Research Paper

Contribution of India to Research on Big Data

Neeraj[1], Prof. Nirmal Kumar Swain[2]
[1]Research Scholar , Department of Library and Information Science,
Maharshi Dayanand University, Rohtak, Haryana, India
[2]Professor, Department of Library and Information Science,
Maharshi Dayanand University, Rohtak, Haryana, India

Abstract

Information has become an essential asset for the growth of organisations in every sector of our society. With the rate at which data has increased, organisations must upgrade their technological architecture for the improvement of their products and services as the traditional data management and processing systems are not capable of handling such a great amount of stream data. Retrieval of value from this data boosts the creation and adoption of advanced technologies. The article focuses on the growth of literature of Indian origin on big data over a period of ten years from 2011 to 2021 and scientometric analysis of data retrieved, top publication sources, different categories of documents published in this domain, prominent subject areas in the big data research and top institutes that contributed most to the big data research during this period.

Keywords: Big data, big data analytics, big data research, scientometric analysis, India.

1. Introduction

In today's era of information, when advances in modern technologies have provided smart devices and equipment within reach of every ordinary being, the internet has become a platform where billions of people communicate and share data. Information has become a quintessential asset to organisations in every sector for efficiently providing effective services. These organisations collect, store and analyse the data of their customers, products, services, transactions, email communications, messaging services, collaboration tools and various other types of data. So, organisations must deal with these floods of data generated within the organisations and by their customers.

The worth of data does not lie in how much data an organisation has at a point in time but rather in how that gathered data is utilised by that organisation. Technological researchers are constantly improving the methods to utilise this enormous amount of available data in order to improve services. Valuable insights can be obtained by analysing this data using big data analytics, which helps organisations in decision-making and benefits them in time and cost reductions, optimisation of product development, failure analysis, and re-calculating risk portfolios and thriving in business communities.

1.1 The Concept of Big Data

According to (Ed Dumbill, 2013) in (Anna & Mannan, 2020), "Big data is the data that cannot be handled by the traditional database and processing systems. Such type of data does not fit the traditional architecture of the database systems and moves at a very high speed. One must choose advanced ways to process the data and extract the value from this data".

Gartner defined big data as: "Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation."

(De Mauro et al., 2016) proposed a definition based on various aspects such as volume, velocity, variety of information, Technology and Analytical methods to derive such information and its value to society. They defined big data as: "the Information asset characterised by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value."

Whereas, (Ahmed & Ameen, 2017) defined big data that: "it is trustable and authentic data of multiple formats generated and delivered from different geospatial locations and large enough not to be processed by particular software, a spreadsheet or a computer and which may create value for organisations." (Zhan & Widén, 2019) reviewed the articles and analysed 35 definitions from different authors. On the bases of their study, they proposed a definition of big data: "Big Data refers to data with large size, fast growing speed, and various types which can complicate data handling techniques but also boost the creation of technological solutions. Value is generated by the proper operation and use of big data."

Big data is not just only the large amount of data. Rather, it can be defined as the huge volume of data that is created in various forms and at a very high speed. The data constantly changes its true nature behaving differently for different context. Storage and processing of such data cannot be handled by traditional systems. Retrieval of value from this data boost the creation and adoption of advanced technologies.

2. Statement of the Problem

The present research article is entitled "Contribution of India to Research on Big Data". With the recent technological developments, the volume of data is continuously growing. This data is available in various forms and is too large and complex to be handled by the traditional processing systems. Since the origin of the word big data, to elucidate the issues related to big

Research Paper

data such as creation, capturing, storage, and analysis, authors from different backgrounds and origin penned articles, books, reports, dissertations, from all over the world. The article focused on the growth of literature on big data in Indian origin over a period of ten years from 2011 to 2021. After the literature survey it was found that numerous studies have been done on "big data" in India during the said period with Scopus Database.

3. Objectives

The objectives of the study are:

1. To examine the year wise growth in Big Data publications;
2. To find out growth rate of Big Data publications using CAGR, AGR, RGR and DT;
3. To find out top ten publication sources (journals/conference proceedings);
4. To find out number of publications on Big Data categorised by different document types;
5. To find the most prominent subject areas in the field of Big Data research;
6. To identify the top ten institutions contributing to Big Data research;

4. Methodology

Big data is relatively a new field while talking about Library and Information Science and it has been observed that this field is not studied in depth previously by researchers with respect to the field of librarianship. The study was conducted to understand the concept of big data and explore its use, implementation and scope in the field of library and information science. Author used secondary research method to explore India's contribution towards the research in the field of Big data and its applications in Library and Information Science, contribution in any form such as journal Articles, conference proceedings, books, book chapters, reviews and many more. Data of the published articles from 2011 to 2021 have been obtained through Scopus database using the search term big data in title, keyword and/or abstract. The search was further refined to country of origin to India. The retrieved data was analysed and interpreted using MS-Excel.

5. Findings

India has contributed significantly to the research and development under the Big Data keyword and its related branches such as Machine Learning, Internet of Things, Data Mining, Cloud Computing, Artificial Intelligence etc. Over the last ten years, i.e. from 2011 to 2021, there has been enormous growth in India in journal articles, conference papers, books, book chapters, review articles and other research publications. Author searched the Scopus Database by using the keyword "Big Data" and limiting the search criteria to the Publication Year range of "2011 to 2021", where selecting "India" as the country of origin of the document. Scopus Database retrieved 47,834 such primary documents based on the search criteria. And further analysed the search results using different metrics showing the following results.

5.1 Documents by Publications Per Year

The following table shows the growth in the number of publications over these years, from only 354 publications in 2011 to 13718 publications in 2021, showing an exponential growth during the period.

Table 1: Year Wise Publication Distribution of Documents from 2011 to 2021 (Source: Scopus)

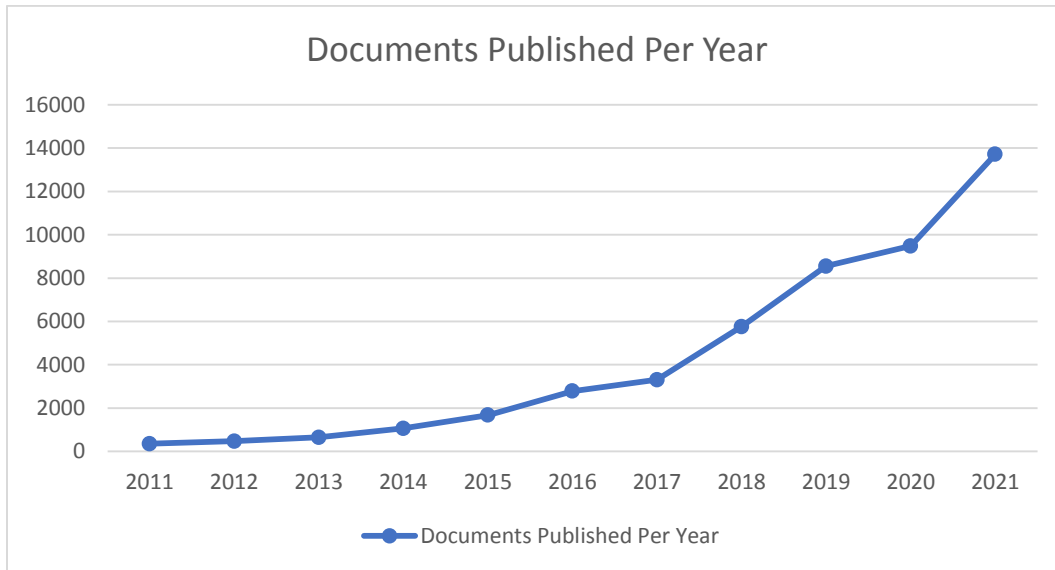| Year	Documents Published
2021	13718
2020	9485
2019	8551
2018	5764
2017	3308
2016	2783
2015	1679
2014	1064
2013	654
2012	474
2011	354

Research Paper

Figure 1: Year Wise Publication Distribution of Documents from 2011 to 2021

The chart shows that the concept of big data in India started to grow after 2010, showing exponential growth in big data in the years following.

5.1.1 Annual Growth Rate

Author used the formula for Annual Growth Rate (AGR) given by (Santha kumar & Kaliyaperumal, 2015). Table 2 shows the calculation of annual big data publications in journals in India from 2011 to 2021.

The formula used for the calculation of AGR is as follows:

$$AGR = \frac{w_2 - w_1}{w_1} \times 100$$

Where W2 represents the total number of publications in the current year, and W1 represents the initial number of publications at the start of a year, i.e., the total number of publications in the previous year.

5.1.2 Compound Annual Growth Rate (CAGR)

The Compound Annual Growth Rate represents a constant growth rate over a period of time through a geometric progression ratio. Generally, CAGR is represented in percentage by calculating the nth root of the total percentage growth rate with 'n' be the number of years in the corresponding period. It can be represented through the following mathematical equation given by (Santha kumar & Kaliyaperumal, 2015):

$$CAGR = \left(\frac{Ending\ Value}{Beginning\ Value}\right)^{\frac{1}{Ending\ Time - Beginning\ Time}} - 1$$

Where, Beginning Value and Ending Value correspond to the value at the beginning of the period and value at the end of the period respectively.

5.1.3 Relative Growth Rate and Doubling Time

The Relative Growth Rate shows the increase in number of articles per unit of time as a proportion of its size at that moment in time. It is also called as Exponential Growth Rate. The mean relative growth rate over a specific period of time interval can be expressed by the following equation given by (Kumar, 2017) (Kumar & Kaliyaperumal, 2015):

$$1 - 2R = \frac{Log\ W_2 - Log\ W_1}{T_2 - T_1}$$

Here,

1-2R = is the mean relative growth rate for the specific time interval;

$Log_e\ W_1$= Natural Log of initial number of articles;

$Log_e\ W_2$= Natural Log of final number of articles after a specific period of time;

T_2-T_1 = Difference between Final Time and Initial Time.

Doubling Time (DT) represents the time in which research publications on big data are doubled and is calculated by the following formula (Bharati et al., 2020) (Narendra & Kannappanavar, 2020).

$$Doubling\ Time\ (DT) = \frac{0.693}{R}$$

Here 0.693 is a constant, and R is the relative growth rate of the corresponding year.

Research Paper

Table 2: Growth of Big Data Publication in India from 2011 to 2021 in terms of CAGR, AGR and RGR and DT

Year	Initial No. of Publications (W1)	New Publications in Year (W2)	Cumulative Sum	Compound Annual Growth Rate (CAGR)	Annual Growth Rate (AGR)	Relative Growth Rate (RGR)	Doubling Time (DT)
2011	-	354	354	Not Defined	Not defined	Not Defined	Not Defined
2012	354	474	828	33.90	33.90	0.85	0.82
2013	474	654	1482	35.92	37.97	0.72	0.97
2014	654	1064	2546	44.32	62.69	0.66	1.05
2015	1064	1679	4225	47.57	57.80	0.62	1.12
2016	1679	2783	7008	51.04	65.75	0.60	1.16
2017	2783	3308	10316	45.13	18.86	0.56	1.23
2018	3308	5764	16080	48.97	74.24	0.55	1.27
2019	5764	8551	24631	48.89	48.35	0.53	1.31
2020	8551	9485	34116	44.10	10.92	0.51	1.37
2021	9485	13718	47834	44.15	44.63	0.49	1.41
	Total	**47834**					
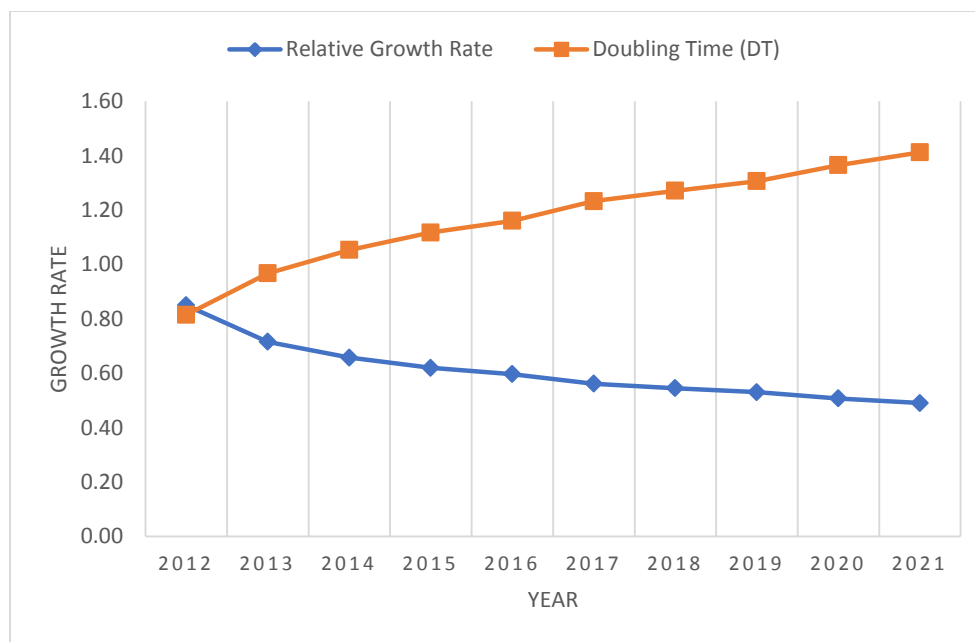
The data in table 2 corresponds to the total publication data from a period of 2011 to 2021 retrieved form Scopus database using keywords "big data" and its related terms having "India" as a place of origin analysed on the basis of metrics to measure growth such as CAGR, AGR, and RG and DT. The growth rate shows a constant increase in the number of publications per year. The highest percentage increase was shown around 74.24 percent from year 2017-2018. Where, the lowest percent increase 10.92 percent between 2019-2020. During rest of the period the percent increase was near to 50 percent per year. The RGR was 0.85 in 2012 and decreased to 0.56 in 2017. There was a small decrease to 0.55 in 2018 and after that it decreased to 0.49 in 2021. In the table 2, the DT shows a gradual increase from 0.82 in 2012 to 1.41 in 2021.

Research Paper

5.2 Documents by Top Publication Sources

Based on the retrieved search results top 10 sources (journals/conference proceedings), having contributed more than 300 publications towards the big data research in India, are identified as follows.

Table 3: Average No. of Documents Published Per Year by Top 10 Sources

Source	Avg. No. of Documents per Year
Advances In Intelligent Systems and Computing	1553
Lecture Notes in Networks and Systems	689
International Journal of Recent Technology and Engineering	608
International Journal of Innovative Technology and Exploring Engineering	575
Journal Of Advanced Research in Dynamical and Control Systems	544
Communications in Computer and Information Science	508
Lecture Notes in Electrical Engineering	501
Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics	447
Lecture Notes on Data Engineering and Communications Technologies	405
International Journal of Engineering and Advanced Technology	403

The following figure shows the number of documents published by these top sources per year from year 2011 to 2021.
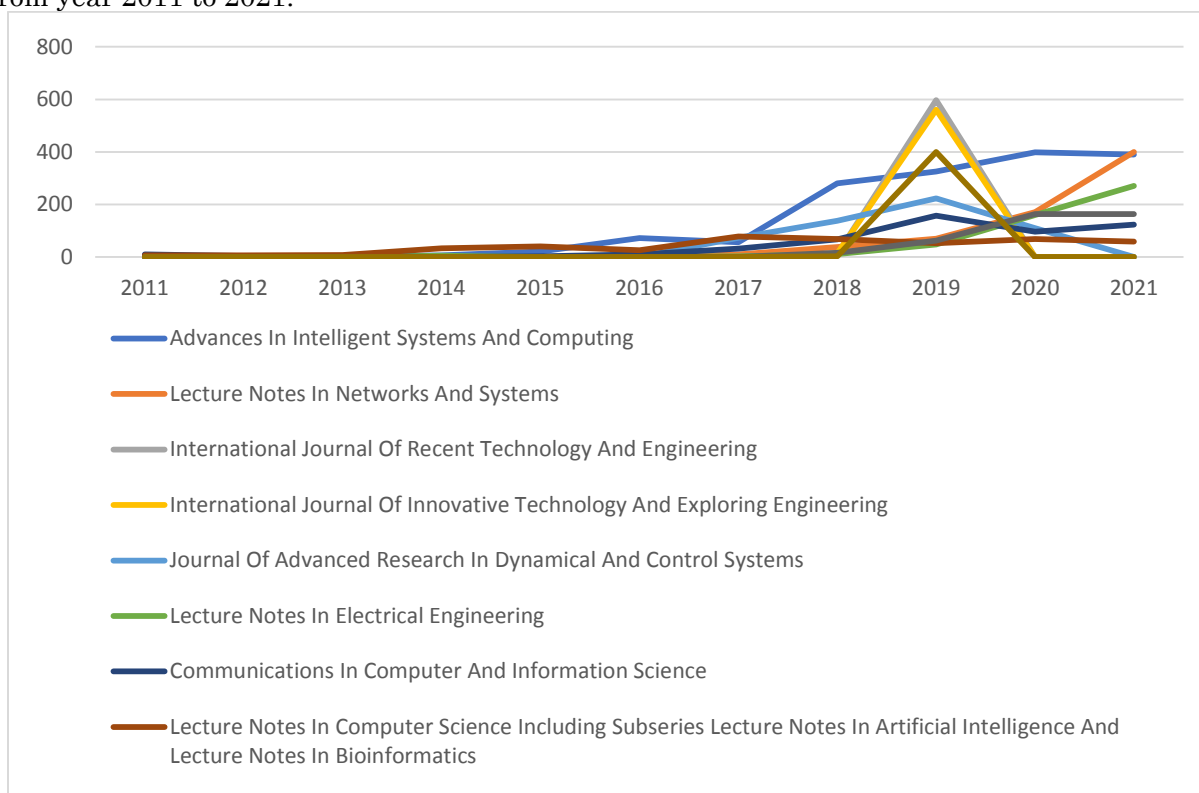

Figure 2: No. of Documents published by Top Sources Per Year from 2011 to 2021

5.3 Documents by Type of Publication

Our search retrieved various types of the documents such as articles, conference papers, book chapters, review articles, books, editorials, notes, letters, short surveys, data papers, erratum, retracted and some under undefined category of documents.

Research Paper

Table 4: Distribution of Documents Published by the Type of Publication

ocument type	Number of Documents Published
Article	24204
Conference Paper	16991
Book Chapter	3415
Review	2476
Book	312
Editorial	205
Note	67
Retracted	55
Letter	52
Short Survey	25
Data Paper	13
Erratum	11
Undefined	8

The distribution of documents published in different forms can be seen in the following pie chart.
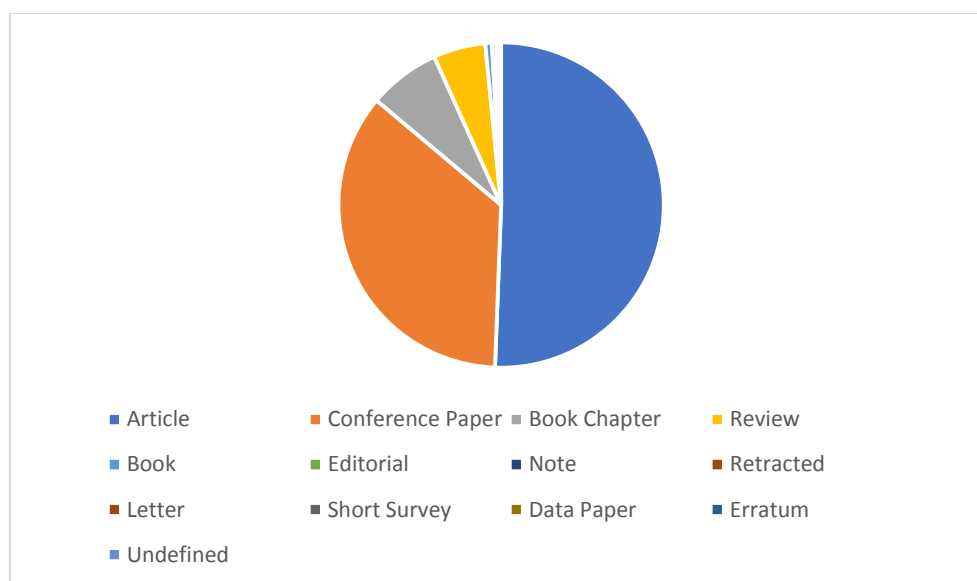

Figure 3: Type-wise Distribution of Documents

Out of these types, Articles top the list with 24204 documents published as journal articles covering 51 percent of the total documents published in all these years. Where 16991 documents were published as Conference papers having 36 percent of all documents published. These two categories of documents cover almost 87 percent of the total documents published on Big Data in India over the selected range of years.

### 5.4 Documents by Subject Area

The documents with the keyword "big data" are published in different subject areas. The following table shows the different subject areas and the number of documents published in these areas over the years of study.

Table 5: Number of Documents published in Different Subject Areas from 2011 to 2021

Subject Area	No. of Documents Published
Computer Science	28919
Engineering	19007
Mathematics	6625
Decision Sciences	5623
Business, Management and Accounting	4727
Medicine	4667
Physics and Astronomy	4020
Social Sciences	3900
Energy	2840
Environmental Science	2784
Biochemistry, Genetics and Molecular	2746

Research Paper

Biology	
Agricultural and Biological Sciences	1757
Materials Science	1752
Earth and Planetary Sciences	1334
Economics, Econometrics and Finance	1289
Chemical Engineering	1082
Pharmacology, Toxicology and Pharmaceutics	920
Chemistry	859
Multidisciplinary	724
Immunology and Microbiology	596
Arts and Humanities	454
Neuroscience	369
Psychology	347
Health Professions	236
Nursing	106
Veterinary	72
Dentistry	33

Figure 4: Subject Wise Distribution of Documents
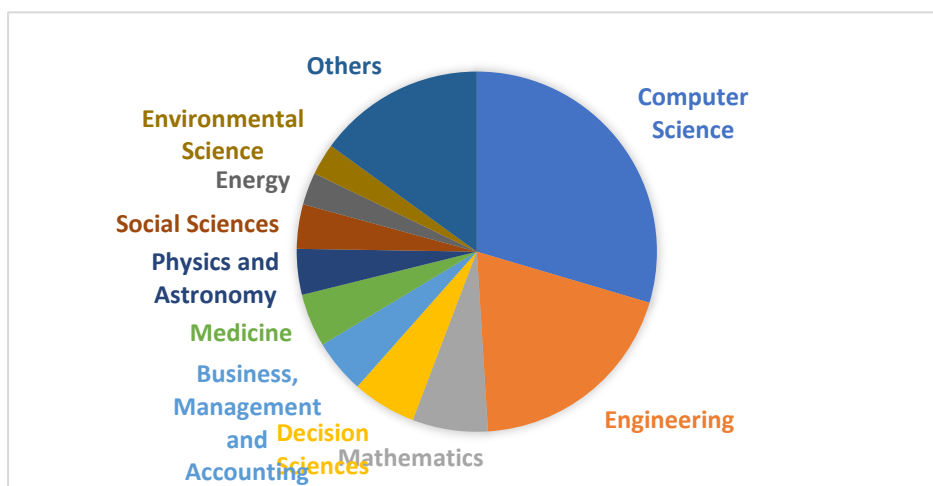
The subject wise distribution of the published documents is shown in the pie chart. The subject wise distribution shows that Computer Science contributed the largest share towards the publication of documents under big data and its related branches with overall 30 percent contribution out of all published documents. After that Engineering field contributed 19 percent publications towards the field. These two subject areas only contributed a total of 49 percent publications towards the evolution of big data concept in India.

5.5 Documents by Affiliation

Indian universities and education institutions play a vital role in introducing and advancing new concepts and technologies with their cutting-edge research contribution in every field. Every now and then, Indian educational institutions have contributed avant-garde in research and advancement of subjects in the country.

Table 6: Number of Documents published by different Universities and Institutions from 2011 to 2021

| Affiliation	No. of Documents
Vellore Institute of Technology	1551
Amity University	1144
Anna University	700
K L Deemed to be University	693
SRM Institute of Science and Technology	616
Thapar Institute of Engineering and Technology	576
Indian Institute of Technology Delhi	546
Indian Institute of Technology Kharagpur	541
University of Delhi	506
Lovely Professional University	492

Research Paper
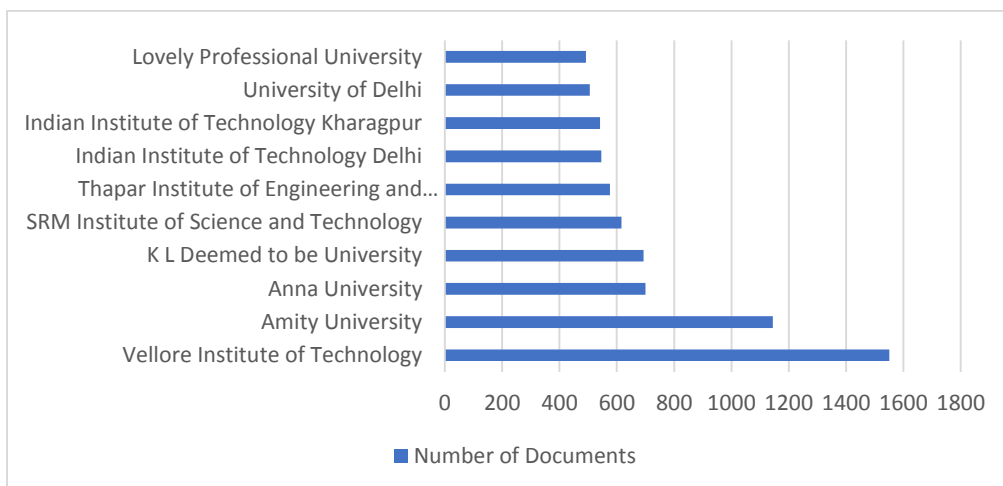
Figure 5: Documents Published by Different Universities and Educational Institutions

6. Debate and Discussion

The data as retrieved from Scopus Database shows a continuous growth in journal articles published with theme of big data. In 2011 only 354 articles were published by different authors whereas in 2021 13718 articles were published showing approximately 3800 percent increase in the total number of publications. Academic journals and conference proceedings have a great relevance to research. They ensure in-depth knowledge over the topic and contribute to the world of knowledge. Top ten sources of big data research were found having more than 300 average number of articles published per year on the concept. Out of these "Advances In Intelligent Systems and Computing" emerged to be at the top with an average of 1553 publications per year, followed by "Lecture Notes in Networks and Systems" and "International Journal of Recent Technology and Engineering" and so on. Journal articles contribute most to the research over the topic. Out of 47834 documents such retrieved, with 24204 documents, journal articles contributed approximately 51 percent of the total documents published during the period, followed with conference papers having 16991 documents covering 36 percent contribution towards the growth of topic.

Over the period, big data has influenced every sector of our society. Data as retrieved showed this influence in different subject areas. With approximately 50 percent of the total publications, Computer science and Engineering are the most prominent subject areas that have contributed to the research in big data in India followed by Mathematics with 7 percent and Decision science with 6% overall contribution. Although big data was first introduced to gain insights in business and decision making, but Business, Management and Accounting sector contritubted only percent of overall publications to the research in big data. Over this period, the growth in big data and its related technologocial innovations are a result of continuous research contribution in all subject areas. Academic institutions play a vital role in research and development of any new area. Indian academic and research institutions have contributed at large in the field of big data research. Institutions such as Vellore Institute of Technology, Amity University, Anna University, K L Deemed to be University, SRM Institute of Science and Technology, Thapar Institute of Engineering and Technology, Indian Institute of Technology Delhi, Indian Institute of Technology Kharagpur, University of Delhi and Lovely Professional University are the top 10 institutions that contributed towards the research of big data during the said period.

7. Conclusion and Future Research

It is quite surprising fact that today we produce more data in just 2 years than the overall data produced in the decades of history. The term Big Data not only describes the data in large volume, coming in fast and in various forms, but it is also the technology and practice. From every self-driving car to cashier-less stores, the data scientists have worked for and implemented the vision of sophisticated big data operations. There has been a significant improvement in the concept, technology and its usage strategies for different industries, since big data first entered in the picture. Libraries are also prevalently using the technologies such as Clouds, Internet of Things (IoT) devices, and edge computing etc., to understand their users and to improve their potential.

The term "Big Data" is a multifaceted term that is rapidly evolving and changing the ways in which our society lives, thinks and works (Garoufallou & Gaitanou, 2021). Big data research in Library and Information Science is still in the phase of its infancy. At one hand, in the fields such as Computer Science, Engineering, Business and Management, big data research has done wonders. But on the other hand it is near to negligible in the field of library and information science. Some of the recent Big Data trends and technologies that can be beneficial for the libraries and can be considered for research and exploration in the field of library and

Research Paper

information science are, Data as Service, artificial intelligence and machine learning enabled smart technologies, Natural Language Processing, Hybrid Clouds, Edge and Quantum Computing, DataOps, Predictive analytics, Data Fabrics, X Analytics, and Augmented Data Management.

The literature indicated huge opportunities for libraries having big data in the picture. Library and information professionals have always played a vital role in the growth and development of our society. And now with the emergence of new technologies in libraries, thye have some new roles to play to facilitate creation and distribution of knowledge in the community. But the one thing we can be sure about is that the real value of data and the ways in which we can utilize the information, is still unexplored in the field of library and information science. As the keepers of information, it is the moral obligation of us, library professional, to keep up with the rapid technological changes. Information and training should be imparted among library professionals to meet the challenges in use and implementation of such technologies in libraries for the benefit of our society.

References

1. Ahmed, W., & Ameen, K. (2017). Defining big data and measuring its associated trends in the field of information and library management. *Library Hi Tech News*, *34*(9), 21–24. https://doi.org/10.1108/LHTN-05-2017-0035
2. Anna, N. E. V., & Mannan, E. F. (2020). Big data adoption in academic libraries: a literature review. *Library Hi Tech News*, *37*(4), 1–5. https://doi.org/10.1108/LHTN-11-2019-0079
3. Bharati, V. K., Singh, M. P., Gautam, S. K., & Pandey, S. K. (2020). Indian Contribution to Big Data : A Scientometric Study Indian Contribution to Big Data : A Scientometric Study. *8th International Library and Information Professional Summit -2019*, *January*, 4–11. https://www.researchgate.net/publication/338423326_Indian_Contribution_to_Big_Data_A_Scientometric_Study
4. De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*, *65*(3), 122–135. https://doi.org/10.1108/LR-06-2015-0061
5. Garoufallou, E., & Gaitanou, P. (2021). Big data: Opportunities and challenges in libraries, a systematic literature review. *College and Research Libraries*, *82*(3), 410–435. https://doi.org/10.5860/crl.82.3.410
6. Kumar, R. S. (2017). Big Data Literature Output of Web of Science : a Scientometric Profile. *KIIT Journal of Library and Information Management*, *4*(2), 93–99. http://www.publishingindia.com/
7. Kumar, R. S., & Kaliyaperumal, K. (2015). Scientometric analysis of global publication output in mobile technology. *DESIDOC Journal of Library and Information Technology*, *35*(4), 287–292. https://doi.org/10.14429/djlit.35.4.7884
8. Narendra, A., & Kannappanavar, B. U. (2020). Relative Growth Rate and Doubling Time of Publications of IASLIC Bulletin: A Scientometrics Study from 2004 to 2018. *Journal of Xi'an University of Architecture & Technology*, *12*(2), 1321–1331.
9. Santha kumar, R., & Kaliyaperumal, K. (2015). A scientometric analysis of mobile technology publications. *Scientometrics*, *105*(2), 921–939. https://doi.org/10.1007/s11192-015-1710-7
10. Zhan, M., & Widén, G. (2019). Understanding big data in librarianship. *Journal of Librarianship and Information Science*, *51*(2), 561–576.
11. https://doi.org/10.1177/0961000617742451