The impact of big data analysis in enhancing sustainable development goals

Basma Kajruba[1], Guma Lakshen[2], Jelena Velickovic[1]

[1]Faculty of Applied Ecology Futura, University of Metropolitan, Belgrade, Serbia
[2]Faculty of Electrical Engineering, University of Belgrade, Belgrade, Serbia

Abstract

Digital technologies witnessed rapid advancements in many domains such as computing, programming, cloud computing, data storing, networking, satellite broadcasting, mobile technologies, but increasingly also in the natural sciences, such as medicine, biology, chemistry, pharmacy, etc. These advancements lead to a massive explosion in data volumes which expected to exceed 175ZB (175×10^{21} Byte) by end of the year 2025. When data size exceeds normal manageable storable sizes, it will be called “big data”. Organizing and analyzing this “big data” with appropriate tools would produce quite valuable results helping decisions and policymakers.

Big data also can assist significantly improve the life quality of much of the world’s population. The United Nations, governments, non-profitable and other organizations are utilizing big data to assist achieving the UN’s Sustainable Development Goals (SDG), a set of 17 targets pertinent to protect the natural environment, fighting hunger and poverty, reducing inequality, improving health outcomes and other things that will make life better around the world.

Big data analysis could be used in many ways to enhance the understanding of the progress towards the SDGs, specify how best to meet SDGs targets, and ensure accountability. The UN has set up a task team to explore how to use big data to help achieve the SDGs. A survey by the task team found that big data projects most frequently focused on the “no poverty” goal, it was found that mobile phone data was the most common data source.

Keywords: Big data, Big data analysis, tools, sustainable development, and 2030 SDGs

1. Introduction

The data produced by new emerging technologies such as the internet, mobile technologies, and cloud computing have experienced a massive increase in data volumes which has been identified as ‘big data’. Big data is quite a buzzword, but its definition is relatively obvious, it refers to any high-volume data that gets collected regularly or embrace a wide variety of topics[1]. It can be defined as “A collection of big and intricate data sets that can be only processed with difficulty using on-hand database management tools” [1]. Big data nowadays is everywhere, and all sorts of businesses, non-profits, governments, and other groups use it to

[1] Big data, big results: The IT buzzword that, we must grudgingly admit, actually works | Ars Technica

extract value for it to improve their understanding of certain topics and improve their practices.

Marketing teams use big data to study and learn more about their customer base and behaviours, healthcare data analysts may use big data to figure out the chances of a person contracting the disease and metropolitan police may utilize big data to optimize traffic flow, and it can also be quite beneficial in wildlife savings[2]. In the last few years, big data play a significant role in science, especially in genetics analysis in many areas, such as biology, ecology, medicine, pharmacology, and microbiology.

In the last few months, in a global coronary virus pandemic, great importance is placed on scientists to understand this virus. One of the main challenges is to sequence the genome of the virus and these days it is mostly done using new genetic technology, known as the Next Generation Sequencing NGS which read the genome of the virus and collects huge data [2]. This big data is a great challenge for many professions, both biological and computer.

Big data also can assist significantly improve the life quality of much of the world's population. The UN, governments, non-profitable and other groups are utilizing big data to assist achieving the UN's Sustainable Development Goals (SDG), a set of 17 targets pertinent to protect the natural environment, reducing inequality, improving health outcomes and other things that will make life better around the world[3].

Big data could be used in many ways to enhance the understanding of the progress towards the SDGs, specify how best to meet SDGs targets, and ensure accountability. The United Nations (UN) has set up a task team to explore how to use big data to help achieve the SDGs[4]. A survey by the task team found that big data projects most frequently focused on the "no poverty" goal, it was found that mobile phone data was the most common data source. There are many possible uses for big data related to the SDGs. Mobile phone data, for instance, could be used to track the movement of populations, such as refugees, to improve proper and suitable preparation procedures. Data analysis could help predict, speculate, and forecast changes in food prices. The list of possibilities of big data analytics is growing at a rapid rate and are virtually endless.

Prospects related to big data analysis utilization are abundant, but there exists also several challenges and risks. Collecting, cleaning, and analyzing big data in medicine and genetic analysis is very important in the sense of not losing data that might have crucial clinical importance for understanding the disease, virus infection, or health of the patient, emphasizing again the enormous importance of adequate maneuvering with big data. It requires advanced technology and infrastructure to overcome it, which can be expensive especially to poorer

---

[2] Best Uses of Big Data in Marketing | Cleverism

[3] Sustainable Development Goals | UNDP

[4] How Can We Use Big Data to Achieve SDGs? | BioEnergy Consult

countries, and self-supporting organizations, which limits their chances of utilizing this technology. In the survey by the UN's big data task team, the team received much higher response rates from high-income countries than lower-income ones.

The UN, through several of its groups, issued recommendations and guidelines[5] for the use of big data related to SDGs. Among the goals of these guidelines is ensuring privacy and increasing access to data globally. The public and private sectors, as well as countries and organizations from around the world, will have to work together to accomplish the UN's SDGs and to ensure best practices and take full advantage of the benefits big data can provide related to achieving them.

In this paper, the authors will emphasize the importance of big data and big data analysis in many areas and professions with all its advantages, but also certain risks and challenges. Additionally, the authors will point out big data analysis tools that enhance greater impacts on achieving many goals and also to introduce a framework that can be implemented in the developing countries in achieving the best uses of big data in enhancing SDGs.

2. Problem statement

The Sustainable Development Goals (SDGs) emphasize the need to monitor each goal through objective targets and indicators based on common denominators in the ability of countries to collect and maintain relevant standardized data. The emergence of big data analysis capabilities has further increased the value of big data that is generated by social networks utilizing GPS, real-time, sensors, Satellite and unmanned aerial vehicles (UAVs) imagery, Web scraping, internet of things, mobile, artificial intelligence, and cloud computing technologies have led to the concept of big social networking. Millions of people can instantly share contents including sports, economy, politics, and the environment is extremely crucial in terms of the comprehensiveness of information. These data groups, which indicate the trends of large groups of people, can be analyzed through big data facilities. The results of this analysis can be used as a new tool to develop effective sustainable investment policies and policies and strategies in areas such as environment, health, and women's rights.

3. Big Data overview

Data in the Web is growing at a tremendous rate according to [3,4,5]; this data represents 2.5 quintillion bytes (Exabyte (EB)= 10^{18} bytes). In 2000, more than 800,000 Petabyte (1 PB= 10^{15} bytes) of data were stored in the Web. By the end of 2019, this volume is expected to reach 35 Zettabytes (1 ZB= 10^{21} bytes) [6], and is also expected to grow 61% and exceeds 175 zettabytes by 20^{25} as per International Data Corporation (IDC) expectations [7]. The 3rd quarter of 2019, showed that 4.33 billion active internet users [8], this represents 8.2% growth in active internet users globally, this translates to 59% of the world population are

[5] Big Data for Sustainable Development | United Nations

online, which is 8 times faster than the world population growth (stands at 1%), and there are today more than 1.7 billion websites [9]. Public attention and literature awareness of the terms *"big data"* has grown tremendously in the last decades, for example, a generic search on a Google search for "big data" resulted in more than 145 million results. According to IDC, every person online will create an average of 1.7 megabytes of new data every second by 2020, and only 37% of all big data could be analyzed, leaving a plethora of untapped information that could be processed and gain value out of it.

McKinsey Global Institute[6], defined big data as *"datasets whose size is exceeding the capacity of classical database software tools to capture, store, manage, and analyze" [11]*. *Douglas Laney* in 2001, envisioned the future changes relating the expanding size of data, through his definition of data by using a three-dimensional view as: *"Big data is high volume, high velocity, and/or high variety information assets that demand new ways of processing to enable enhancing decision making, insight discovery, and process optimization"* [12] (figure 1). *Loukides* emphasized on data volume by defining big data as *"when the size of the data itself becomes part of the problem and traditional techniques for working with data run out of steam" [13]*. *Stonebraker* in 2012, defined big data as *"Big data can mean big volume, big velocity, or big variety" [14]*. Various additional definitions of big data have been elaborated over the last decade [1].

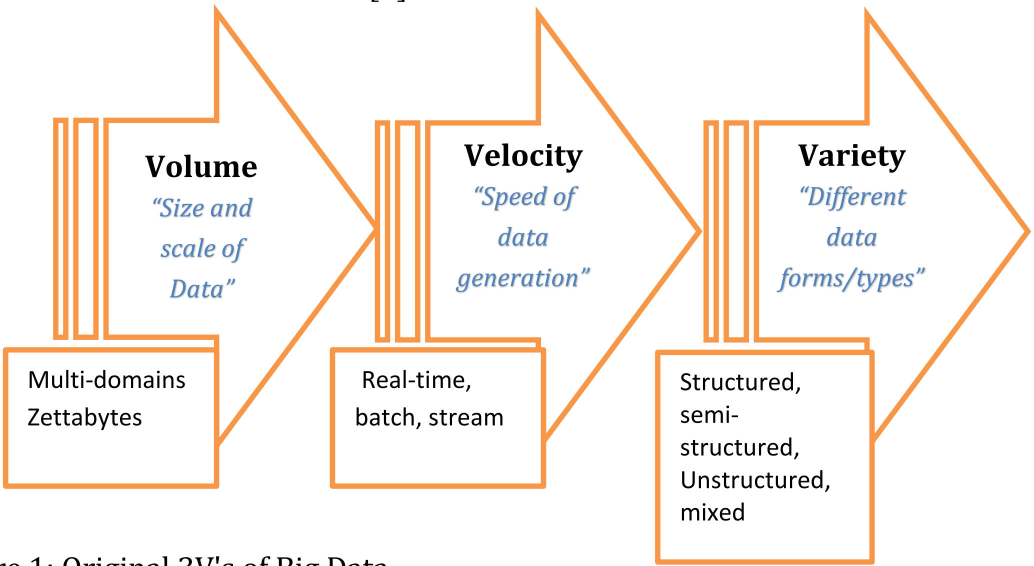

Figure 1: Original 3V's of Big Data

The above 3V's of big data challenge urges the need for innovative ways of data processing to enable enhanced insight discovery, decision-making, process optimization, and data visualization. As the field of big data developed, additional V's of data have been added over the years. Table 1 gives a brief description of the most important V's of big data.

[6] https://www.mckinsey.com/

Table 1: Short Description of the 10 V's of big data

Volume	measures the *amount* of data available to an organization, which does not necessarily have to own all of it as long as it can access it
Velocity	measures the *speed* of data creation, streaming, and aggregation
Verity	measures the *richness* of the data representation – text, images, video, audio, etc.
Veracity	measures the *understandability* of the data – biases, noise, abnormality, etc.
Variability	measures the *inconsistencies* of the data.
Validity	measures the accuracy and correctness of the data is for its intended use
Vulnerability	refers to new *security concerns* of big data
Volatility	refers to how long is data *valid* and how long should it be *stored*
Visualization	refers to the challenges of how big *data visualized*
Value	measures the *usefulness* of data in making decisions

4. Big data analysis

Big data analytics refers to the process of gathering and analyzing large volumes of data from various sources, then organizing it in a meaningful and understandable way to create big datasets to explore meaningful facts. Big data analysis, in addition to exploring facts and figures of information in the, gathered big data, also it categorizes data or data ranking concerning significances of the provided information. Big data and big data analysis tools have gained reiterated and significant roles as tools for realizing sustainable development.

The appearance of big data analysis capabilities has additionally raised big data value. Millions of people can instantly share the web content including economy, sports, environment, and politics are highly decisive in terms of the holistic of information. These data groups, which demonstrates the directions of substantial groups of people, can be analyzed through big data analytical tools.

As an illustrative example, Twitter as a social media organization processes approximately 10000 tweets per second before publishing them for the public to ensure that every tweet is decent and filtering out restricted words from tweets. The analyzing process carried out in real-time to avert delays in publishing twits lives for the public. Therefore, big data analytics requires unique platforms, tools, and services that reduce time and can offer distributed and scalable solutions, such as those included in the Apache Hadoop ecosystem [23,24]. Large amounts of different types of data have been generated at an accelerating rate that coincides with the invention of computers. Advances technologies such as mobile devices, digital sensors, communications, computing, and storage have provided means to collect data [25]. This situation motivates researches since then to dig in deep into this phenomenon seeking solutions for the expanding challenges facing it.

5. Big data analytics and sustainable development

Big data is a popular field aimed at providing alternatives to conventional solutions based on databases and data analysis. In addition to providing data storage and data access, big data aims to perform analysis to produce required solutions, comprehend data, and realize its value. Owing to sustainable development is a critical issue and analyzing big data was considerably helpful in different studies. Recently, many researchers attempted to measure the extent to which Big data analysis could increase sustainability, also, innovations in Information Technology (IT) services has been discussed in [2].

Cuquet and Fensel stated that using big data greatly assisted the economic, social and ethical, legal and political benefits in Europe which can apply to other countries as well [3]. Strong and reliable models are required to use big data analysis to forecast different situations. Predictive analytics capabilities have a significant effect on the environment, social and economic performance of the supply chain [4]. *Papadopoulos et al.* developed a theoretical framework to assess resilience in the supply chain with the use of big data [5].

New technologies have resulted in a high-level increase in current data volume and types, and have created unprecedented possibilities for informing and transforming society, and for protecting the environment. Corporations, researchers, and citizens are attempting to become familiar with the new data world via experiences and innovations. This change is referred to as the data revolution [4]. Before the great data revolution, companies could not store all of their archives for long periods and effectively manage large data sets. Conventional technologies have limited storage capacities and inflexible management tools, which are considerably expensive. Their lack of scalability, flexibility, and performance has been resolved through new technologies in the context of big data [5]

6. Existing Big Data Analysis Tools

Big data contains large amounts of multidimensional unstructured, semi-structured, and structured data which cannot be processed using classic methods anymore. The most popular technology used to analyze and process big data today is Hadoop [26], an open-source software. Hadoop can process big data fast and free of charge. Many distributors, such as Cloudera [27] and Hortonworks [28] provide big data platform services by using the Hadoop framework. These distributors provide a simplified Hadoop experience to users by combining various Hadoop components.

The Hadoop project contains many separate sub-projects. Big data can be processed with considerable ease under these different projects, and different conclusions can be drawn. The Hadoop project contains the following main modules (see figure 2):

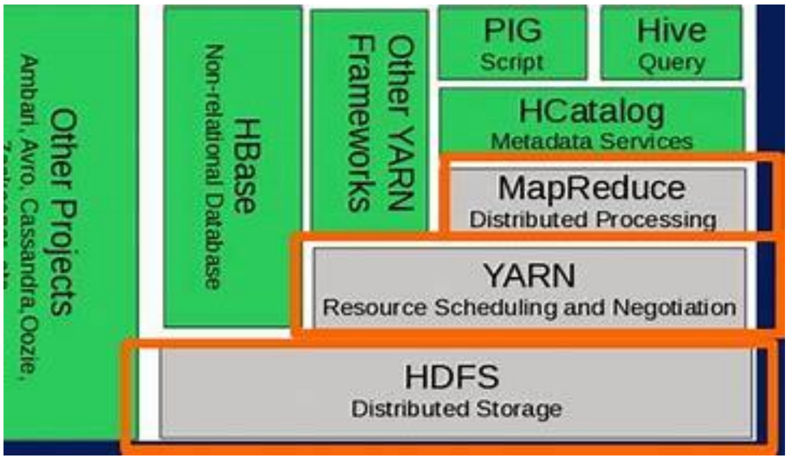

Figure 2: Hadoop main modules

- *Hadoop Common:* Common tools that support other Hadoop modules [26].
- *Hadoop Distributed File System (HDFS):* A distributed file system providing high-efficiency access to application data [26,29].
- *Hadoop YARN:* A framework for business planning and cluster source management [26].
- *Hadoop MapReduce:* A YARN-based parallel processing and programming model used for processing big datasets and retrieving information from them [26,30].

additional big data analytics tools:

- *Ambari:* A web-based tool containing support for Hadoop HDFS, Hadoop MapReduce, Hive, HCatalog, HBase, ZooKeeper, Oozie, Pig, and Sqoop projects and for enabling one to monitor and manage Apache Hadoop clusters. Ambari also provides a dashboard to monitor heat maps and applications like MapReduce, Pig, and Hive and to display the cluster efficiency [26].
- Avro: A data serialization system [26].
- *Cassandra*: A distributed, open source, unrelated, column-oriented database that was developed by Facebook to store a vast amount of structured data [31].
- *Chukwa*: A data collection system used for managing big distributed systems [36].
- *HBase*: A scalable and distributed database supporting structural data storage for big tables. Google's BigTable is an important Apache-Hadoop-based project that was developed recently and modelled using HBase. Hbase adds a fault-tolerant scalable database that is installed and distributed on an HDFS file system and that has random real-time read/write access [26,32].

- *Hive*: A data warehouse built on Hadoop that enables summarizing, querying, and analyzing big data clusters stored in Hadoop files. It is not designed to offer real-time queries; however, it may support text files and queue files [33].
- *Mahout*: A scalable machine learning and data mining library. Mahout currently focuses on algorithms for clustering, classification, data mining (frequent item set), and evolutionary programming [26,34].
- *Pig*: Provides a high-level parallel mechanism to execute MapReduce works on Hadoop clusters. It uses a command file language called Pig Latin; the data streaming language is collaterally directed to data processing [33].
- *Spark*: A fast and general calculation engine for Hadoop data. It provides a simple and effective programming model that supports a wide range of applications such as Spark, ETL, machine learning, stream processing, and graph calculation [36].
- *TEZ*: A generalized data stream programming framework installed on Hadoop YARN that provides a powerful and flexible engine to execute the directed acyclic gap (DAG) of tasks to process data for collective and interactive usage [36,35].
- *ZooKeeper*: Provides a high-performance coordination service for distributed applications [26].

7. Big Social Network Data and Sustainable Development

The widespread use of the internet is one of the most important new technological developments of the century. Later, online social networks were created which greatly impacted sustainable development and are also one of the main sources responsible for producing big data, which is an important economic parameter.

This paper focuses on the large volumes of data (i.e., big data) produced by governmental agencies and social networks, how new technological resources can be utilized to measure the vast majority's opinions, and how this would influence sustainable development. Big data can provide very valuable information once it is analyzed. In fact, in 2012, attendees of the World Economic Forum held in Davos, Switzerland[7], declared big data to be a strategic economic resource, which is as important as money, oil, and gold [36,37].

Social network data plays an important role in providing a clearer and more up-to-date picture of the world, planning required policies and programs together, monitoring and assessing these programs, and evaluating the processes of sharing resources that could affect people's lives and influence political decision-making [28]. Social networks are important in a lot of sustainability issues, such as information transfer, cooperation on management of shared resources, and the formulation of policies aimed at influencing various behaviours [36]. There exist

7 2012, Attendees Of The World Economic Forum Held In Davos, Switzerland

many social network analysis methods that can be utilized efficiently. Methods used for network analyses, such as anomaly detection [36], discrimination discovery [37], opinion leaders' detection [38], event detection [39], role mining [40], rumor propagation detection system [41], conflict detection [42], and topic detection [43], can also greatly contribute in the field of sustainable development.

Many companies actively use social network platforms. The statistical graph in Figure 3 presents leading social media platforms used worldwide as of January 2020.

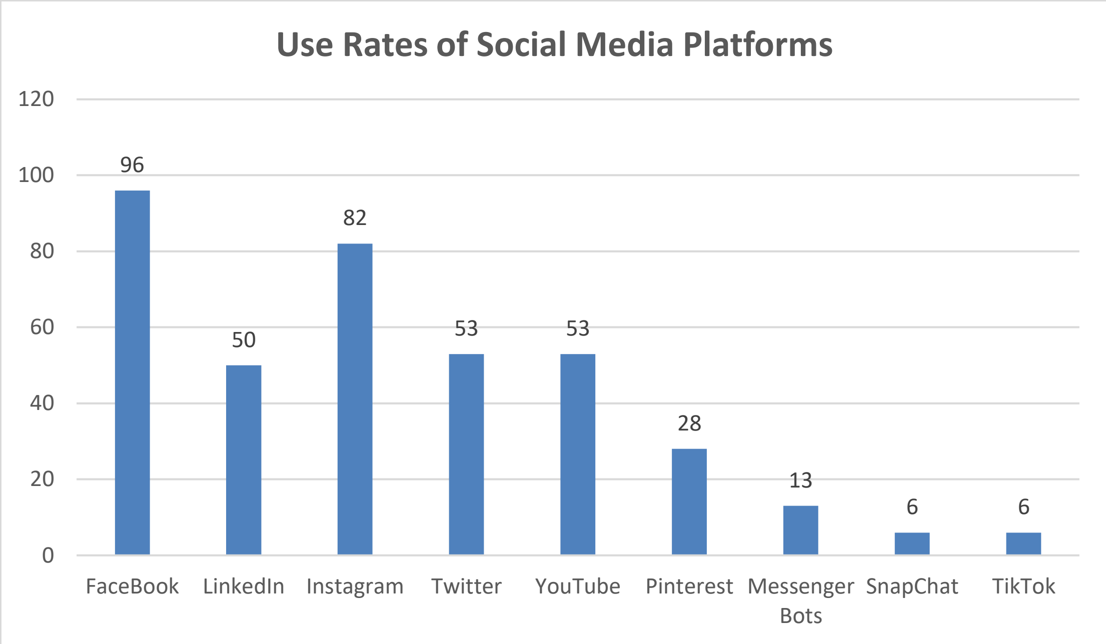

Figure 3: Leading social media platforms used worldwide as of January 2020. Source: Social media platforms used by B2B and B2C marketers worldwide 2020 | Statista

As the above figure shows, Facebook is the most important social platform for its establishments [7]. These data show the significance and inclusion of the social platforms and the big data they produce as they enable establishments to rapidly reach more than 2 billion people. This widespread use of social networks may lead to the use of more efficient technologies that are appropriate for sustainable development in various sectors. The efficient utilization of data in the social network will boost productivity in many sectors and participate in sustainable development. This should have a great impact on economic productivity which means that the efficient utilization of big social network data may influence and support developmental goals 7–10 specified in the sustainable goals for 2030.

8. Sustainable Development as a goal

The perception of sustainable development has been suggested as a substitute to the economic process in many studies and venues that started with the industrial revolution and still exists today, that prohibits long-term environmental development, and only depending on the economic advantages earned while processing crude materials into finished goods. The sustainable development concept started to be accepted in the 70s and 80s as was discussed in many international meetings and venues. The concept was first coined by the Limits of Growth report[8], it was the first to prove the paradox between unlimited and uncontrolled growth and the world's limited resources, and it accentuated the choice for the society to establish a sustainable development process that would be coherent with environmental limitations [44].

Then, in June 1972, the United Nations Conference on the Human Environment held in Stockholm[9], that resulted in the development of the concepts of sustainability and sustainable development. The fears about the protection and optimization of the ecological environment populated by humans and the portability of the environment to the future generations were on the agenda.

Therefore, they constituted a fundamental for the development of the concepts of sustainability and sustainable development [45]. The notions of "sustainability of the world system in a balanced manner" and "eco-development", which were submitted to the agenda by Dennis and Donella Meadows, were embedded initially in an official conceptual framework in the "Our Common Future" report published in 1987 by the World Commission on Environment and Development (also known as the Brundtland Commission) [46]. There exist several definitions of sustainable development, but the most frequently used definition is "a development model that meets the requirements of the present without compromising the ability of next generations to satisfy their requirements" [46].

Most of the international organizations that administer the global economic and political processes have started utilizing the concept of sustainable development. The UN Conference on Environment and Development (known as the Earth Summit[10]), which was held in Rio de Janeiro in June 1992, used the conceptual context of Brundtland Commission's report[11] for sustainability and sustainable development. The Earth Summit is substantial because five points deal with the international consensus on the concepts of sustainability and sustainable development was announced [47]. Sustainable development gained global importance and became a significant international phenomenon since the Rio summit [48].

Currently, the 2030 sustainable development goals [49] constitute the bulk updated goals of sustainable development. Although the goals cover different themes, they also constitute the three dimensions that are typically recognized

---

[8] The Limits to Growth

[9] United Nations Conference on the Human Environment held in Stockholm

[10] United Nations Conference on Environment and Development, Rio de Janeiro, Brazil, 3-14 June 1992 | United Nations

[11] Brundtland Commission's report

since the international community has begun using and discussing sustainable development [50]:

- ***The economical dimension:*** An economically sustainable system must have the ability to produce goods and services based on continuing tenets; the economy should guide the government and debts and block disruptions between the sectors that would harm industrial and agricultural production.
- ***The environmental dimension:*** The environmentally sustainable system must intake resources that are adequately sedentary, so that the base of the resources remains, blocking exploitation of renewable resource systems or environmental investment tasks, and only use non-renewable resources. Additionally, this operation must encompass the biodiversity conservation, atmospheric equilibrium
- ***The social dimension:*** A socially sustainable system should ensure a balanced and fair distribution of health, education, gender equality, political responsibility, participation, and the sufficient delivery of social services.

These above dimensions are closely related to each other and consistently affecting each other. Thus, the concept of sustainable development does not demand economic success only. It is essential to concurrently examine the "economic", "environmental", and "social" constituents for sustainable development. The goals described in the above dimensions of sustainable development were assigned as new and updated targets at the United Nations meeting in 2015 for the international community.

Recently, the UN Sustainable Development Summit held on 24–25 September 2019[12], the goals of sustainable development for 2030 were accepted with the signatures of 193 countries [49]. Sustainable development must be assessed to the context of the following goals, and policies should be produced towards the direction of these goals.

The shared indicators and statistical frameworks of the SDGs tend to assist organizations to measure the fulfilment of a specific goal locally or compared to others. Similarly, it is understood that for a particular government to outline and observe the impact of its policies, it should benchmark its data and measure the progress on yearly basis on particular development indicators. Based on this understanding, Sustainable Development Goals (SDGs) have been structured upon the Millennium Development Goals (MDGs) and been adopted on September 25, 2015.

The emphasis on scientifically grounded indicators for each SDGs is regarded as one of the most significant dimensions of the SDGs that can assist to monitor progress, implement strategies, allocate resources, and increase the

[12] SDG Summit | UNITED NATIONS ECONOMIC and SOCIAL COUNCIL

accountability of stakeholders involved. With this fundamental motivation, SDGs includes 17 goals covering 169 targets to be achieved by 2030 for achieving sustainable development. The 17 SDGs (see table 3) seek to exterminate all forms of poverty, fight inequalities, tackle climate change, and address a range of social needs like education, health, social protection, and job opportunities by 2030. The UNDP specified SDGs as “a universal call to action to end poverty, protect the planet, and ensure that all people enjoy peace and prosperity”[13] and all the participating countries are committed to achieving the same goals.

9. Utilizing Big Data Analytics to enhance governmental decisions

Achieving a decent quality of life and a fair happiness index for any country is one of the important goals for any government delivery agency targets. Accomplishing this goal, the utilization of real-time data is a must. Data analytics could be used for aiding decision-makers in the ever-changing and dynamic environment. Massive data continuously being generated by diverse sources such as social media, sensors, RFID devices, government departments and so on. This big data can be very beneficial and supportive to governmental decision-makers If it is correctly analyzed via a variety of techniques of big data analytics, therefore it can provide substantial support (Figure 4).

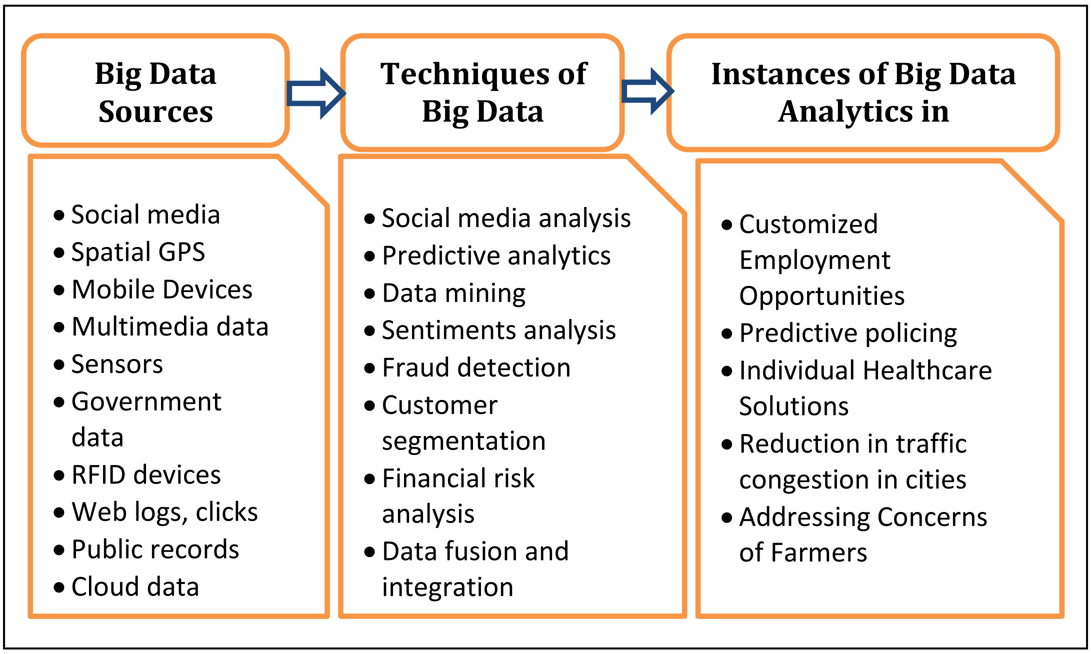

Figure 4: Big Data Analytics related to Governance

People nowadays are more claimants and more linked than ever before as a result of the widespread of numerous interfaces available via arising technologies. Big Data helps to ‘gather’ this data and ‘Data analytics’ helps to ‘mine’ this data. The application of big data analytics could be utilized in many ways such as the steady stream of the collated baseline data from the grassroots can help design

[13] UNDP - United Nations Development Programme

customized public services and also for strategizing future grassroots innovations in the economies. The baseline data, therefore, collected from the governance tasks undertaken at the grassroots when combined with unstructured data created by the user whilst using the Internet, mobile Social, Media, cloud etc., can later assist in not just continuous effect assessment of these tasks but can also help to reveal hidden patterns, unknown correlations, citizens' preferences, and other actionable insights that help public organizations make more-informed evidence-based governance decisions.

Therefore, it is inevitable that the citizens contextualized aspirations in a community of a constituency in a country may be mapped successfully by the delivery/governance agencies, by application of big data analytics. Thus, keeping insight the SDG targets, the cautious application of Big Data analytics can serve as the basement of sustainable future efforts to create need-based public services and tailored governance models, and further, tempt indigenous innovations to satisfy the specific localized needs.

The UN is already recommending the necessity for partnerships to influence the potential of the 'data ecosystem', including big data applications through the sectors, in a totalitarian way. Successful accomplishment of SDGs by 2030 would demand leveraging arising technologies, including Big Data Analytics, for backing the relevant processes of governance, some of which are evident in this paper (Table 2), only as an illustration and not in any way as a comprehensive list.

Table 2: The UN 2030 Sustainable Development Goals

SDG No.	SDG Goal	Logo	Brief Description	Suggested Application of Big Data Analytics
1	No Poverty		Ending all kinds of poverty, no matter where they are.	Creating poverty maps[14] by Mobile Call Data Records (MCDRs) giving an insight into the communication and mobility patterns of the affected population.
2	Zero Hunger		Ending hunger, providing food safety, developing nutritional resources, and supporting sustainable agriculture.	Through precision agriculture[15] , by using Big Data Analytics on the data collected from various sources e.g. such as aerial images, embedded local sensors, weather departments, etc.
3	Good Health and Well-being		Ensuring a healthy life for all people and prosperity for	Using mobile phone data mobility patterns to forecast and suppress the

[14] Big Data for improved diagnosis of poverty: A case study of Senegal (brookings.edu); Retrieved on December 15, 2020

[15] https://cio.economictimes.indiatimes.com/dobig/news/detail/1623?redirect=1 ; Retrieved on July 26, 2017

			everybody, at every age.	proliferation of infectious diseases
4	Quality education		Providing everybody with equally quality education and the possibility of life-long education.	Analyzing the educational sector's big data can expose various defects in the education system
5	Gender Equality		Providing gender equality in society and strengthening the social status of women and girls.	Financial transactions analysis can detect spending patterns of men and women; further, these patterns can reveal the differences in financial and economic well-being, independence, growth, etc.
6	Clean water and sanitation		Ensuring the availability and sustainable management of water and sanitation for all	Sensors connected to water pumps can track access to clean water.
7	Affordable and clean energy		Providing accessible, reliable, sustainable, and modern energy to everyone	Smart metering permits utility companies to increase or restrict the flow of electricity, gas or water to minimize waste and guarantee sufficient supply at peak periods.
8	Decent work and economy growth		Ensuring sustainable and inclusive economic development and ensuring full and productive employment and decent jobs for human dignity.	Patterns generated via analysis of Internet-of-Things based solutions could provide indicators such as economic growth, transfers, commerce and GDP
9	Industry, innovation and infrastructure		Constructing durable infrastructure and encouraging sustainable and inclusive industrialization and new inventions	GPS Data devices could be used to control traffic and enhance public transport
10	Reduce inequality		Reducing inequalities in and between countries	Speech-to-text analytics may uncover discrimination concerns and support policy response
11	Sustainable cities and communities		Making cities and human settlements strong, secure, and sustainable.	Encroachment on public lands, forests, parks can be tracked through Satellite remote sensing
12	Response consumption and production		Providing sustainable consumption and production.	Online search patterns or e-commerce transactions can disclose the transition frequency to energy-efficient products
13	Climate action		Taking emergent steps to address climate change and its impacts.	Combining satellite imagery, crowd-sourced witness accounts and open data can

				help track deforestation and reduce encroachment
14	Life below water		Protecting oceans, seas, and marine resources for sustainable development and sustainably using them.	Maritime vessel tracking data can reveal illegal, unregulated and unreported fishing activities
15	Life on Land		Preserving and restoring terrestrial ecosystems, ensuring their sustainable use, and addressing desertification.	Social media monitoring could support calamities management with real-time information on the victim's location
16	Peace, Justice and strong institutions		Encouraging peaceful and embracing communities for sustainable development, ensuring access to justice for everyone, and establishing effective, accountable, and embracing institutions at every level	Sentiment analysis of social media may reveal public opinion on efficient governance, public service delivery or human rights issues
17	Partnerships For the goals		Strengthening the application tools of a global partnership for sustainable development and reviving global partnership.	Partnerships to empower the combining of statistics, mobile and internet data can offer a better and real-time comprehension of today's hyperconnected world

The assignment of SDGs and big data analytics pinpoints the need to demystify technological possibilities, potential avenues, policy opportunities, and related challenges that may appear while involving with techniques and technologies pertinent to big data and big data analytics at the local level. To identify these opportunities and challenges, many innovative models, frameworks, and fresh possibilities required to be comprehended to fix systemic governance issues, to realize sustainable development.

10. Case Study: Sustainable development and goals in Libya

The previous sections presented a glimpse of applications of big data and big data analysis to address the challenges of SDGs, Libya requires to assess its strategic opportunities for using big data to attain the SDG 2030 agenda[16]. Libya is currently suffering from loads of problems starting from the political instability and the economical chaos as well as the internal fighting all this have led to the social unrest and the continues fall of index of well beings of the people. Some Libyan statistic compared with some Arabic countries are presented in table 3. Due to the divided political and governmental organization and agencies with the country

[16] ليبيا والتقرير الدوري عن أهداف التنمية المستدامة في قارة أفريقيا للعام 2019 "الجزء الثاني" – صدى (sada.ly)

and monetary institutions, the concentrations of development and sustainable development has been very weak except some guideline put by the ministry of planning, unfortunately, it will take a considerable time before it will be put into practice due to the reasons mentioned above. Nevertheless, Libyan people have the well and the resource to stand for their country once this current chaos is over.

Table 3: Different statistics about Libya and the 17 SDG goals
[Source: https://hunalibya.com/local-affairs/10369/]

No	Category	Libya compared to some Arabic countries
1	Ranking	14/21 position according to the Arabic SDGs goal index
2	Employee's poverty	10.4% compared to 2.2% in Tunis and 2.0% in Algeria
3	Stunting	21% for children under the age of 5 years
4	Obesity	32.5% compared to 35.1% in Qatar and 29.8% in Bahrain
5	Suicide	5.5/100000 person compared to 5.8/100000 in Qatar and 4.4% in Egypt
6	Pre-school education (age 5 years)	10%
7	Education	Average study years female/male from age more than 25
8	Political equality (female/male)	Parliamentary seats 16%, Ministerial positions 3.6%
9	Water sanitation in human activity	9.6%
10	Electricity	98%
11	Scientific and technical journal articles and periodicals	0% to 1000 person
12	Unemployment	41.9% for youth 15-24 years
13	Internet usage	21.8% Libya, 34.3% Syria, 30.9% Sudan, and 26.7% Yemen
14	Political stability and no terrorist environment	Libya -2.3%, Tunis -1.1%, and Algeria -1.0%
15	Media freedom (0-100 worst)	Libya 56.8, Egypt 56.7, and Tunis 30.9

4. Concluding Remarks

Big data and its tools are not the golden keys to all problems, there are various challenges and risks implicated in employing big data techniques/tools in governance and decision-making process and these challenges need to be acknowledged and processed through big data strategies. The most common type of challenges is infrastructure concerns, data deficit, data governance, reliability, quality, privacy, data anonymity, data security, rightful ownership of data, data-price/value/cost for the people as well as for the user-agency and so on. Data privacy emerges as one of the main considerations, with various stakeholders calling for a unified legal and ethical approach. At present, more specifically the capabilities of governmental administrators and general people to analyze big data is minimal and should be enhanced to unleash the capabilities of big data in the policy and decision-making cycle. This paper attempts to present various developments related to big data, data analytics, and their role in governmental processes towards achieving sustainable development goals. After examining the subject area and issues highlighted in the paper, it is realized that if big data is properly analyzed with appropriate techniques and technologies, it can give a boost to many ideas and will assist in decision-making. It can be concluded that if the governmental agencies use big data analytics for providing services to the citizens then certainly it will give rise to not just more effective services but also help to embrace more responsive local governance systems. In the same context, it is relevant to say that there exists a variety of risks and data security concerns with big data technology-based applications. The study proposes that the government should focus on protecting the privacy rights of its people, i.e.it is relevant to protect the privacy of the people who are associated with the data by any means.

References

1. Curry E. (2016) The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy. Springer, Cham. https://doi.org/10.1007/978-3-319-21569-3_3
2. Marston, D.A., McElhinney, L.M., Ellis, R.J. *et al.* Next generation sequencing of viral RNA genomes. *BMC Genomics* **14,** 444 (2013). https://doi.org/10.1186/1471-2164-14-444
3. J. Manyika *et al.*, "Big data : The next frontier for innovation, competition, and productivity," 2011.
4. Gartner Inc., "What Is Big Data? - Gartner IT Glossary - Big Data," *Gartner IT Glossary*, 2013.
5. M. A. Beyer and D. Laney, "The importance of 'big data': a definition," *Stamford, CT Gart.*, 2012.
6. Zikopoulos Paul and Eaton Chris, *Understanding Big Data*, vol. 1, no. 2019.
7. A. Patrizio, "Expect 175 zettabytes of data worldwide by 2025," *NetworkWorld*, 2019.
8. Ying Lin, "10 Internet Statistics Every Marketer Should Know in 2020 [Infographic]," 2019.
9. S. Kemp, "Digital 2020: Global Digital Overview — DataReportal – Global Digital Insights," *Datareportal.com*, 2020. [Online]. Available: https://datareportal.com/reports/digital-2020-global-digital-overview%0Ahttps://datareportal.com/reports/digital-2020-global-digital-overview%0Ahttps://datareportal.com/reports/digital-2020-ecuador%0Ahttps://datareportal.com/reports/digital-2020-global-di.
10. K. D. Foote, "A Brief History of Big Data - DATAVERSITY." 201
11. J. Manyika *et al.*, "Big data : The next frontier for innovation , competition , and

productivity," 2011.
12. Doug Laney, "3D Data Management: Controlling Data Volume, Velocity, and Variety," *META Gr.*, no. February 2001, 2001.
13. M. Loukides, "What is data science – O'Reilly." O'Reily Radar, 2010.
14. M. Stonebraker, "What Does ' Big Data ' Mean and Who Will Win ?," *Xldb*, 2012.
15. MIKE2.0, "Open Framework, Information Management Strategy & Collaborative Governance | Data & Social Methodology - MIKE2.0 Methodology.".
16. Wikipedia, "Big data - Wikipedia," *Wikipedia, The Free Encyclopedia*, 2019. [Online]. Available: http://en.wikipedia.org/wiki/Big_data.
17. NESSI, "Big Data_ A New World for Opportunities," *insideBIGDATA*, p. 25, 2012.
18. Microsoft, "The Big Bang: How the Big Data Explosion Is Changing the World," *Microsoft News Center*. pp. 1–35, 2013.
19. R. L. Villars, C. W. Olofson, and M. Eastwood, "Big Data: What It Is and Why You Should Care," *IDC White Pap.*, pp. 7–8, 2011.
20. A. Brust, "Big Data: Defining its definition," *ZDNet*, 2012.
21. A. Jacobs, "The Pathologies of Big Data," 2009.
22. YourDictionary, "Data dictionary definition _ data defined."
23. M. Lněnička and J. Komárková, "The Impact of Cloud Computing and Open (Big) Data on the Enterprise Architecture Framework," in *Proceedings of the 26th International Business-Information-Management-Association Conference (pp. 1679–1683). Norristown: IBIMA.*, 2015, no. June, pp. 1679–1683.
24. G. Vossen, "Big data as the new enabler in business and other intelligence," *Vietnam J. Comput. Sci.*, vol. 1, no. 1, pp. 3–14, 2014.
25. R. Bryant, R. Katz, and E. Lazowska, "Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society in Computing Research Initiatives for the 21st Century.," *Comput. Res. Assoc.*, 2008.
26. Apache Hadoop. What Is Apache Hadoop? Available online: http://hadoop.apache.org/ (accessed on 13 September 2017).
27. Cloudera. Available online: https://www.cloudera.com/ (accessed on 14 September 2017).
28. Hortonworks. Available online: https://hortonworks.com/ (accessed on 14 September 2017).
29. Shvachko, K.; Kuang, H.; Radia, S.; Chansler, R. The Hadoop distributed file system. In Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), Incline Village, NV, USA, 3–7 May 2010.
30. Dean, J.; Ghemawat, S. MapReduce: A flexible data processing tool. Commun. ACM 2010, 53, 72–77. [CrossRef]
31. Lakshman, A.; Malik, P. Cassandra: Structured storage system on a p2p network. In Proceedings of the 28th ACM symposium on Principles of distributed computing, Calgary, AB, Canada, 10–12 August 2009.
32. Chang, F.; Dean, J.; Ghemawat, S.; Hsieh, W.C.; Wallach, D.A.; Burrows, M.; Chandra, T.; Fikes, A.; Gruber, R.E. Bigtable: A distributed storage system for structured data. ACM Trans. Comput. Syst. (TOCS) 2008, 26, 4. [CrossRef]
33. Kulkarni, A.P.; Khandelwal, M. Survey on Hadoop and Introduction to YARN. Int. J. Emerg. Technol. Adv. Eng. 2014, 4, 82–87.
34. Taylor, R.C. An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics. BMC Bioinform. 2010, 11, S1. [CrossRef] [PubMed]
35. Saha, B.; Shah, H.; Seth, S.; Vijayaraghavan, G.; Murthy, A.; Curino, C. Apache Tez: A unifying framework for modelling and building data processing applications. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Australia, 31 May–4 June 2015.
36. M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mob. Networks Appl.*, vol. 19, no. 2, pp. 171–209, 2014.

37. J. C. Bertot and H. Choi, “Big data and e-government: Issues, policies, and recommendations,” *ACM Int. Conf. Proceeding Ser.*, pp. 1–10, 2013.
38. A. Gandomi and M. Haider, “Beyond the hype: Big data concepts, methods, and analytics,” *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, 2015.
39. P. Lake and R. Drake, *Information Systems Management in the Big Data Era*. Springer International Publishing, 2014.
40. J. Manyika *et al.*, “Big data : The next frontier for innovation, competition, and productivity,” 2011.
41. C. W. Tsai, C. F. Lai, H. C. Chao, and A. V. Vasilakos, “Big data analytics: a survey,” *J. Big Data*, vol. 2, no. 1, pp. 1–32, 2015.
42. B. Klievink, B. J. Romijn, S. Cunningham, and H. de Bruijn, “Big data in the public sector: Uncertainties and readiness,” *Inf90. Syst. Front.*, vol. 19, no. 2, pp. 267–283, 2017.
43. Doug Laney, “3D Data Management: Controlling Data Volume, Velocity, and Variety,” *META Gr.*, no. February 2001, 2001.
44. Meadows, D.H.; Meadows, D.L.; Randers, J.; Behrens, W.W. The Limits to Growth, 5th ed.; Universe Books: New York, NY, USA, 1972, ISBN 0-87663-165-0.
45. Declaration of the United Conference on the Human Environment. Available online: http://www.undocuments.net/aconf48-14r1.pdf (accessed on 25 September 2017).
46. UN Documents, Report of the World Commission on Environment and Development: Our Common Future. Available online: http://www.un-documents.net/wced-ocf.htm (accessed on 25 September 2017).
47. The Rio Declaration on Environment and Development. Available online: http://www.unesco.org/ education/pdf/RIO_W.PDF (accessed on 26 September 2017).
48. Wheeler, S.M.; Beatley, T. The Sustainable Urban Development Reader, 3rd ed.; Routledge: New York, NY, USA, 2014; pp. 79–87, ISBN 978-0-415-70775-6.
49. Sustainable Development Knowledge Platform, Transforming Our World: The 2030 Agenda for Sustainable Development. Available online: https://sustainabledevelopment.un.org/post2015/transformingourworld (accessed on 5 September 2017).
50. Harris, J.M. Basic principles of sustainable development. In Dimensions of Sustainable Development; Seidler, R., Bawa, K.S., Eds.; Eolss Publishers Co. Ltd.: Oxford, UK, 2009; Volume 1, pp. 21–41, ISBN 978-1-84826-207-2.