Use of Big Data to Improve Environmental Sustainability in Developing Countries

Maged Adel Abdo Mukred[1] & Zheng Jianguo[2]

[1]The Glorious Sun School of Business and Management, Dong Hua University, Shanghai, China

Correspondence: Maged Adel Abdo Mukred, The Glorious Sun School of Business and Management, Dong Hua University, Shanghai, China. E-mail: mukred@qq.com

Abstract

Big data inhibits the ability to significantly impact a wide range of fields in an economy, from the government sector to commercial sectors like retail and healthcare. Not only has it altered the way companies assess their product's demand and supply patterns but has also phenomenally helped in making the environment healthier in recent years. It carries the ability to identify valuable data from a huge dataset with exceptional parallel processing. This study presents the general introduction of big data bringing forth its various features and advantages along with the challenges which organizations face while using with respect to environmental sustainability. Observations have also been made on the findings of various researches, and studies and surveys performed by some international organizations in the recent years on the urgent need of taking necessary measures and initiatives to prevent further depletion of natural resources thus making the environment sustainable. Making the issue the study aim, future studies must intend to explore how multinational corporations can enhance environmental sustainability through big data analytics. Lastly, recommendations have been made to organisations– private and public in hiring adequate expertise and set-up, thereby making big data analytics more efficient and reliable.

Keywords: big data, environmental sustainability, developing countries, technology, expertise

1. Introduction

Big data and its applications have been instrumental in solving a number of problems in organizations in various fields for over a decade now. In the computing era, datasets which are very large and too complex to be processed by traditional software are processed through big data analytics. Patterns and relations between the pieces of data available in structured or unstructured form are identified first, then they are processed by the means of big data software. This process is sometimes also referred to as just "analysis", depending upon the scale and complexity of the data. Big data analytics can be defined as "the process of testing or analysing large sets of data uncovering hidden patterns, preferences of consumers, market trends, and, unidentifiable correlations".[30] The growing relevance and use of big data among leading tech-giants provide a platform for improvement in many operational aspects, including demand forecasting, inventory management, and supply chain management using popular applications.

Big data is, usually, characterized by 4V's, as follows:

Volume	The quantity of retrieved and stored data which defines if it should be considered big data or not.
Velocity	The speed at which data is processed to meet the demands.
Variety	Various forms of data to be analysed according their types, data structures and different data formats.
Veracity	In this context, the challenge is to handle the quality of data.

The phenomenon has seen tremendous success over the past few years due to the exponential growth in the volumes of structured and unstructured data due to the advent of social media. Companies are realising the value this form of data holds in their optimising internal and external environment, and are putting it to different uses by experimenting with big data applications such as Hadoop, IBM Watson Analytics, R, and Python. As per the report from IDC, the big data and analytics market is expected to be valued at more than $203 billion in 2020, up

from $130 billion in 2016 with a compound annual growth rate (CAGR) of 11.7%. Every year, an increasing number of companies are realising the vast potential in big data in various spheres of operational management such as predicting user behaviour, refining product search, identifying patterns in supply chain issues, and optimising operational sustainability. According to a Gartner survey conducted in 2016 with the participation of 199 companies, about 48% of them invested in big data in the year of 2016, whereas in 2015, that number was only around 3%.

Of late, a few organizations such as UNSD, Global Pulse and INEGI have also started to pursue the sustainable development goals using big data and its applications. These analyses bring to the forefront the relevant statistics about the environment as well as the economy, which, in turn, helps them in decision making, while also making some efforts towards environmental development. Developed countries are usually the first to implement useful strategies in this case, while the developing countries lag behind due to obvious reasons such as, inferior technology, financial constraints, infrastructural requirements, requirement of better planning. Moreover many developing countries like India, Brazil, China and others, have been facing some serious environmental issues like air pollution, water scarcity, forest depletion, land degradation, biodiversity loss, overpopulation and others. With increasing population, the resources have been put under more stress over the recent years, calling for an urgent need and attention of concerned authorities to take necessary measures to prevent further depletion of natural and non-renewable resources. With the formulation of millennium development goals and sustainable development goals by individual nations, including developing ones, there has been an increasing effort in development of urban-rural metrics through big data innovations in achieving sustainability. Big data innovations and their deployment has enabled developing nations to address their rampant environmental challenges through environmentally models of economic development by addressing climate change issues as well as improving energy efficiency. For instance, while recommending strategies of utilizing Big Data in China, Cheng (2014) recommends practitioners to apply the concept in e-waste disposal, map poverty, enhance urban transport planning and identify pollution hotspots in cities. However, such implementation is not without challenges which include, data protection and privacy issues affecting fundamental rights of denizens. Another significant challenge lies in lack of coalition between adoption of big data and sustainability efforts by developing countries.

1.1 Aim of the Paper

The research paper aims to examine the ways in which big data can be used to improve sustainability, especially in developing countries, given the environmental issues degrading the living standard and human health in the concerned nations.

2. Features and Advantages of Big Data

Companies adopt Big Data because of its benefits and power, which come along with it, to process tons of data simultaneously. It comes with its own challenges and privacy concerns which are usually overlooked because of its large number of advantages. It benefits the companies with its fast parallel processing and analytics, helping them in making better and quicker decisions and which is one of the reasons of its huge adoption rate. Data to be processed has increased exponentially over a decade which used to be in gigabytes earlier, but now, it is reaching in range of petabytes. It is therefore impossible for traditional software to process those big amounts of data with ease and that's where big data analysis comes into play. Big Data processing is less expensive than common relational databases and has high throughout, owing to its parallel processing. Its storage is also less expense and big data analytics makes it easier to find valuable information in heaps of data with much reliability and quality. Various companies like Amazon, Netflix, Ebay and many others use big data as a tool to improve their customers' experience.

For instance, in 2013, EBay announced a big data-driven search engine, naming it as Cassini, for their e-commerce website– eBay.com which was a significant improvement over the basic search bar they had been using for years. Not only that improved users' experience, but it also resulted in improved business as almost every purchase on eBay starts with a search. A different company, Netflix, uses big data analysis for a different purpose. With over 93 million subscribers, Netflix uses analysis to improvise the experience of their subscribers by getting insights into their viewing habits. Based on those insights, shows of similar interests are recommended to their subscribers.

The main advantage of big data analytics is that it helps companies in making better decisions, even in real-time. The data can come from various sources such as social media, scientific applications, sensors, surveillance, and, video and image archives. Due to its parallel processing, data can be imported and processed in real-time, and insights can be given in real-time making it quicker for the companies to take necessary steps for improvisation.

Results incurred through the analysis are something which a company can rely on as it finds out all the relevant data with correlations and patterns between data pieces which would not have been possible without the big data software. Analysis can also bring more revenues to the company with the helpful insights it provides. Various types of expenses can be saved and services can be improved drastically, which in turn result in increase in company revenue. With the use of big data analytics, an organization can have better insights in sales, better monitoring of various factors, less use of other resources, insights of competitive offerings and promotions and many others, which would save company's expenses resulting in increase in revenue.

It goes without saying that there are also some challenges and privacy concerns faced by big data users and programmers, but given the heaps of features that big data brings, they are usually overlooked or taken care of to some extent.

3. Sustainability as a Growing Concern in Developing Countries

Environmental sustainability is, for sure, one of the major concerns in developing countries like India, Brazil, China, Korea and many others all over the globe. Just like the increase in usage of big data, the need for resources is also increasing proportionally with the increasing population. In general, environmental sustainability talks about maintaining the current natural resources without using them excessively. It has been a major concern for long enough compelling several organizations to jump into the pool to solve the sustainability problems using scientific technologies like big data analytics. The threat to the environment is that the resources which humans want to sustain may not last long enough such as clean water and air, good climate, quality of life. There is also a risk to the diminishing non-renewable resources which are limited and human beings have been using them excessively from decades. All that excessive usage cannot be reversed, so the least one can do is to sustain what currently is left for sustenance [23].
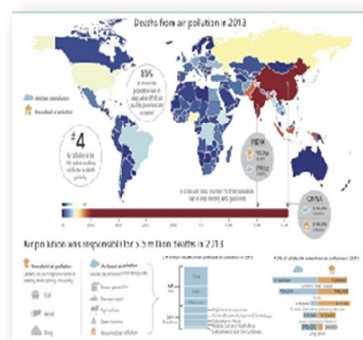

Figure 1: Projection of Air Pollution in the Developing Nations, especially India and China

Source: Watts (2016).

In developing countries like India, there has been a growing concern regarding the sustainability of the environment. In May 2014, New Delhi was declared as the most polluted city in the world by the World Health Organization, and then again in November 2016, New Delhi saw severe air pollution which lasted for few days and was considered worst of all in the last 17 years. Although the World Bank experts say that India has made one of the fastest progress in improvising its environment, there is still a long way to go to match the developed countries and India is still nowhere near them [14]. This issue is not uncommon in developing countries. Many other developing economies like Pakistan, Iraq, and Bangladesh face this issue.
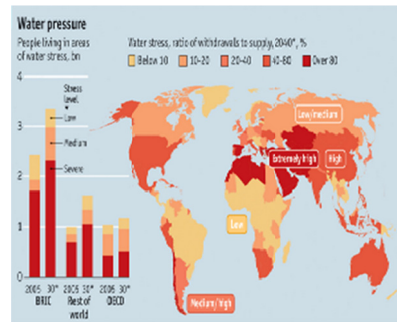
Figure 2. Projection of High Intensity Water Scarcity in the Developing Nations

Source: The Economist (2016).

Water scarcity is another issue in the developing economies, as seen in the above figure. According to a recent UN report, about 50% of people going through the water scarcity problems reside in India and China. The water resources in these countries are already stressed to an extent. Another recent report from UNICEF concluded that one out of every 4 children all over the globe will be living in water scarce areas by 2040 (United Nations 2017). Well, that's not all. The WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation estimated in 2014 that more than 1.8 billion people worldwide drink contaminated water on a daily basis. The number increases to a higher extent if the people, who get drinking water through means which are not protected again sanitary hazards, are to be considered.
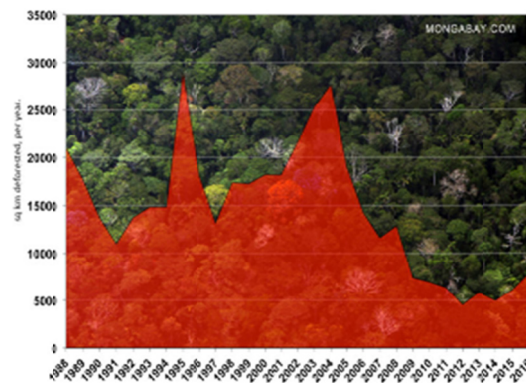

Figure 3. Projection of Deforestation in Amazon Basin of Brazil, 1988-2016

Source: Butler (2017)

Apart from India and China, other developing nations like Brazil have also been struggling through various environmental issues like deforestation in the Amazon basin, risk of extinction of various endangered species, illegal wildlife trade, land degradation, air and water population in large cities like Rio de Janeiro and Sao Paulo, and water pollution cause by various mining activities in the area. Among the other developing economies, South Korea is also one of those struggling with some environmental issues. South Korea has recently managed to be in top 150 countries to have severe air pollution with more than 50% of the nation's population exposed to dangerous levels of fine dust.

All this clearly indicate that there is a serious need to take necessary steps to conserve natural resources and prevent them further depletion before it is too late, only understanding the definition of "Environmental Sustainability" would not help.

3.1 Use of Big Data in Improving Sustainability

Big data not only helps companies in generating more revenues, but also in continuous improvement of environment through sustainable methods. Initial setup may be expensive, but in long term, it is going to benefit the company and the environment. According to a recent conference on "Sustainability in the Age of Big Data"

held by IGEL (the Initiative for Global Environmental Leadership), many companies have found that they can make big savings using big data analytics. Cities have reduced road congestion and traffic by optimizing the traffic signal timings, while companies have reduced energy costs by using big data analytics for better management, thus, reducing their overall expenses. With its fast parallel processing and the real-time sights, analytics can be used to deliver driving directions in real-time which also requires insights about traffic congestion in near-by areas, reports of traffic incidents, traffic camera records and similar insights. All this data can be used to find the best way to the destination for the user, hence, saving his time and fuel. Inrix– a SaaS and DaaS based company, does a similar job. They produce and provide various insights and useful information such as real-time data about traffic, speed and travel time data information about recent traffic collisions, parking data and more, provides real-time traffic information, state-wide traffic analytics, traffic collisions, parking data and analytics, connected car services, traffic count and population movement insights to various other companies which use those data insights to give directions their users. Microsoft's initiative "Microsoft Smart Campus" is another example on how big data can help in reducing energy costs. IBM has also started helping cities in reducing traffics, improving water managements and its collection with its own initiative "Smarter Cities" based on big data. Not only that, UN's Food and Agriculture Organization is also planning to increase the global food supply by up to 70% in the next 35 years to feed the population which is going hungry along with the increase in rate in that matter.

Another important use of big data in environment sustainability is in reduction of environment problems and risks with its real-time insights. Aqueduct, WRI's interactive water-risk mapping tool, carries the ability to calculate the overall water risk anywhere on the planet by generating insights about the quality and quantity of water. The tool can also provide insights about the changes going to take place in future in terms of water risk.

Not only for food, water and traffic, weather predictions and risk assessments are also being made using big data's real-time analytics by crunching tons of data instantaneously. Deep Thunder from IBM offers similar insights to various US-based companies. Deep Thunder can predict the locations where the public is going to face outages due to weather conditions. Thus, the company can take necessary steps before hand to prevent that or can send their staff to reduce the outage time by starting fixing it right on time. That saves cost for the company and optimizes its energy use.

The usability of big data towards environmental sustainability is not limited to the domains explained above. There are many other applications where big data is helping in keeping the environment healthy and sustained such as environmental research, risk assessment, public education, capacity building, and, monitoring progress.

3.2 Challenges in the Usage

Along with solving many problems for the organizations and the environment, big data also brings some challenges for the programmers and data scientists, data capacity being the major challenge among all. Since the data grows at exponential rate, it becomes a problem for the organization to keep and manage that old data. Such data is also supposed to be made public recognised as "open data". Growing data demands more and more infrastructure time-to-time, hence, the organizations should have some policies defining the life-cycle and expiration of old data to maintain manageability and operability.

There are many other problems that the organizations have to go through with big data applications. Big data is currently only for specific type of set-up and expertise to be able to understand and process it. For others, it may become almost impossible to understand the data. Thus, the data scientists and programmers have to work hard to fix that issue, but unfortunately, there is a dearth of skilled personnel such as data scientists, analysts and programmers to handle the data which is growing at such a rate creating another concern for the organizations to cope up with. The IGEL (Initiative for Global Environmental Leadership) conference held in 2014 noted an issue that transmitting data wirelessly is very expensive at the current stage. Although new scientific systems in future might reduce the costs, but as of now, it costs much for the companies to transmit huge amounts of data from one point to another.

Big data also poses some privacy and security concerns as mentioned above. With the rapid growth of big data, the security issues are also getting boosted. Instance can be seen from the report of BBC's in 2014, where, a Russian crime ring stole login credentials of around 1.2 billion usernames and passwords which belonged to more than 500 million email addresses by hacking 420,000 websites. Biggest data breach in the history took place in 2014 where over 1 billion user accounts were hacked in Yahoo. The company disclosed that their systems were hacked 2014 but they got to know about it in 2016 which means that the hackers had access to their databases for two years. This clearly indicated that the current security in big data applications is still inadequate. Hardware and software encryption, and round the monitoring are not enough in keeping the data

secure enough.

Since the application is same, these privacy and security concerns create flaws in the world of environmental sustainability also. Big data management authorities have to cope with more security challenges such as limiting its access to unauthorized users and hiring only trustworthy personnel to access the database. Secure software and hardware encryption technologies must be used to protect confidential data as traditional methods are not adequate to solve big data problems.

3.3 Empirical Review: Use of Big Data for Environmental Sustainability

The above exploration on cases of environmental sustainability in developing countries presented the urgent need of the hour of the Government, public and private enterprises in taking concrete steps, tackling the groping issue. In line with such requirement, various surveys and studies have been conducted so far regarding the usage of big data analytics in the field of environmental sustainability. Most of them noted similar challenges faced in the application with some different findings, while some came up with different challenges in this approach.

A survey was conducted by Breggin and Amselem (2014)on "Big Data and Environmental Protection: An Initial Survey of Public and Private Initiatives" surveyed many government entities, non-profit organizations and corporations such as US Environmental Protection Agency, US Geological Survey agency, IBM, Microsoft and many others about how they are using big data analytics to make a positive impact on the environment [3]. The survey resulted in a variety of purposes for which the organizations were using big data analytics such as IBM uses them to manage water and electrical transmission and distribution in a more efficient way whereas US EPA works on improvising the compliance with environmental regulations.

Keeso's, (2014) study on "Big Data and Environmental Sustainability" interviewed 14 organizations in various sectors with the purpose to find out how they employed and enabled big data to make the environment sustainable. The findings were that various organizations such as CI's Tropical Ecology Assessment Monitoring (TEAM) Network which works on producing real-time insights with HP like a warning system for the nature and also monitors biodiversity in 16 tropical forests, and others are making a great contribution in maintain the environment and making it more sustainable with the help of big data. However, the study also noted the different barriers being faced by the organizations in adopting big data for environmental sustainability such as lack of expertise (similar to other studies above), financial constraints, lack of data and inequality, security issues, cultural legacies, confidentiality, and lack of support and regulations. The barriers were mostly observed in NGOs and international institutes, while least in government entities and corporations.

A recent study "Big Data, Management, and Sustainability: Strategic Opportunities Ahead" conducted by Dror Etzion and Aragon-Correa in 2016) focused on the impact that big data can make on various processes in corporations and the opportunities it brings for researchers and for sustaining the environment. Many things were observed in the study such as the big data analytics can be used to conserve resources by monitoring parameters like markets, behaviours and availability, producing useful insights from wide variety of data from different sensors and seamless communication between them thereby bringing better efficiency and management and many other aspects like risk management.

Another study from 2016 was conducted by Ali Kharrazi, Hua Qin and Yi Zhang on "Urban Big Data and Sustainable Development Goals: Challenges and Opportunities". The study emphasized on making improvements in the current system to pursue Sustainable Development Goals (SDGs) with the use of big data.[13] The study suggested the creation of mayoral institutions for processing, transparency and better management of big data, spreading awareness to understand common skill sets and knowledge about big data and its management, and corporations making investments in research for big data and in bringing more expertise and skilled people to the corporation.

4. Conclusion

All the harm that has been done to the non-renewable natural resources cannot be reversed, but the necessary steps can be taken to prevent the further depletion of those resources. Different initiatives are being taken by a large number of organizations all over the world to make a positive impact on the environment. Most of them have adopted big data for this purpose, while small organizations, which cannot afford, have to rely on basic ways to do that and that is one of the barriers mentioned in a study above. Lack of expertise, along with the other challenges, also prevents the small organizations to adopt big data applications.

As observed in the findings from the discussion above, while big data is playing a major role in fixing various environmental issues such as early weather warnings, monitoring the progress, land degradation, preventing deforestation and others, small firms are still being left behind because of the barriers. Big corporations have

good income streams to adopt and analyse big data along with the required expertise. Thus more conferences and meetings should be held to give directions and support to the organizations which are not able to cope up with big data and its challenges. Addressing this issue would result in thousands of small firms all over the world to come forward and make a bigger contribution in making the environment a healthier place which is more necessary in developing countries.

Organizations in developing countries usually struggle in infrastructural requirements for storing big amounts of data and skills required to process it which makes them lag behind even more in comparison to the big corporations in the developed economies. New approaches should be experimented to evaluate the datasets, overcoming those challenges and addressing the environmental sustainability issue. Those organizations need to hire people with proper expertise to bring down the big data challenges and to compete with the organizations in developed countries. There is also a technology-divide between the developing and developed economies where the new technologies are launched in the already developed countries and others just keep waiting for them for years. Government should come forward to solve this issue and make new technology and scientific systems available in the developing nations without much delay for their own good, along with a reliable internet connection which does not seem to be there as of now.

There is also a need for the technology to evolve even more, globally, so that those large data streams can be transmitted wirelessly in a cost-effective manner as the current technology turns out to be expensive to the organizations resulting in increased revenues if they use the technology for small purposes.

Despite extensive exploration of secondary literature on big data application and challenges faced by developing nations in meeting environmental sustainability, the study is not without certain limitation. First being the lack of empirical investigation based on primary data on perception of various organisations' representatives of developing nations. Further, no representative from the government bodies of developing countries were approached to understand the strategies implemented at ground level. Lack of primary investigation disables the study to establishing observation of what is happening, in actuality. Future studies is therefore recommended to act of these lacunae and developed studies on grounded realities on the usage of and challenges associated with application of big data innovation in environmental sustainability.

Acknowledgements

References

BBC. (2014). Russia gang hacks 1.2 billion usernames and passwords.

Boyd, D., & Crawford, K. (201). Six Provocations for Big Data.

Breggin, L., Amselem, J. (2014). Big Data and Environmental Protection : An Initial Survey of Public and Private Initiatives.

Butler, R. (2017). Calculating Deforestation Figures for the Amazon. Rio de Janeiro.

Cheng, J. W. (2014). *Big Data for Development in China.* Washington: UNDP China Working Paper.

CIA. (2011). The World FactBook Environment - Current Issues.

eBay Inc. (2013). Big Data is Reshaping Commerce, as eBay Inc. Innovates.

Etzion, D., & Aragon-Correa, J. A. (2016). Big Data, Management, and Sustainability: Strategic Opportunities Ahead. https://doi.org/10.1177/1086026616650437.

Gartner. (2016). Gartner Survey Reveals Investment in Big Data Is Up but Fewer Organizations Plan to Invest.

IDC. (2016). Double-Digit Growth Forecast for the Worldwide Big Data and Business Analytics Market Through 2020 Led by Banking and Manufacturing Investments. 2016.

IGEL. (2014). Sustainability in the Age of Big Data.

Keeso, A. (2014). Big data and environmental sustainability: A conversation starter.

Kharrazi, A., Qin, H., & Zhang, Y. (2016). Urban Big Data and Sustainable Development Goals: Challenges and Opportunities. https://doi.org/10.3390/su8121293

Lomate, V. U. (2015). Environmental Problems at National Level and their Solutions.

Marr, B. (2015). Big Data: How Netflix Uses It to Drive Business Success.

Moura, J., & Serrão, C. (2016). Security and Privacy Issues of Big Data.

Nandi, J. (2016). Delhi no more the most polluted city in the world, says WHO report.

Neves, P. C., Schmerl, B., Camara, J., & Bernardino, J. (2016). Big data in cloud computing: Features and issues.

PTI, United Nations. (2017). Half of people facing water scarcity live in India, China.

Rijmenam, M. V. (2016). The Advantages and Disadvantages Of Real-Time Big Data Analytics.

Sikos, L. F. (2015). Big Data Applications. In: Mastering Structured Data on the Semantic Web. https://doi.org/10.1007/978-1-4842-1049-9_8

Stockton, N. (2016). Deep Thunder Can Forecast the Weather—Down to a City Block.

Sutton, P. (2004). A Perspective on Environmental Sustainability.

The Economist. (2016). Water scarcity: Liquidity crisis.

The Korea Herald. (2016). South Korea near bottom of world survey of air quality.

Thielman, S. (2016). Yahoo hack: 1bn accounts compromised by biggest data breach in history.

UN News Centre. (2013). World must sustainably produce 70 per cent more food by mid-century-UN report.

United Nations GWG. (2014). Using Big Data for the Sustainable Development Goals.

United Nations. (2014). Water quality report.

Watson, H. J. (2014). Tutorial: Big Data Analytics: Concepts, Technologies, and Applications. http://aisel.aisnet.org/cais/vol34/iss1/65

Watts, A. (2016). Poor air quality kills 5.5 million worldwide annually.

White, J. B. (2008). New Services Gather Data In an Effort to Track Current And Future Traffic Jams.

WRI. (2013). Measuring, mapping and understanding water risks around the globe.