding Agencies	Total Count	% Of all identified publication
1	National Natural Science Foundation of China (NSFC)	964	5.18
2	National Science Foundation (NSF)	936	5.03
3	European Commission	664	3.568
4	United States Department of Health Human Services	342	1.838

5	National Institutes of Health (NIH USA)	324	1.741
6	United States Department of Energy Doe	281	1.51
7	UK Research Innovation (UKRI)	261	1.402
8	Science Foundation Ireland	223	1.198
9	Natural Sciences and Engineering Research Council of Canada (NSERC)	185	0.994
10	Engineering Physical Sciences Research Council (EPSRC)	179	0.962

Overall, as per the tables 1 to 3 shown above, Data Analytics is an emerging global topic. Firstly, as per the findings, Data Analytics had become an international wide topic that involved researchers' publications from different affiliations and different countries. Secondly, in terms of funding, the topic Data Analytics had also successfully obtained sponsored from funding organisations internationally. In fact, we also found that many related publications involved collaborations between different affiliations from different countries. Therefore, we consider that above findings could provide a useful real-world reference to support future collaborative research directions.

Types of documents

Table 4 shows the top five types of document types of publication related to Data Analytics. It demonstrates that proceeding paper involved more Data Analytics related publication. Overall, 48.6% of the works were published as conference proceeding papers. On the other hand, 44.4% of the works were published as articles.

As per the figure, more works related to “Data Analytics” were published as proceeding paper during the period. This finding may reflect that “Data Analytics” was welcoming many new and innovative topics from researchers, including preliminary works. This suggestion comes from the general difference between the natures of conference proceeding article and journal paper. In a nutshell, a conference proceedings paper is published in company with a conference. By nature, conference proceeding article often refers to an earlier-term research work, such as preliminary findings, or a new idea that has emerged in course of the further research study.

Table 4: Types of documents

Rank	**Document Types**	**Total Count**	**% Of all identified publication**
1	Proceedings Papers	9053	48.646
2	Articles	8279	44.487
3	Review Articles	906	4.868
4	Early Access	495	2.66

5	Editorial Materials	379	2.037

Languages

In terms of language, as per table 5, 18 types of languages were found across all publications in which English is the main of language used in the publication. In total, 99.43% of related publications are in English.

Table 5: Languages

Rank	Languages	Total Count	% Of all identified publication
1	English	18504	99.43
2	Spanish	33	0.177
3	German	14	0.075
4	Portuguese	14	0.075
5	Turkish	11	0.059
6	Chinese	8	0.043
7	Russian	8	0.043
8	Unspecified	4	0.021
9	French	2	0.011
10	Hungarian	2	0.011
11	Italian	2	0.011
12	Korean	2	0.011
13	Bulgarian	1	0.005
14	Catalan	1	0.005
15	Serbian	1	0.005
16	Slovenian	1	0.005
17	Ukrainian	1	0.005
18	Welsh	1	0.005

Research Areas

According to table 6, “Data Analytics” had drawn much of attention from computer science and engineering related research. However, it’s worth mentioning that many “Data Analytics” related research are Business or Management related. In overall, “Data Analytics” should be considered as a topic that involve research from different areas.

Table 6: Research Areas

Rank	Research Areas	Total Count	% Of all identified publication
1	Computer Science	10746	57.743
2	Engineering	5673	30.484
3	Telecommunications	1864	10.016
4	Business Economics	1365	7.335
5	Science Technology Other Topics	822	4.417

6	Operations Research Management Science	697	3.745
7	Environmental Sciences Ecology	492	2.644
8	Information Science Library Science	465	2.499
9	Automation Control Systems	455	2.445
10	Energy Fuels	432	2.321
11	Mathematics	363	1.951
12	Education Educational Research	360	1.934
13	Transportation	342	1.838
14	Medical Informatics	325	1.746
15	Physics	314	1.687
16	Chemistry	303	1.628
17	Health Care Sciences Services	285	1.531
18	Materials Science	264	1.419
19	Instruments Instrumentation	219	1.177
20	Optics	209	1.123
21	Social Sciences Other Topics	209	1.123
22	Remote Sensing	178	0.956
23	Construction Building Technology	166	0.892
24	Public Environmental Occupational Health	165	0.887
25	Imaging Science Photographic Technology	154	0.828

Discussion

In total, 18 years of real-world publication data obtained from Web of Science database were analysed in this paper. These data include the first relevant publication found in the database since 2004.

In overall, the analysis of this research provides snapshots of Data Analytics related publication in seven perspectives.

By analysing the identified publications, we suggest that Data Analytics is a glowing international topic involving affiliations and funding organisations from different countries across the world. Moreover, the annual numbers of related publication were showing an increasing trend since the first related publication found in 2004, although the figures were showing decrease in 2020 and 2021. In fact, although United State was the key sources of related publication in terms of countries, location of affiliations, funding agencies, etc, we are still able to find related research from many other countries.

In line with many other research disciplines, English is the main of language used in the publication. Moreover, it's also worth mentioning that many "Data Analytics" related research are Business or Management related. In addition, we found that proceeding paper was the major source type of publication for â€œData Analyticsâ€. This finding may indicate that Data Analytics was welcoming many new and innovative topics from researchers, including preliminary works.

Conclusion

Technology is fast changing how businesses operate and the development of Data Analytics. On top of the industrial voice saying Data Analytics is an emerging topic, the findings from this paper provide an additional reference for industrial participators, policy makers, and academia to conduct, promote and support the Data Analytics related research.

In overall, we conclude this study is the first time that specific research has been conducted to systematically review the development trends of Data Analytics. We hope that this study can provide new insights on this emerging research topic.

Conflict of Interests

The authors declare that they have no conflict of interests.

Acknowledgement

The authors are thankful to the institutional authority for completion of the work.

References

Alkatheeri, Y., Ameen, A., Isaac, O., Nusari, M., Duraisamy, B., & Khalifa, G. S. (2020). The effect of big data on the quality of decision-making in Abu Dhabi Government organisations. In *Data management, analytics and innovation* (pp. 231-248). Springer, Singapore. https://doi.org/10.1007/978-981-13-9364-8_18

Baharuden, A. F., Isaac, O., & Ameen, A. (2019). Factors influencing big data & analytics (BD&A) learning intentions with transformational leadership as moderator variable: Malaysian SME perspective. *International Journal of Management and Human Science (IJMHS), 3*(1), 10-20.

Borasi, P., Khan, S., & Kumar, V. (2021, Sep). *Big Data and Business Analytics Market 2022*, Allied Market Research. https://www.alliedmarketresearch.com/big-data-and-business-analytics-market

Columbus, L. (2018, Aug 8). *Global State of Enterprise Analytics*. Forbes. https://www.forbes.com/sites/louiscolumbus/2018/08/08/global-state-of-enterprise-analytics-2018/?sh=3472147a6361

Fisch, C., & Block, J. (2018). Six tips for your (systematic) literature review in business and management research. *Management Review Quarterly, 68*(2), 103-106. https://doi.org/10.1007/s11301-018-0142-x

Ghosal, I. (2018). Consumer Buying Behavior on E-Marketing and its Operations: A case study on Amazon, India. *International Journal on Recent Trends in Business and Tourism, 2*(3), 1-7.

Leong, K., Sung, A., Au, D., & Blanchard, C. (2021). A review of the trend of microlearning. *Journal of Work-Applied Management, 13*(1), 88–102. https://doi.org/10.1108/JWAM-10-2020-0044

Liao, J. Kickul, J. & Ma, H. (2017). Organizational Dynamic Capability and Innovation: An Empirical Examination of Internet Firms. *Journal of Small Business Management, 47*(3), 263-286. https://doi.org/10.1111/j.1540-627X.2009.00271.x

Wang, C.C. & Chen, C.C. (2010). Electronic commerce research in latest decade: a literature review. *International Journal of Electronic Commerce Studies, 1*(1), 1-14. http://dx.doi.org/10.7903/ijecs.898

White, H.D. & McCain, K.W. (1998). Visualizing a discipline: an author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science, 49*(4), 327-355.