ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 6, Issue 6, June 2019

Using Big Data Analytics In Order To Understand and Take Care of Environmental Emergencies

[1] Dr. Yogesh Kumar Sharma, [2] S.V.G.Sridevi
[1] Associate Professor, Department of Computer Science & Engineering, [2] Research Scholar
[1][2] Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan

Abstract: With its growing roots, the Big Data Analytics is finding its way to help the mankind in every walk of their life. To understand big data, its principles, its features and its usage in various walks of critical fields is a commendable job. The major purpose of this paper is to throw some light on what is Big Data, how Big it really is. The paper also highlights how the Environmental Agencies are trying to move conveniently towards the Big Data technologies in order for a better performance. The usage of Big Data in various fields of science shows, unknowingly how do we use Big Data? It also tries to highlight, the advantages and limitations that are involved. Two articles are disaster management using Big Data have been reviewed to focus on the reason why Big Data is being the buzzword.

I. INTRODUCTION

the user can draw some conclusions or infer some facts. The term data science refers to the study of data. With big data, the process is bit advanced as with traditional processes and techniques the data processing becomes complex and tedious. It should be noted that the term big data doesn't imply because of the size of the data, but it is termed so based on five important metrics which are collectively termed as the 5 Vs of big data:
i) Volume
ii) Variety
iii) Veracity
iv) Velocity
v) Value

Volume indicates how much data is being generated in a specific time period. As per the current statics, the data is getting generated in yotta bytes.

Variety indicates what types of data is being generated. Earlier the data type was confined to only text. But now we see data of various data types like images, audio, video etc.,

Veracity is the new feature now being added with every application to check the data that is being generated is accurate. If it is not from a trusted source, the data can be rejected.

Velocity concentrates on with what speed the data is getting generated. It not only checks the speed of data generation but also considers the speed of capturing and speed of data processing.

Value is the measure of usefulness of the extracted data to the users of big data.

I.1 HOW BIG THE DATA REALLY IS???

Counting Very Large Numbers		
Byte (B) =	1 byte =	1=One character of text
Kilobyte (KB) =	10^3 bytes =	1,000 = One page of text
Megabyte (MB) =	10^6bytes =	1,000,000 = One small photo
Gigabyte (GB) =	10^9bytes =	1,000,000,000 = One hour of High-Definition video
Terabyte (TB) =	10^{12}bytes =	1,000,000,000,000 = The largest current hard drive
Petabyte (PB) =	10^{15}bytes =	1,000,000,000,000,000 = AT&T currently carries about 18.7 PB of data traffic on an average business day
Exabyte (EB) =	10^{18}bytes =	1,000,000,000,000,000,000 = Approximately all of the hard drives in home computers in Minnesota, which has a population of 5.1M
Zettabyte (ZB) =	10^{21}bytes =	1,000,000,000,000,000,000,00= Approximately data in 174 news papers

Table 1: data measurements

I.2 WHERE BIGDATA IS USED???

Big data is used in various fields of business such as finance, marketing, urban planning, public health administration, epidemic outbreak etc., In genetic sciences and ecology, there is a disparity on the amount of data getting generated and the techniques for processing that enormous amounts of data.

With the indispensable social networking applications it is evident that a voluminous and variety of data is generated every second. Processing is almost in competition with the growth of data. According to the publication, How Much Information? 2009 Report on American Consumers, the rate of growth of information usage is

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 6, Issue 6, June 2019

approximately 5.4 percent. Put together all data uses, this percentage can be higher.

I.3 USES OF BIG DATA IN ENVIRONMENTAL SCIENCES

It is no surprise that all over the world, almost all government bodies and university research departments are using Big Data to assist in research and decision-making. Furnishing below some of the areas where Big Data finds its existence extensively.

I.3.1. The Environmental Protection Agency (EPA) and Public Health

One of the biggest areas big data with environmental science is public and environmental health. We are already witnessing improvements in monitoring and mitigation of issues of industrial chemicals released into the atmosphere and the levels of damage they are causing. Monitoring had always used traditional methods of environmental sampling, but now data can be processed through computational methods, of which the results are more accurate, more up-to-date. The results are produced much faster produced, with more options for analysis that allow the environmental science experts to make an informed decision. Big Data allows for high throughput, combined data sets and meta-analysis, and deeper analysis of the results produced from these studies.

EPA is presently using such data acquired through Big Data Analytics to synthesize more accurate predictions for areas where data either does not exist or is difficult to acquire. Also, researchers can identify gaps in the data and potential vulnerabilities in the system and process of investigation. Overall, this mitigates the problems and enhances data for better decision making for public health concerns. They are now working with NCDS (National Consortium for Data Science) to identify current challenges that they hope to address through big data science.

I.3.2. Geographic data

Few of the big data tools have proven as useful to so many environmental sciences as the map. For example, the cartography used for naval navigation, geographic surveying. The uses include, the Geographic Information Systems that thrive on Big Data. The GIS strength lies in its ability to summarize, utilize and present various statistical data reports. The more the data is from a geographic area, the better the quality of the output and the more a proper decision is likely to come. So far the GIS is majorly contributing to spatial analytics, and that's good news for GIS technicians and for those people charged with making decisions based on the outputs of their data.

As an example we can quote the disaster recovery. As recently as 2017, a researcher showed in a seminal study that it would be possible in future to parse textual references to GIS databases for up-to-the-minute problem areas currently suffering from tsunamis, flooding, and earthquakes. This would not have been possible if we still stick to the age old methods. Satellite data and aerial imagery have already informed GIS in disaster management, with Hurricane Katrina being one of the first and best-known choices in using the technology. In future, Big Data will further enhance its efficacy and also prove its amalgamation is essential with the data in every possible field.

The EPA is using geographic data to inform research into public health through the Environmental Quality Index.

I.3.3. Climate Change and Planetary Monitoring

In 2013, the UK government announced large-scale investment in Big Data infrastructure for science, particularly in the environmental sector. Of particular note to global research was a commitment to maintaining funding for a program called CEMS (Climate and Environmental Monitoring from Space). This led to the creation of larger databases to cope with the increasing Big Data revolution and to allow research partner organizations to work with more data and produce more promising and useful results. With a specific focus on climate change and planetary monitoring, CEMS storage removed the need to download enormous data sets while reducing the costs of access. It provides the tools as well as the data, allowing for greater efficiency, sharing in the academic community, and providing resources once beyond the reach of many institutes due to budgetary restrictions alone. Along with Cloud data, this is now the standard accepted globally for some of the world's top research institutes.

Also, one of the popular universities of UK, announced its plans to open a Big Data center exclusively for environmental science research and analysis in order to bridge the "data gap" between the people who do research global environmental problems and those who have the authority to make decisions to remedy such issues. Working in collaboration to see how big data analytics can be applied to a variety of issues that prevail in risk management and natural disasters. In situations of increased frequency of erratic and extreme weather, Lighthill is now working with commitment to develop global databases and preparing all possible case studies for sharing data Such cross-government and partnerships

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 6, Issue 6, June 2019

between industry and government are working as shown with the previously discussed EPA programs and the EU-wide Copernicus Climate Change Service which recently went live.
As far as the sciences are concerned, climate modeling could be the single most important area of academia for Big Data applications. Early 2010, NASA started to capture and store Big Data to create climate models and to give accurate climatic predictions. It was estimated that these models maintained a data of almost 32-peta bytes.

I.3.4. AGRITECH

The ever-growing global population is putting more pressure on availability of resources. This leads to the investments on some important developments in one of the major sectors, the agritech. It is projected that barring major disasters, the global population will reach 9 billion by 2060, with the highest growth rate in poorer countries. This means a lot of investment in agricultural systems to cope. One of these is Genetic Modification(GM) technology, which is expected to help the world's poorest communities grow disaster resilient crops for sustainable food supply and economies. However, GM alone will not solve this problem.
Essential resource management plans must be made to ensure we can make the most out of agricultural land and effectively use ground nutrients yet limiting the dangerous deforestation, properly managing normal as well as ground water resources and developing new methods of cultivation that could use even less space than before. In the US, some notable agricultural organizations are already using crowdsourced data in conjunction with data from remote sensors and the data that is publicly available such as weather forecast information. This leads to the creation of Big Data sets so that domestic farmers can improve land utilisation to the maximum extent, there by maximizing productivity and revenue stream. Here, Big Data is used in environmental engineering to inform farmer what crops they should grow this year and even the likely event of when their machinery will break down. This information may be used for crop management and/or order spare parts ahead of time so that work is not lost in the second.
This is expected to be even more important for people who live in "marginal landscapes". These areas suffer from lower agricultural growth and low production. This is due to erratic water supply, less precipitation and acidic / alkaline soils. The use of Big Data here is two-fold: firstly, providing mitigation and management tools for marginal landscapes already in use. Second, identifying the best uses for marginal landscapes not already turned over to agriculture.

I.3.5. GENETIC STUDIES

Although this field is not technically an environmental science, it has many uses that are beneficial to the environment from Genetic Modification technology to facilitate gene mapping, examination of the already spread and likelihood of transmission of infectious diseases in vital food crops, for example, Panama Disease in bananas. It's useful in a wide range of biological sciences. We can expect many advances in genetics to come with the advent of Big Data.
The process of decoding human genome took almost 10 years. Now, with Big Data analytics, the process takes 24 hours as estimated by OECD (Organization for Economic Cooperation and Development). Faster research of genetic structures means faster reaction and identification to problematic genes and faster implementation for prevention / correction measures.

I.3.6. CITIZEN SCIENCE

One of the unexpected benefits of Big Data to any field of science, but particularly the environment is the so-called "Citizen Science". This is the accumulation of data as reported by the people in geographic locations all over the world voluntarily willing to share information on their living conditions. It is often beyond the financial and time resources of researchers to investigate all claims directly, so they rely on local people to report such information. This is not new, but the term "citizen science" and the over public engagement is new. Indeed, there are many examples of successful citizen science projects already such as the Christmas Bird Census of 1900 and that came long before global communication, cloud storage and mobile technology. These three are the technologies that have enabled public engagement like no other.
When many people report a phenomenon, it reduces the possibility of hoax, chances of misinterpretation and likeliness fake reporting.

I.3.7. ANTHROPOLOGY & ARCHAEOLOGY

The study of people of the past and their mortal / material remains may not considered as a Big Data application, mostly because they always tend to study small groups of individuals on specific sites. However, compiling such data can have benefits to a larger extent to determine the spread of technology, Cultural Revolution, and even track the spread of ancient farming practices such as slash and burn. In 2017, it was suggested that Big Data could be

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 6, Issue 6, June 2019

used to explore through the old excavation reports to "data mine" in a hope of extracting new information. Archaeologists and anthropologists often deal with complex data, comparing site analyses and trying to link up otherwise seemingly disparate data sets.

I.3.8. IN ENVIRONMENTAL CONSERVATION

It was reported in 2014 that Big Data was not yet part of the world of sustainability and environmental conservation. Yet some applications of Big Data analytics in the field of climate science and climate modeling, there are still few areas where Big Data is useful in such areas as land conservation, sustainability and local environmental diminution. The report was acknowledging a wide range of important areas that could definitely get benefitted from the application of Big Data and Big Data Analytics. And these areas of prospective applications of Big Data analytics include:

- Environmental NGOs may use Big Data in protecting individual landscapes from encroaching.
- Subject Matter Experts can accumulate and provide information for all clients who need it.
- Corporate bodies may use Big Data for two major reasons: firstly as a proof that they are complying with government regulation pertaining to their industry and sector; secondly to investigate and take necessary measures that may be caused by their respective industrial operations.
- International organizations that wish to work with environmental areas / policies, can use the data for analyzing before starting their practices.
- Government bodies in determining policy and bills on environmental regulation and sustainability. At present, the US is working with the Dutch government in ensuring open data policy for Big Data analytics in this area.

I.3.9. IN REGIONAL PLANNING

Urban landscapes are often neglected when the issues related to environmental sciences are considered. The urban centers are themselves having their own ecosystems. They are a crucial and may affect the environment. Yet studies in urbanism represent some of the best and earliest examples of the application of Big Data. In 2014, a report on China's applied statistics and Big Data to examine urban systems and urban-rural planning highlighted the project (begun in the year 2000) as a major success. 2014 was the year they engaged in rapid expansion of the practice. It requires a unification of data between information technologists, geographers, logistics and urban planning.

- Big Data can be applied to examine problems of traffic and assist in decision making on where to lay roads and where to go for extensions of roads, where to make it as a one way and so on. Establishment of centers to deal with crimes and the areas of enforcing law and order. Understand Health problems caused due to pollution.
- In order to deal with all the above stated issues, the available standard data sets may be insufficient and a proper urban planning requires to take information from various sources such as demographics, geographical regions, availability of resources, employment status indicators, pollution indices, employability issues, factors related to medical and health and many more to understand the complex parts that go into making an urban center planning and functionalities.
- Big Data should assist to improve the process of urban planning and resource allocation. In fact, it's already doing so. More recently, studies have shown the usefulness of Big Data in planning "smart urban planning" through large data sets, and the relative usefulness of doing so in future. It is almost proving to be both a time saver and a money saver.

II. BIG DATA USAGE IN DISASTER ANAGEMENT - the literature survey

In their article "Big data and disaster management: a systematic review and agenda for future research", Akter, Shahriar and Fosso Wamba and Samuel, discussed that the era of big data and analytics is opening up new possibilities for disaster management (DM). Due to its ability to visualize, analyze and predict disasters, big data is changing the humanitarian operations and crisis management dramatically. Yet, the relevant literature is diverse and fragmented, which calls for its review in order to ascertain its development. A number of publications have dealt with the subject of big data and its applications for minimizing disasters. Based on a systematic literature review, this study examines big data in DM to present main contributions, gaps, challenges and future research agenda. The study presents the findings in terms of yearly distribution, main journals, and most cited papers. The findings also show a classification of publications, an analysis of the trends and the impact of published research in the DM context. Overall the study contributes to a better understanding of the importance of big data in disaster management.

Rahman, Md & Di, Liping & Esraz-Ul-Zannat, Md. (2017), in their article, "The role of big data in disaster management", highlighted that the Loss of life and property damage from disaster is one of the major concerns around the globe. Effective disaster management is prime attention for all governments and policymakers. Computer generated data, data from vast sensor networks, and data from social media bring the opportunity to improve disaster management. Big data from these sources is getting attention from disaster managers and researchers. Management and analytics of big data can play a vital role in all phases of disaster management. Though the challenges are also remaining to use these big data in an emergency, but big data already proves its usefulness. As big data application in disaster management is comparatively new, many improvements in capacity level, management level, and research level will help to get the maximum benefits from this modern opportunity.

III. ADVANTAGES OF USING BIG DATA IN ENVIRONMENTAL SCIENCES

The advantages of Big Data are numerous:

i) For collecting, Sorting, Analyzing, Presenting Quickly

As discussed in Introduction section, Big Data always takes huge volumes of data. The collected data can be subjected for sorting / dividing into clusters which may help to analyze and present any kids of reports.

ii) Error Mitigation

It helps in handling errors in data, presenting reports, rogue data and anomalous results which is one of the major problems in any field of science. Studies are often limited by sample size alone because of the sample collecting resources and sample generating resources. As data set grows, the error rate in the collected data that can damage the overall result. The data error rate is also coupled with the cost incurring in the application planning, development and its operation. But with Big Data, all these errors can be predicted and reduced as much as possible.

iii) Better Environmental Management

This applies to urban management as our cities continue to undergo rapid and vast changes in line with changing technology and demands of residents. In one study, the Norwegian capital of Oslo was able to reduce its energy consumption through the application of Big Data Analytics when examining its energy resources. At the same time, predictive reporting and risk analysis can reduce the crime rate. Portland in Oregon uses a system to analyze stop light changes at intersections in order to manage traffic flow better. After just six years, the city eliminated 157,000 metric tonnes of CO_2 emissions. Traffic flow varies as a city grows; what was once a sufficient stoplight pattern can change.

iv) Better Decision Making

It is almost a proven fact that Big Data is making the data management, study and summarizing activities simpler which was a tedious job once. It was limited due to the availability of resources, but with its increased access and availability, it is expected to permit easier presentation and reporting, delivering more confident results and therefore, better to aid decision makers and policy development professionals. Scientists and government can work together more efficiently in future, not just to react to the environmental problems of today, but work with greater foresight today to make better decisions for tomorrow.

IV. CHALLENGES FOR BIG DATA IN ENVIRONMENTAL SCIENCES

As environmental changes cannot be kept constant, in spite of its usefulness, Big Data suffers from certain limitations:

- Methods of data gathering
- The availability of storage technology and devices to store volumes of data.
- Analyzing data during the time of it capture.
- Ability to search, share and process the data.
- Ability to query the data.
- Updating the information in line with recent changes
- Data security, privacy issues and the sources of storage
- Problems with human error - wrongly entered data

V. CONCLUSION

Despite the problems, it is an undoubted fact that Big Data is finding its way in almost all the areas we are associated with, for our day to day activities. It is becoming indispensable in every area where it is used. Majorly, the contribution of Big Data analytics is commendable in studying, protecting and taking corrective measures of the environment we live.

VI. REFERENCES

1. https://www.nature.com/articles/455001a
2.https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know/

3.http://highscalability.com/blog/2012/9/11/how-big-is-a-petabyte-exabyte-zettabyte-or-a-yottabyte.html
4.https://www.coursera.org/learn/big-data-introduction/lecture/IIsZJ/characteristics-of-big-data-velocity
5.https://www.sciencedirect.com/science/article/pii/S1364815216304194
6.https://www.weforum.org/agenda/2015/02/a-brief-history-of-big-data-everyone-should-read/
7.http://maajournal.com/Issues/2002/Vol02-1/Full3.pdf
8.https://www.census.gov/history/www/innovations/technology/the_hollerith_tabulator.html
9.https://en.oxforddictionaries.com/definition/information_explosion
10.https://blogs.scientificamerican.com/guest-blog/how-alan-turing-invented-the-computer-age/
11.https://www.census.gov/history/pdf/kraus-natdatacenter.pdf
12.https://dl.acm.org/citation.cfm?id=363790.363813&coll=DL&dl=GUIDE
13.https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#79c899d165a1
14.https://hbr.org/2012/11/2012-the-first-big-data-electi
15.https://obamawhitehouse.archives.gov/blog/2012/03/29/big-data-big-deal
16.https://blog.epa.gov/blog/2016/10/filling-the-gaps-in-environmental-science-with-big-data/
17.https://www.integratesustainability.com.au/blog/view-more.php?id=79
18.https://catalog.data.gov/dataset/usepa-environmental-quality-index-eqi-air-water-land-built-and-sociodemographic-domains-transf
19.http://www.nerc.ac.uk/innovation/activities/environmentaldata/bigdatacapital/
20.https://sa.catapult.org.uk/facilities/cems/
21.https://www.reading.ac.uk/news-and-events/releases/PR604426.aspx
22.https://lighthillrisknetwork.org/research-priorities/
23.http://ilsirf.org/wpcontent/uploads/sites/5/2017/08/2017_WorldBank_Chapter_15.pdf
24.http://onlinelibrary.wiley.com/doi/10.1111/gcbb.12078/full
25.https://www.wur.nl/en/newsarticle/World-firstPanama-disease-resistant-Cavendish-bananas.htm
26.http://www.oecd.org/sti/ieconomy/Session_3_Delort.pdf#page=6
27.https://jcom.sissa.it/sites/default/files/documents/JCOM_1602_2017_C05.pdf
28.https://www.nesta.org.uk/digital-social-innovation/citizen-science
29.https://www.israel21c.org/5-israeli-precision-ag-technologies-making-farms-smarter/
30.https://www.nasa.gov/centers/goddard/news/releases/2010/10-051.html
31.https://www.universiteitleiden.nl/en/news/2017/05/exploring-the-opportunities-of-big-data-in-archaeology
32.https://www.academia.edu/14362660/Think_big_about_data_Archaeology_and_the_Big_Data_challenge
34.http://www.smithschool.ox.ac.uk/publications/wpapers/workingpaper14-04.pdf
35.https://www.sciencedirect.com/science/article/pii/S2226585615000217
36.https://www.sciencedirect.com/science/article/pii/S0167739X17308993
37.http://www.pnas.org/content/113/39/10729
38.https://datafloq.com/read/the-power-of-real-time-big-data/225
39.http://datascienceseries.com/stories/ten-practical-big-data-benefits
40.https://ico.org.uk/media/fororganisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf
41.Akter, Shahriar & Fosso Wamba, Samuel. (2017). Big data and disaster management: a systematic review and agenda for future research. Annals of Operations Research. 10.1007/s10479-017-2584-2.
42.Rahman, Md & Di, Liping & Esraz-Ul-Zannat, Md. (2017). The role of big data in disaster management.
43. Bohn, Roger & E. Short, James. (2009). How Much Information? 2009 Report on American Consumers.