

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

---

Winter 12-8-2019

## Research Performance in the field of big data in India: A Scientometric analysis based on Scopus database

Dr.Suchetan Kumar  
suchetan.sah@gmail.com

md safiqur Rahaman  
safirahaman2007@gmail.com

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>

 Part of the [Library and Information Science Commons](#)

---

Kumar, Dr.Suchetan and Rahaman, md safiqur, "Research Performance in the field of big data in India: A Scientometric analysis based on Scopus database" (2019). *Library Philosophy and Practice (e-journal)*. 3736.

<https://digitalcommons.unl.edu/libphilprac/3736>

## **Research Performance in the field of big data in India: A Scientometric analysis based on Scopus database**

**Dr. Suchetan Kumar**

Assistant Librarian & Course Coordinator of Department of library and Information Science,  
DSB Campus, Kumaun University, Nainital, 263001

**Md. Safiqur Rahaman**

Research Scholar, Department of Library and Information science, DSB Campus, Kumaun  
University, Nainital, 263001

E-mail: [suchetan.sah@gmail.com](mailto:suchetan.sah@gmail.com); [safirahaman2007@gmail.com](mailto:safirahaman2007@gmail.com)

**Abstract:** The present study has been carried out to find out the research performance in the field of big data in India during 2010-2019 as indexed in the Scopus database. A total 7502 research publication has been found in the Scopus during the 10 years period of 2010-2019 and were downloaded for analyze with the help of Microsoft excel and vosviewer. The study also provide comparative performance of different Scientometric parameters including total research productivity, yearly research output, authorship pattern, International collaboration, Institution wise collaboration, Degree of collaboration, RGR & DT, Source wise research publication, Subject wise publication, and most prolific authors etc. The study revealed that publication in the field of big data has increased exponentially during the study periods of 2010-2019. Most of the researcher published in conference papers i. e. 53.57% (2019). It is also observed that most preferred subject of the researcher is “Computer Science”. The relative growth rate decreased from 2010 (0.73) to 2019 (0.25) While the doubling time of the publication increased from 0.95 in 2010 to 2.72 in 2019. The Advances in Intelligent Systems and Computing has 422 (5.62%) publications and is the more popular source of publications on big data research publication in India and the degree of collaboration is 0.93. It has been considered that the scientific research production in the field of big data is increasing exponentially due to the collaboration among the researcher from different subjects.

**Keywords:** Big data, Scientometric, Scopus, Degree of collaboration, authorship pattern, Relative growth rate & doubling time.

### **Introduction:**

The term “**Big data**” etymologically can be traced during mid-1990, used to the term to refer to managing, handling and analyzing very large volume of datasets. It is said that father of the word “**Big data**” is John Mashey, who was the chief Scientist at silicon graphics in 1990 (Diebold,

2012). The term big data always remembered as a buzzword in the Industry and commerce only after 2008 (Vossen, 2014). With the invention of World Wide Web and Web 2.0 technologies, there has been huge amount Structure and unstructured data is generated on continuous basis from different sources like emails, different social network platform (such as Facebook, WhatsApp, LinkedIn etc.), blogs, online transactions, articles and forums. Further, different kind of sensor data produce from different sources such as health care Science, environmental organization, metrological department, business data, census data, and company data etc. in larger volume and enormous velocity is termed as Big data. All data are analyzed to get meaningful information and these data are outcome of digital technologies like Smart phone, camera, remote sensing, RFID technology and microphones etc. Data is generated in various formats such as audio, video, text, images and social media post data. These huge data is not static but increase every day exponentially. As a result, traditional computational tools and algorithm as well as data storage and management techniques are not able to deal with all these big data. Hence so many scalable database tools and techniques have developed such as Hadoop, Map Reduce, NoSQL database; HPCC and apache Hive are more reliable, flexible and Scalable. All these technologies generally manage huge volume of data in the form of KB, MB, GB, TB, YB, PB, EB and ZB. So research on big data become very popular in government sector, Academic fields and Industry all over the universe. Hence we have attempted to highlight the research performance in the field of big data in India during the year 2010 to 2019. We have performed Scientometric analysis of big data as indexed in the Scopus database during 2010-2019. The word Scientometric is originate from the Russian language, which means application of Quantitative methods to the sociology of science. Malimov and Mulcheko define “Scientometric as a quantitative of investigating the development of science as an information process”(Nalimov, V.V., Mulchenko, 1969). So Research productivity or output is a measure of quantity of scientific research activity and publication generated by the research scientist during a specific period of time.

## **Review of Related Literature**

*Trivedi, (2019)* explores the present status of Agriculture big data research publication with help of network analysis and visualization. All 379 publications data were downloaded from web of

science-Clarivate analytics with the entire time span and all the data are analyzed using vosviewer, MS-excel and R software. Scientometric parameters are used to analyze the result based on yearly scientific production, highly cited papers, most cited authors, highly cited affiliations, Prolific country, h-index, keyword co-occurrence analysis and density visualization. It is basically investigate the proliferation of agriculture big data which can be help the research scholars and policy maker and other people who are directly or indirectly associated with Agriculture big data for future study.

***Soleimani-Roozbahani, Rajabzadeh Ghatari, & Radfar, (2019)*** have described the role of big data in the evolution of healthcare practices and Research activity. The paper analyzed meta-analysis on published research methodology from 2008 to 2018. It was found that the most used methodology in the published research is “Meta-analysis and Evidence”. The study also revealed Data mining techniques used for predicting knowledge discovering in the field. Rapid miner has been applied in this study. It was examined eight main categories for words used in publications while “developing methods to evaluate of care”. It was found that average is the most highly methodology for publishing papers and “Agent-Based modeling” in nature is maximum used methodology.

***R.Balasubramani and G.T.Khila, (2018)*** have reviewed research productivity on big data with specially authorship pattern during 1998 to 2017. Its aims to analyze the research performance of the scientist in the field of big data. The study used Scientometric analytical tool to analyze the big data publication over the period of 1998-2017 according to the objectives of the study. The study focused on Scientometric tools as pattern of authorship, year wise distribution of authorship pattern, degree of collaboration (DC), most prolific authors in the field of big data research publications during the year 1998-2017. The study examined total 10641 publications on big data research. The study explored that 1064 publication contributed by 27900 authors in 3193 Journals; it also found that 10641 publications were contributed by 7463 institutions, located in 109 countries.

***Akhilesh, Varma, & Singh, (2017)*** have conducted Scientometric study on big data during 2012 to 2016. The main purpose of the study was to revealed find out the research growth and trends on big data research in India over the 5 years. A total 1265 data have been retrieved from Scopus database and all the data have been analyzed with help of different Scientometric parameters

such as Yearly distribution of publications, form wise distribution of papers, Subject wise distribution of publications, most productive authors, and Subject wise distribution of papers, most productive authors and top most ranking of Journals etc.

*Alagu, Thanuskodi, & Dhanya, (2017)* have investigated the research growth and characteristics of big data in India during 2008 to 2017. The bibliographic data has been downloaded from the web of science database on big data research during the study time. A total 714 articles downloaded and indexed in the web of science on big data. All the data were exported to MS-excel for further analysis of data according to the objective of the study. The study revealed the most productive Journal names are “Journal of pharmaceutical biological and chemical science research” among the scientist of India, most of the papers are written by multiple authors, contribution by authors are Single authors accounted for 67 (9.38), Two authors 221 (30.95) and Three authors 149 (20.87%) etc. The Study also focused on year wise distribution of publications, Document wise distribution, Author wise distribution, and top range collaborating institution in the field of big data research.

*Singh, Banshal, Singhal, & Uddin, (2015)* have conducted Scientometrics mapping of research publication in the field of big data during the year 2010 to 2014 as indexed in the database of Scopus and Web of Science. The research analysis focused on different parameters like total research publications, growth of literature on big data, pattern of authorship, country wise collaboration pattern, country wise contribution, Institution wise contribution , top publication sources and thematic trends and emerging themes in the field of big data research.

## **Objectives**

The study deals with the following objectives;

1. To find out relative growth rate and doubling time of literature in big data
2. To find out the Year wise distribution of Research publication of in the field of big data in India.
3. To identify the most prolific author in India on the field of big data
4. To know the pattern of authorship in the field of big data literature.
5. To find out total number of scientific research papers.

6. To know the subject wise scattering of Publications.
7. To find out highly cited paper of big data.
8. To examine degree of collaboration pattern of the researchers at national and international levels.
9. To identify the most productive Source in big data.
10. To find out Published documents types.
11. To reveal the most productive institution and organization

### **Research Methodology**

For this study the required data has been collected from Scopus database during the study period from 2010 to 2019. Scopus is the only source of data for this study. The database covers huge range of article, conference papers, books, reviews, case studies, letters etc. Scopus is an Elsevier product, which launched in 2004, The Scopus database have 75 + million records, 24,600 active titles and 194,000 books from approximately 500 publishers. Its content coverage by the subject area are Health Science 7568 titles (25%), Physical Science 8102 titles (27%), Life Science 4883 titles (16%) and Social sciences 9692 titles (32%). Scopus database cover 3 types of sources such as Book series, trade journals and Journals. It's an abstract and citation data base cover all the international scientific publication to meet the objectives of the study, required data has been collected in the form of Scientometric parameters like types of documents, most cited journals, yearly growth of literature, most prolific author etc. Search Strategy used for data collection in Scopus Database is "Your query : (TITLE-ABS-KEY(big data) AND ( LIMIT-TO ( AFFILCOUNTRY,""India"" ) ) AND ( LIMIT-TO ( PUBYEAR,2019) OR LIMIT-TO ( PUBYEAR,2018) OR LIMIT-TO ( PUBYEAR,2017) OR LIMIT-TO ( PUBYEAR,2016) OR LIMIT-TO ( PUBYEAR,2015) OR LIMIT-TO ( PUBYEAR,2014) OR LIMIT-TO ( PUBYEAR,2013) OR LIMIT-TO ( PUBYEAR,2012) OR LIMIT-TO ( PUBYEAR,2011) OR LIMIT-TO ( PUBYEAR,2010) ) )".The researcher has been searched "Big data "in the search box to get necessary data for the present study.

### **Data Analysis and Interpretations**

A total 130,564 records were found, which have been refined by selecting affiliation country "India", total 7829 documents were found and further researcher refined by the study period

(2010-2019), finally 7502 Publication were found on big data research in India during 2010-2019. The data has been downloaded on 9<sup>th</sup> November, 2019 and further analyze with help of MS-Excel and vosviewer software to meet the objectives of the study.

**Yearly Publication growth:** The following the revealed the year wise distribution of publication from 2010 to 2019.

Table 1: Year wise research publication on big data in India

S. No	Year	Number of Publications	Percentage	Cumulative of publications	Cumulative percentage
1	2010	79	1.05	79	1.05
2	2011	85	1.13	164	2.18
3	2012	128	1.70	292	3.89
4	2013	192	2.55	484	6.45
5	2014	366	4.90	850	11.33
6	2015	742	9.89	1592	21.22
7	2016	1270	16.92	2862	38.14
8	2017	1230	16.40	4092	54.54
9	2018	1768	23.56	5860	78.11
10	2019	1642	21.90	7502	100.00
	<b>Total</b>	<b>7502</b>	<b>100</b>		

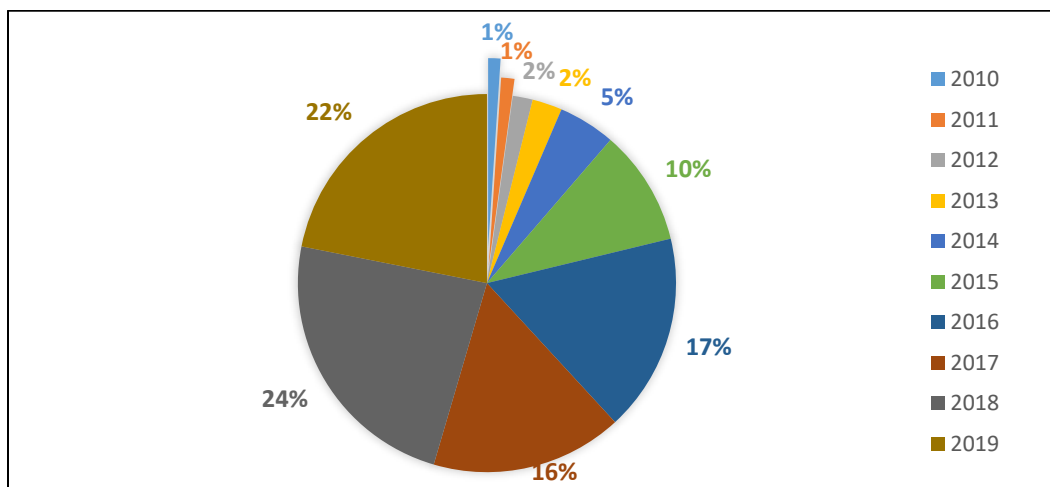


Figure 1: Number of publication per year (Percentage)

The table 1 portrays that author/s of India have contributed 7502 research publications during the period of 2010 to 2019. The annual research output on big data in India is about 750.20 papers per year. The research productivity of initial 6 years was below annual average of 750.20 papers but the research productivity increased exponentially after 2016. During the study period, the

year 2018 have maximum research productivity that accounted for 1768 publications (23.56%) while lowest research recorded in 2010 which only 79 publications (1.05%). The table shows that there is ups and down in research growth in the field of big data in India.

**Types of Research Published:** Following table show the type of research published on big data in India.

Table 2: Type of research published big data in India

S. No	Type of Documents	Number of Publications	Percentage	Cumulative of publications	Cumulative percentage
1	Conference paper	4019	53.57	4019	53.57
2	Articles	2766	36.87	6785	90.44
3	Book Chapter	416	5.54	7201	95.98
4	Review	140	1.86	7341	97.85
5	Book	42	0.55	7383	98.41
6	Editorial	40	0.53	7423	98.94
7	Note	7	0.093	7430	99.04
8	Letter	5	0.066	7435	99.10
9	Short Survey	5	0.066	74340	99.17
10	Erratum	1	0.013	7441	99.18
11	Undefined	61	0.81	7502	100.00
	<b>Total</b>	<b>7502</b>			

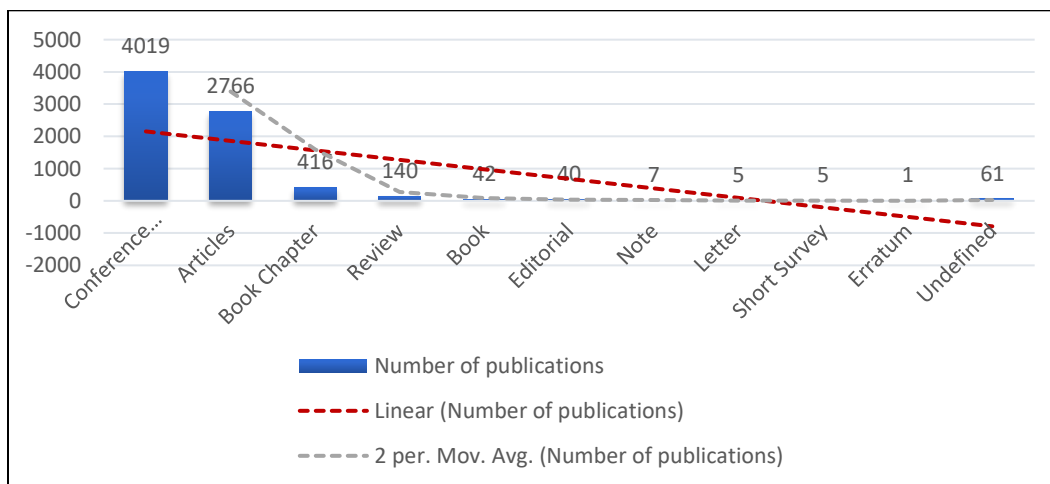


Figure 2: Type of publications



The table no.2 indicates that the types of research paper being published on big data in India, there are 4019 conference papers being produced i.e. (53.57%) which is highest number of research publications during the study period 2010 to 2019, 2766 (36.87%) research articles being published, therefore the research articles are at 2<sup>nd</sup> position. 416 (5.54%) are book chapters and 140 Review articles being produced. From the above analysis, it is observed that researchers like to published their research work more in the form of conference papers i. e. 53.57% of total publication followed by Journal articles 36.87%, book chapter 5.54%, Review 1.86%, book 0.55%, Editorial 0.53% , Note 0.093%, Short survey 0.066% and Erratum 0.013%.

**Relative growth rate and doubling time:** The relative growth rate and doubling time model was advanced by Mahapatra in 1985 and this model is applied to examine relative growth rate of research publications (Mahapatra, 1985).

**Relative growth rate:**

Relative growth rate is the increased in the number of research publications/articles or pages per unit of time and it can be calculated with the following equations.

$$R (1 - 2) = \frac{W1-W2}{T2-T1}$$

Where, R (1-2) means the relative growth rate over specified period of interval

W1=Log w1 (Natural log of initial number of publications/Pages)

W2=Log W2 (Natural log of final number of publications or pages)

T2-T1= the unit difference between the initial time and final time

The relative growth rate for both publications and pages can be calculated separately.

Therefore,

R (a) = Relative growth rate per unit of time (Year)

R (p) =Relative growth rate per unit of pages, per unit of time (year)

**Doubling Time:**

It is found from the equation that there is direct relation between relative growth rate and doubling time. If the number of research publications/ articles/pages of a subject double during the given period then the difference between the logarithms of numbers at the beginning and end

of this period must be the logarithms of number two. If one uses natural logarithms, then this difference has a value of 0.693. Thus the corresponding doubling time for publications and pages can be calculated by the following equations.

$$Dt = \frac{0.693}{R}$$

Where,  $DT$  = Double time,  $R$  = Relative growth

Therefore,

Doubling time for publications data can be calculated as

$$Dt(a) = \frac{0.693}{R(a)}$$

Doubling time for pages can be calculated as

$$Dt(p) = \frac{0.693}{R(p)}$$

Table 3: Relative growth rate and doubling time of research productivity

S. No	Year	Number of Publications	Cumulative	W1	W2	Relative growth rate	Doubling time (Dt)
1	2010	79	79	-	4.37	-	-
2	2011	85	164	4.37	5.10	0.73	0.95
3	2012	128	292	5.10	5.68	0.58	1.19
4	2013	192	484	5.68	6.18	0.50	1.39
5	2014	366	850	6.18	6.74	0.56	1.23
6	2015	742	1592	6.74	7.37	0.63	1.10
7	2016	1270	2862	7.37	7.95	0.58	1.19
8	2017	1230	4092	7.96	8.31	0.35	1.98
9	2018	1768	5860	8.31	8.67	0.36	1.92
10	2019	1642	7502	8.67	8.92	0.25	2.72
	<b>Total</b>	<b>7502</b>					

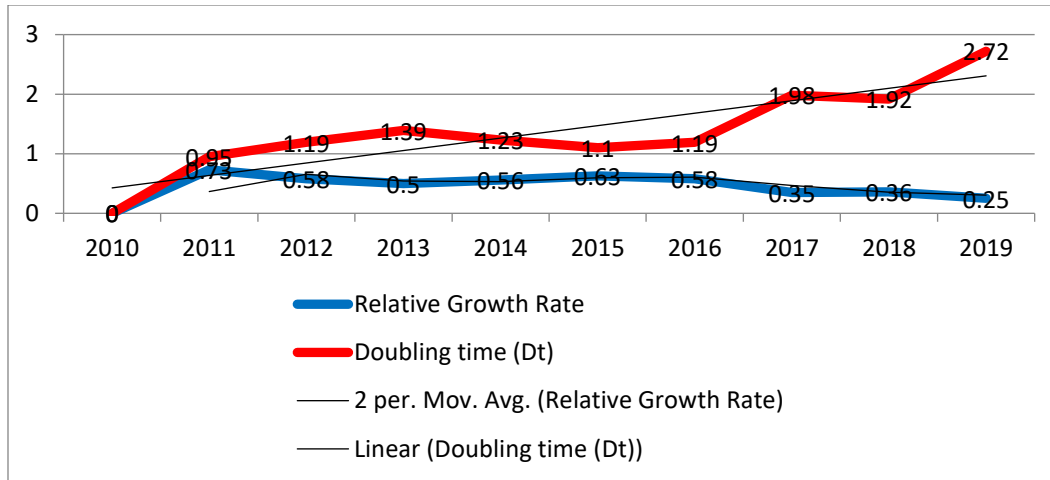


Figure 3: Relative growth rate and double time

The table 3 shows that relative growth rate decreased from 0.73 in 2010 to 0.50 in 2013. While increased from 0.56 in 2014 to 0.63 in 2015 and it is decreased from 0.63 in 2015 to 0.25 in 2019. It is very precise from the above analysis that the relative growth rate decreased from 2010 to 2019. The doubling time of the publication increased from 0.95 in 2010 to 1.39 in 2013 but it's decreased from 1.23 in 2014 to 1.10 in 2015 and again the double time increased from 1.19 in 2016 to 2.72 in 2019. The study revealed that both relative growth rate and double time are fluctuating but relative growth rate decreasing while doubling time of research publication of big data increasing.

**Prolific author/s on big data:** The following table 4 recognizes the top 20 author/s with their names, number of publication and their ranks.

Table 4: Most prolific author/s

S. No	Author Name	Number of Publication	Rank
1	Pandey, M	35	1
2	Kumar, N	22	2
3	Simmhan, Y	22	2
4	Vijayakumar, V	21	3
5	Alam, M	20	4
6	Rautaray, S.S	20	4
7	Manogaran, G	19	5
8	Raj, P	19	5
9	Lopez, G	18	6
10	Rani, R	18	6

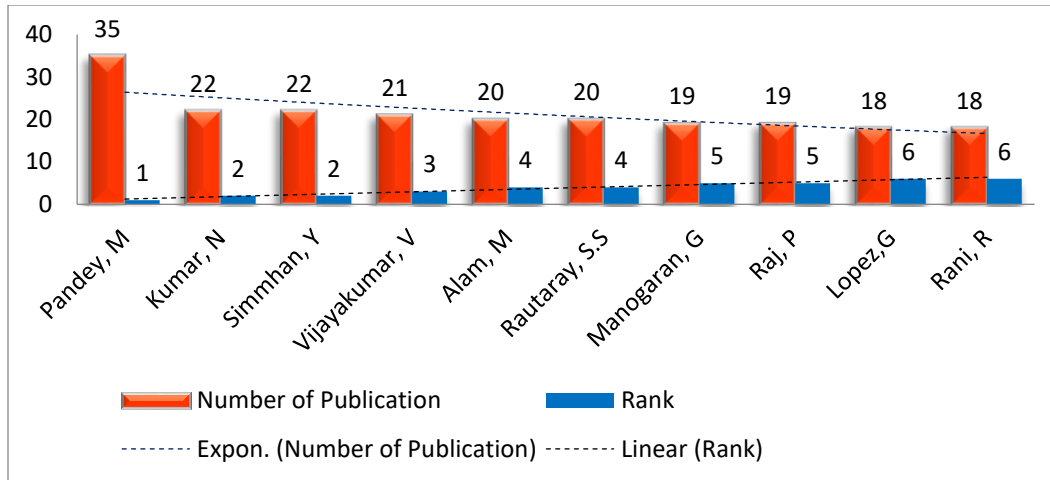


Figure 4: Most prolific Author/s

The table 4 shows the list of most productive author/s of India in contributing big data research publications during 2010 to 2019. India has produced a total of 7502 research publications on big data. It has been found that the most productive author is Panday, M, who has 35 publications and which makes him number one rank holder. Kumar, N and Simmhan, Y has 22 publications each and both have secured 2<sup>nd</sup> rank in producing big data research in India. Vijyakumar, V holds the 3<sup>rd</sup> rank with 21 publications while Alam, M and Rautaray, S.S both have (20) publications each. Manogaran, G and Raj. P has 19 Publications while Lopex, G and Rani, R with 18 publications.

**Year wise Citation on big data research Publications:** The following table shows the yearly citation on big data publication in India.

Table 5: Citations per year

S. No	Year	Number of Publications	Citations	% of Citations
1	2010	79	174	0.44
2	2011	85	1557	3.90
3	2012	128	2167	5.42
4	2013	192	981	2.45
5	2014	366	1115	2.79
6	2015	742	3563	8.92
7	2016	1270	5275	13.22
8	2017	1230	6237	15.62

9	2018	1768	8984	22.50
10	2019	1642	9874	24.74
	<b>Total</b>	<b>7502</b>	<b>39927</b>	<b>100.00</b>

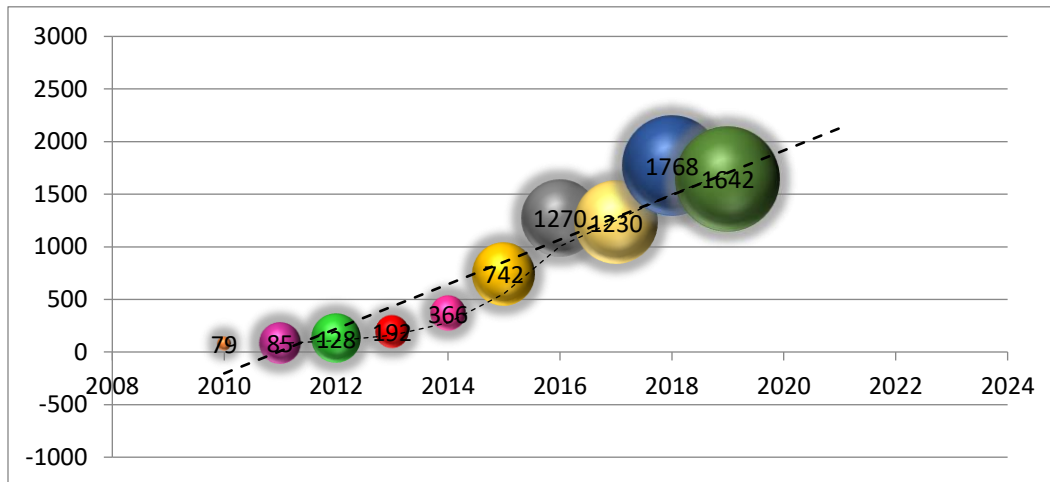


Figure 5: Citations per year

From the table 5 it is observed that during the study period of 10 years (2010-2019) total 39927 citations received against 7502 Publication. The average citations recorded per documents is 5.32. It has been also observed that 2019 has maximum citation with 9874 (24.74%) against 1642 research publications which means 6.01 average citations per documents on big data in India. The lowest citation recorded in 2010 with 174 citations (0.44%) followed by 2011 received 1557 cations (3.90%) with 85 documents, the analysis revealed that Citations are increasing over the study year from 2010 with 174 citation to 2019 with 9874 citations.

**Top 20 most cited of Authors:** Following table 6 study the top 20 most cited authors, the table includes Authors, number of publications produced and number citation received.

Table 6: Top 10 most cited authors

S. No	Author Name	Number of Publication	Citations
1	Dubey, R	11	793
2	Gunasekaran, A	14	768
3	Childe, S.J	8	750
4	Papadopoulos,T	8	389
5	Manogaran,G	19	388
6	Kumar, N	49	348

7	Goudar,R.H	7	324
8	Lopez, D	18	314
9	Kumar, A	80	276
10	Kumar, R	41	231

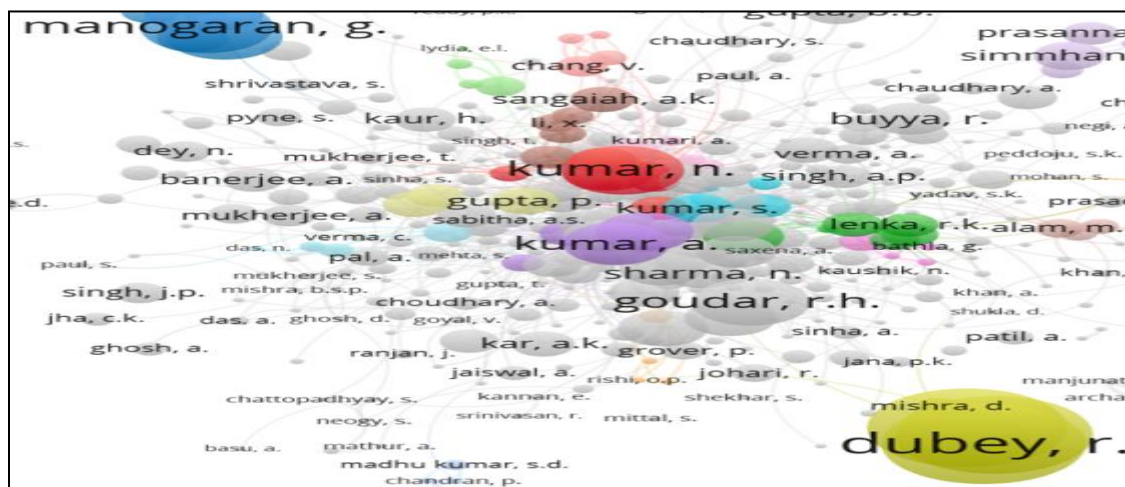


Figure 6: Top most cited authors

The table 6 shows that Dubey, R is the most cited authors; he received total 793 citations with 11 research publications which means 72 average citation per document, followed by Gunasekaran, A, received 768 Citations with 14 documents means 54.85 average citations per document. Childe, S. J has 750 citations with 8 documents, which means 93.75 average citations per documents. Therefore we can conclude that Dubey, R has more citations but Childe, S. J has highest average citations per year.

**Authorship pattern on big data in India:** Following table 7 examines the authorship pattern.

Table 7: Authorship pattern

S.No.	Year	Single Author	Joint authorship	Total
1	2010	11	68	79
2	2011	08	77	85
3	2012	12	116	128
4	2013	25	167	192
5	2014	46	320	366
6	2015	51	691	742
7	2016	74	1196	1270

8	2017	75	1155	1230
9	2018	93	1675	1768
10	2019	60	1582	1642
	<b>Total</b>	<b>455</b>	<b>7047</b>	<b>7502</b>

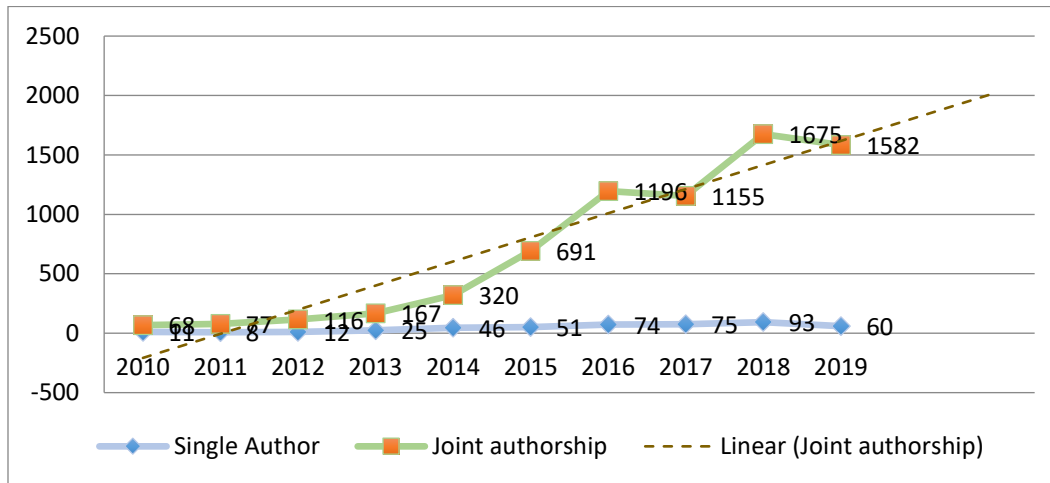


Figure 7: Authorship pattern

Table 7 revealed that there is total 7502 Publication on big data on in India during the year 2010 to 2019. Single author Produced 455 (6.06%) publications out of 7502 and multiple or Joint Authors produced 7047 (93.93%) publications. Indian researcher/s produced 79 publications in 2010 with single author (11) and Joint authors' produces (68) publications while 2019 Single author produced 60 publications and Joint authors produced 1582 research publication. From above analysis it is confirmed that researcher of big data prefer collaborative research. It is revealed from the study that only 6% researchers interested to publish their work under single authorship while 93% scientist of big data prefer joint authorship.

**Degree of Collaboration:** The degree of collaboration is the ratio of the number of co-author or collaborative research publications to the total number of research publications in particular subject area during a particular period of time. The analysis of collaboration of author was given by Subramanyan to find out the degree of collaboration (Subramanyan, 1983). Formula for degree of collaboration  $DC = \frac{Nm}{Nm+Ns}$

Following table 8 studies the degree of collaboration during the time period 2010 to 2019 on big data research publications.

Table 8: Degree of Collaboration

S. No	Years	Single author (Ns)	Multi-Authors (Nm)	Total no. of articles (Ns + Nm)	DC
1	2010	11	68	79	0.86
2	2011	08	77	85	0.90
3	2012	12	116	128	0.90
4	2013	25	167	192	0.86
5	2014	46	320	366	0.87
6	2015	51	691	742	0.93
7	2016	74	1196	1270	0.94
8	2017	75	1155	1230	0.93
9	2018	93	1675	1768	0.94
10	2019	60	1582	1642	0.96
	<b>Total</b>	<b>455</b>	<b>7047</b>	<b>7502</b>	<b>0.93</b>

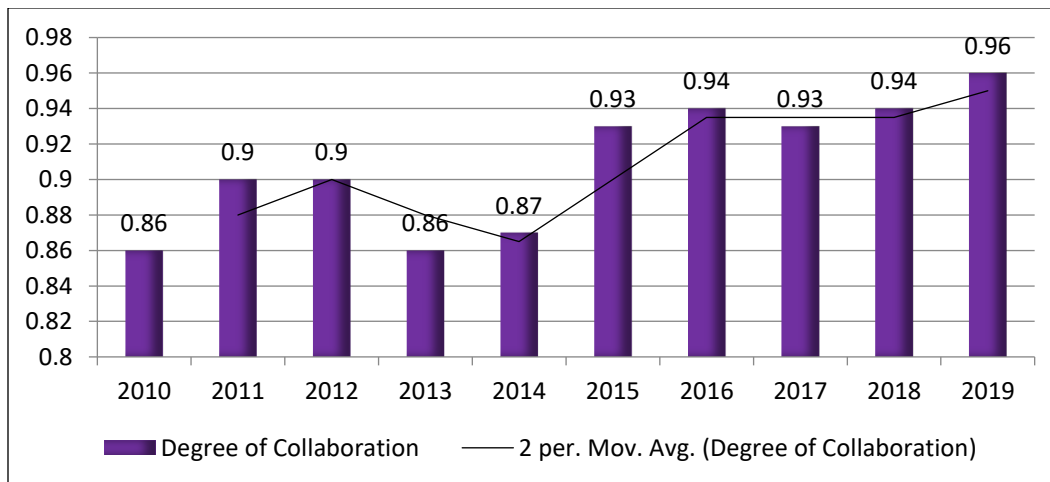


Figure 8: Degree of Collaboration

The table 8 indicated the of collaboration on big data in India during the 10 year study period lies from the range of 0.86 in 2010 to 0.96 in 2019. The Study revealed that multi-authored research produced 93% publication on big data in India during the 2010-2019. Hence the study concludes



that collaborative researches are more active than the Single author research which produces (6%) publications.

The degree of collaboration for the 10 years study period can be calculated as:

$$DC = \frac{Nm}{Nm+Ns}$$

( $Nm$  = multi-authored research;  $Ns$ = Single authored research)

$$DC = \frac{7047}{7502}$$

DC = 0.93 (The degree of collaboration during 2010 to 2019 = 0.93).

**Publication of research by top 9 highly productive Source:** The following the revealed the Source of publication.

Table 9: Sources of Publications

S. No	Source of Publication	Number of Publication	Percentage (%)
1	Advances in Intelligent Systems and Computing	422	5.62
2	International Journal of Recent Technology and Engineering	186	2.47
3	International Journal of Innovative Technology and Exploring Engineering	180	2.39
4	Lecture notes in computer Science Including Subseries lecture notes in artificial intelligence and lecture notes in Bioinformatics	177	2.35
5	Procedia Computer Science	169	2.25
6	International Journal of Applied Engineering Research	157	2.09
7	ACM International Conference Proceeding series	129	1.71
8	Communications in computer and Information Science	129	1.71
9	Journal of Advanced research in Dynamical and control systems	126	1.67
10	Smart Innovation systems and Technologies	103	1.37

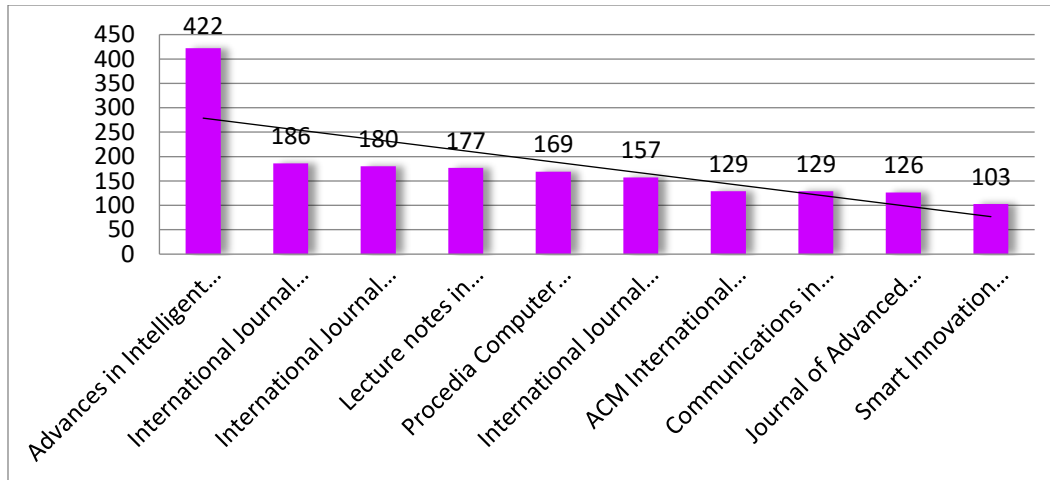


Figure 9: Sources of Publications

It is found that the Advances in Intelligent Systems and Computing having 422 (5.62%) publications followed by International Journal of Recent Technology and Engineering with 186 i.e. (2.47%) publications, International Journal of Innovative Technology and Exploring Engineering with 180 (2.39%) papers etc. Therefore from the above observation, it can conclude that the Advances in Intelligent Systems and Computing is the more popular source of publications on big data research in India.

**Institution wise of research Publication on big data in India:** The following table examines the institutional wise big data research publications produced in India during 2010-2019.

Table 10: Institution wise distributions of research papers

S. No	Name of Institutions	Number of Publication	Percentage (%)
1	Vellore Institute Technology, Vellore	291	3.87
2	Amity University, Noida	271	3.61
3	Anna University	148	1.97
4	Vellore Institute of Technology, Chennai	120	1.59
5	K. L. Deemed to be University	114	1.51
6	Sathyabama Institute of Science and Technology	109	1.45
7	Thapar Institute of Engineering and Technology	90	1.19
8	Kalinga Institute of Industrial Technology, Bhuvaneshwar	79	1.05
9	Bharathiar University	73	0.97
10	Amrita Vishwa Vidyapeetham	70	0.93

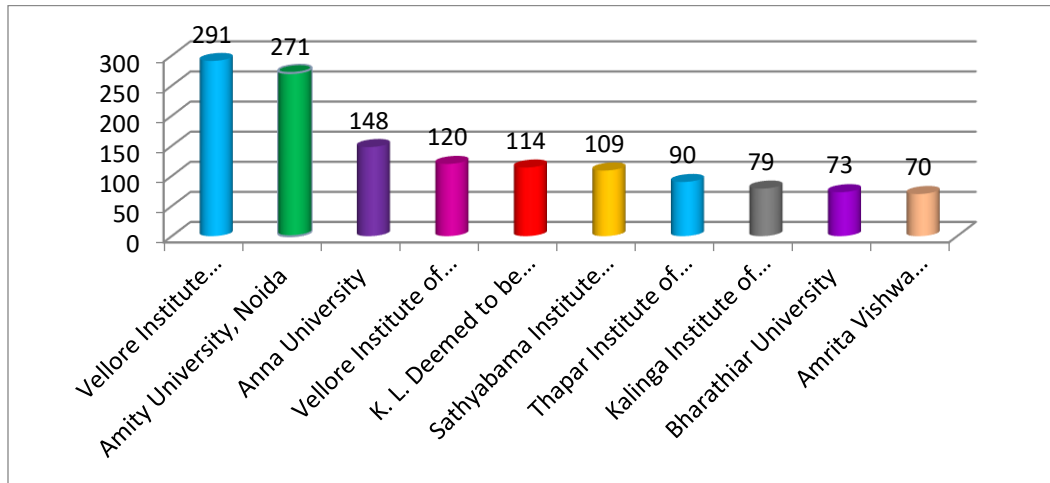


Figure 10: Institution wise distributions of research papers

From the table 11 it is found that Vellore Institute Technology, Vellore has contributed maximum number of research paper in big data amongst all the other institutions in India during the study period of 2010-2019, It has 291 i.e. 3.87% of research publications, followed by Amity University, Noida with 271 research papers i.e. 3.61% publications While Anna University hold 3<sup>rd</sup> position in producing 148 research papers i.e. 1.97%.

**Country wise collaboration of research on big data:** Following table study the country wise collaboration on big data research in India from 2010 to 2019.

Table 11: Country wise collaboration

S. No	Name of County	Number of Publication	Percentage
1	India	7502	100
2	United State of America	346	4.61
3	United Kingdom	145	1.93
4	Australia	96	1.27
5	China	95	1.26
6	France	60	0.79
7	Canada	55	0.73
8	Germany	52	0.69
9	South Korea	50	0.66
10	Japan	44	0.58

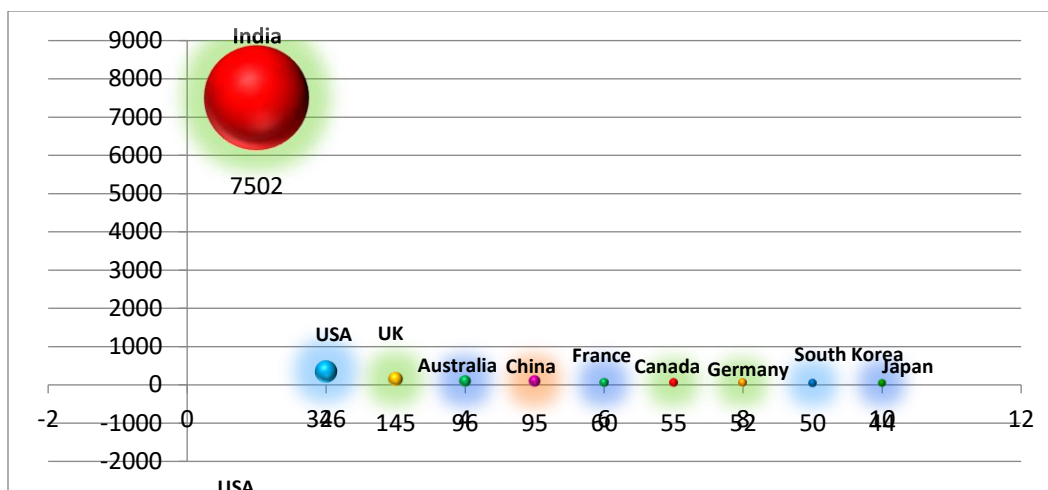


Figure 11: Country wise collaboration

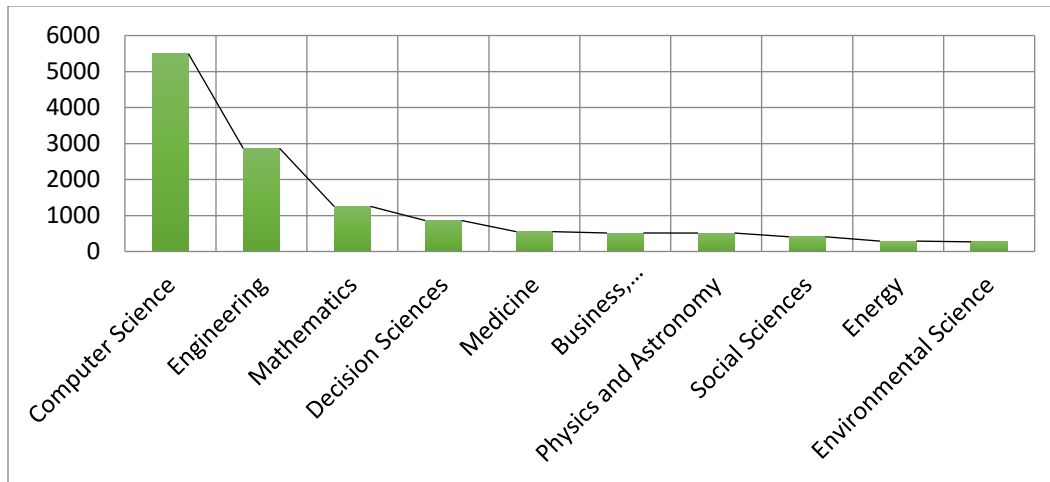
From the above table no. 12 it has been found that the researcher/s of India collaborated with the researcher/s of the countries like *United State of America, United Kingdom, and Australia* etc. The highest research collaboration of Indian author/s have been found with United State of America; these collaboration have resulted 346 i.e. 4.61% research publications followed by United Kingdom 145 (1.93%) publications, while Japan contributed 44 (0.58%) collaborated publications.

**Top 25 Subject wise Publication of big data:** Following table represents the subject or discipline wise distribution of publications by the researcher/s of India on big data 2010-2019.

Table 12: Subject wise distribution of research papers

S. No	Subject area	Number of Publication	Percentage (%)
1	Computer Science	5490	73.18
2	Engineering	2867	38.21
3	Mathematics	1253	16.70
4	Decision Sciences	861	11.47
5	Medicine	551	07.34
6	Business, Management and Accounting	513	06.83
7	Physics and Astronomy	512	06.82
8	Social Sciences	408	05.43

9	Energy	288	03.83
10	Environmental Science	267	03.55



*Figure 12: Subject wise distribution of research paper*

From the above table no.13, it is absolutely clear that Computer science is the most preferred discipline on which big data researches have produced, there are 5490 (73.18 %) publications on Computer science followed by Engineering on which 2867 (38.21%) publications produced etc.

### **Findings**

- The table 4.1 portrays that author/s of India have contributed 7502 research publications during the period of 2010 to 2019 as indexed in the Scopus database.
- The year 2018 have maximum research productivity that accounted for 1768 publications (23.56%) while lowest research recorded in 2010 which only 79 publications (1.05%)
- It is observed that researchers like to published their research work more in the form of conference papers i. e. 53.57% (4019) while Erratum has minimum publication i.e 0.013% (1)
- It is observed that the relative growth rate decreased from 2010 (0.73) to 2019 (0.25)
- The doubling time of the publication increased from 0.95 in 2010 to 2.72 in 2019.
- It is found that the most productive author is Panday, M, who has 35 publications and which makes him number one rank holder.

- The analysis revealed that Citations are increasing over the study year from 2010 with 174 citations to 2019 with 9874 citations.
- The table 4.6 shows that Dubey, R is the most cited authors; he received total 793 citations with 11 research publications which means 72 average citation per document, followed by Gunasekaran, A, received 768 Citations with 14 documents means 54.85 average citations per document
- It's revealed from the study that only 6 %( 455) researchers interested to publish their work under single authorship while 93% (7047) scientist of big data prefer joint authorship.
- The table 4.8 indicated the of collaboration on big data in India during the 10 year study period lies from the range of 0.86 in 2010 to 0.96 in 2019.
- The degree of collaboration during 2010 to 2019 = 0.93
- The *Advances in Intelligent Systems and Computing* having 422 (5.62%) publications and is the more popular source of publications on big data research publication in India.
- *Vellore Institute Technology, Vellore* has contributed maximum number of research paper in big data amongst all the other institutions in India during the study period of 2010-2019 with 291 i.e. 3.87% of research publications.
- The highest research collaboration of Indian author/s have been found with United State of America with 346 i.e. 4.61% research publications followed by United Kingdom 145 (1.93%) publications.
- It is found that Computer science is the most preferred discipline on which big data researches have produced, there are 5490 (73.18 %) publications on Computer science followed by Engineering on which 2867 (38.21%) publications produced.

## **Conclusion**

The Research productivity successfully analyzed on big data publications in India for past 10 years focusing on different Scientometric parameters and on the basis of results and finding following conclusion been drawn. The research on big data in India is growing very fast in the present decade. The result discloses that total number of 7502 publications produced on big data during 2010-2019. It shows that the conferences papers (53.57%) are the highest ranked in publishing researches since conferences are conducted very frequently on big data in India. It has

been found that Indian Scientist collaborated with other foreign national in their research; most collaborated country is USA (4.61%), UK (1.93%), Australia (1.27%), France, and Canada etc. It's been revealed from the study that only 6% scientist interested to publish their work in single authorship while 93% scientist of big data prefer in joint authorship. Therefore, there's an increasing trend on Joint authorship research in big data in India. Moreover the study also revealed that Vellore Institute Technology, Vellore is the most productive institute that Contributed 291 (3.87%) research papers. It is also observed that most preferred subject of the researcher is "Computer Science" with 5490 (73.18 %) Publications. Therefore, it can be stated that computer science professional have more tendency to conduct research on big data. It has been found from the result that most productive author is Panday, M with 35 publications and ranked 1. From the yearly analysis it can be concluded that the scientific research production in the field of big data is increasing exponentially due to the collaboration among the researcher from different subjects. Finally, a details analysis and picturizing of research on the field of big data have been presented, which is very informative, useful and 1<sup>st</sup> of its kind on the theme of big data.

## References

- Akhilesh, M., Varma, K., & Singh, K. (2017). Scientometric Analysis of Literature on Big Data Research Output in India Based on Scopus Database. *International Journal of Library and Information Studies*, 7(4), 90–96. Retrieved from <http://www.ijlis.org>
- Alagu, A., Thanuskodi, S., & Dhanya, P. (2017). Big Data : A Scientometric Analysis Based on Indian Publications, 29, 84.
- Diebold, F. X. (2012). On the Origin(s) and Development of the Term "Big Data." *SSRN Electronic Journal*. <http://doi.org/10.2139/ssrn.2152421>
- Mahapatra, M. (1985). On the Validity of the Theory of Exponential Growth of Scientific Literature. In *Proceedings of the 15th IASLIC Conference* (pp. 61–70). Bangalore.
- Nalimov, V.V., Mulchenko, Z. . (1969). *Naukometriya Izucheniye Razvitiya nauki kak Informatsionnogo protsessat*. Moscow.

R.Balasubramani and G.T.Khila. (2018). Authorship Pattern of Big Data Research Output :, 7(1), 15–19.

Singh, V. K., Banshal, S. K., Singhal, K., & Uddin, A. (2015). Scientometric mapping of research on 'Big Data.' *Scientometrics*, 105(2), 727–741. <http://doi.org/10.1007/s11192-015-1729-9>

Soleimani-Roozbahani, F., Rajabzadeh Ghatari, A., & Radfar, R. (2019). Knowledge discovery from a more than a decade studies on healthcare Big Data systems: a scientometrics study. *Journal of Big Data*, 6(1). <http://doi.org/10.1186/s40537-018-0167-y>

Subramanyan, K. (1983). Bibliometric studies of research collaboration: a review. *Journal of Information Science*, 6(1), 33–38.

Trivedi, G. (2019). Visualization and Scientometric Mapping of Global Agriculture Big Data Research.

Vossen, G. (2014). Big data as the new enabler in business and other intelligence. *Vietnam Journal of Computer Science*, 1(1), 3–14. <http://doi.org/10.1007/s40595-013-0001-6>