Journal of Applied Sciences, Information and Computing
School of Mathematics and Computing, Kampala International University

ISSN: 1813-3509 https://doi.org/10.59568/JASIC-2022-3-2-02

Assessing the Impact of Big Data Analytics in the Telecommunications Sector

[1]M. A. Zayyad
[1]Department of Information Technology & Systems
School of Mathematics and Computing
Kampala International University
PO Box 20000, Kampala Uganda
zayyad.musa@kiu.ac.ug

Abstract
Big data analytics is a significant way of improving telecommunications operators' Quality of Service (QoS) and improve their businesses. This study was aimed at determining the impact of big data analysis in telecommunication industry and also to arouse their interest towards big data. The study used four domestic telecommunication companies as case study, namely MTN, GLO, 9Mobile and AIRTEL. The study uses survey method of data collection by using structured questionnaire. Sixteen network nodes were used in the study from four clusters. Analysis of Variance (ANOVA), Cronbach's Alpha Coefficient using Smart – PLS, DBSCAN Algorithm was employed using google python as data analysis method. The study revealed that there is high traffic data from the telecoms and this is as a result of various user activities like internet usage and highly dense network clusters, which happened at a specific time or within a particular duration at a specified location. The findings led to a conclusion that the big data analytics can be considered as a solution to directional small cells in millimeter wave cellular networks to improve QoS. The study recommends that big data should be given due consideration and be used in making national budget and policy formulation. The study further recommends that tertiary institutions should create big data analytics training to create a platform for training and development of manpower mainly on big data analytics as is obtainable in developing and IT-driven countries.

Keywords: Analytics, Big, Data, QoS, Telecommunications.

1. Introduction

With the advent of digital technology and smart devices, a large amount of digital data is being generated every day. Advances in digital sensors and communication technology have enormously added to this huge amount of data, capturing valuable information for enterprises and businesses. The Big data is hard to process using conventional technologies and calls for massive parallel processing. Technologies that are able to store and process terabytes or petabytes of data without tremendously raising the data warehousing cost is a need of time. Ability to derive insights from this massive data has the potential to transform how we live, think and work. Benefits from Big data analysis range from healthcare domain to government to finance to marketing and many more (Günther, Mehrizi, Huysman & Feldberg, 2017)

Big data open source technologies have gained quite a bit of attraction due to the demonstrated ability to parallel processing of large amounts of data. Both parallel processing and technique of bringing computation to data has made it possible to process large datasets at high speed. These key features and ability to process vast data has been a great motivation to take a look into the architecture of the industry leading big data processing framework by Apache, Hadoop. Understand how this big data storage and analysis is achieved and experimenting with RDBMS vs Hadoop environment has proven to provide a

great insight into much talked about technology (Mercier, 2019).

Big data analytics is not just a passing trend; it is becoming an important part in every aspect of a communication service in Nigeria and all over the world. Every time we talk about the electronic communication, it means we are talking about the production of data by using wired or wireless medium (Lee, 2017). The speedy growth of internet and the availability of technology everywhere, internet users are increasing day by day. On the other hand, the boom in smartphone industry makes it easier for users to access network on the move. This ease of access is becoming great threat for telecom industry. People are diverting from wired to wireless medium especially on GSM to access the internet. Mobile applications such as Google, Facebook, Instagram, WhatsApp, Google+, Twitter, etc., providing ease to connect, at the same time are generating huge amount of data to tackle with (Wilmer, Sherman & Chein, 2017).

According to Sivarajah, Kamal, Irani, & Weerakkody (2017), big data is an aggregate of data sets that are large and complex, thus overwhelming the traditional data mining tools. Big data broadly refers to the new methods and technologies for collecting, managing, and analyzing in real time the vast amount of both structured and unstructured data. Curry (2016) defined big data as "high-volume, high-velocity and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization". Big data analytics is a term encompassing the new methods, tools and technologies for collecting, managing and analyzing, in real-time, the vast increase in both structured and unstructured data for insightful and effective decision making. Big data offers telecommunication business a real opportunity to gain a much more complete picture of their operations and their customers, and to further their innovation efforts. Big data demands of every industry a very different and unconventional approach to business development. The operators that can incorporate new agile strategies into their organizational processes will gain a more competitive advantage that their slower rivals.

Several researchers (Bao, Liu & Chen (2018); Raban & Gordon (2020); Jensen, Nguyen, Do & Årnes (2017); and Gandomi & Haider (2015)) from different parts of the world are showing interest towards big data analytics. Raban & Gordon (2020) lamented that data is growing exponentially attracting researchers to work of a popular term, the Big Data. Big Data is observed in various fields, such as information technology, telecommunication, theoretical computing, mathematics, data mining and data warehousing. Data science is frequently referred with Big Data as it uses methods to scale down the Big Data. Currently more than 3.2 billion of the world population is connected to internet out of which 46% are connected via smart phones (Khan, Khan, Alam & Ali, 2018). Over 5.5 billion people are using cell phones. As technology is rapidly shifting from ordinary cell phones towards smart phones, therefore proportion of using internet is also growing. There is a forecast that by 2020 around 7 billion people at the globe will be using internet out of which 52% will be using their smart phones to connect. In year 2050 that figure will be touching 95% of world population. Their work gave a comparison of amount of Big Data generated by telecom industry. Based on the collected data used, forecasting tools to predict the amount of Big Data to be generated in future and also identify threats that telecom industry will be facing from that huge amount of Big Data (Khan, et. al., 2018).

Several researches discussed big data analytics but projected their interest in different areas. For instance, Raban & Gordon (2020) conducted a bibliometric analysis on the evolution of data science and big data research. Jensen, et al., (2017) conducted a study on big data analytics approach to combat telecommunication vulnerabilities. Similarly, Bao, Liu & Chen (2018) developed a framework for learning-based automatic parameter tuning for big data analytics. However, there was no research that cover the whole big data analytics in general. In view of that, this study sought to investigate the big data analytics based on a telecommunication approach, which generally cover the whole big data analytics processes from big data error, failure, interruption, damage, mining, coding, and security.

Despite the interest of international researchers towards big data analytics, there is limited interest towards the subject matter in Nigeria. Very few researches have been found that show interest towards big data analytics in telecoms industry in Nigeria but no research has been found that generally investigate big data in telecoms industry. Therefore, this study sought to investigate big data analytics: a telecommunication approach in Nigeria. The study used combine research methodology, survey research design was used to collect big data resources from the four telecommunication companies the study used whereas descriptive research design was used to define big data analytics system. The objective of the study is to assess the impact of big data analytics for the telecommunication industry.

1.1 Statement of the Problem

Consequently, the vast amount of data that operators are generating have potentially outpace the ability of existing processing system. Telecommunications data can be used to optimize operations and drive operational business intelligence to realize immediate business opportunities. Multiple foreign mainstream telecom operators have already applied big data for their own development. Therefore, there is need for domestic operators to also engage into innovating their business models by exploring the use of big data. Several researches are already adding value to telecommunication industries in developing countries like Orange Business Services for instance, used big data to enhance the accuracy of their churn detection. Spain's Telefonica Dynamic Insights obtained reliable predictions of user behavior by packaging and analyzing data. In 2014, Verizon built data centers in California to implement precision marketing. But there has been limited interest of Nigerian telecoms industry towards big data analytics. Therefore in an effort to arose their interest, this study sought to investigate the big data analytics: a telecommunication approach. The study adopted combined research methodology and used different algorithms to ensure efficient and effective outcomes.

The study will have significant impact in structuring findings from different sources of data collection using telecommunication industries. The study will also help in providing businesses with business intelligence that can reduce cost and improve the efficiency of operations. This approach will help to analyse past data to make predictions about the future.

2. Literature Review

The term Big Data applies to information that cannot be processed or analyzed using traditional processes or tools (Oussous, Benjelloun, Lahcen & Belfkih, 2018). "Big Data", because the term is a bit of a misnomer since it implies that pre-existing data is somehow small (it isn't) or that the only challenge is its sheer size (size is one of them, but there are often more). In short, increasingly, organizations today are facing more and more Big Data challenges. They have access to a wealth of information, but they don't know how to get value out of it because it is sitting in its most raw form or in a semi structured or unstructured format; and as a result, they don't even know whether it's worth keeping (or even able to keep it for that matter). An IBM survey found that over half of the business leaders today realize they don't have access to the insights they need to do their jobs. Companies are facing these challenges in a climate where they have the ability to store anything and they are generating data like never before in history; combined, this presents a real information challenge. It's a conundrum: today's business has more access to potential insight than ever before, yet as this potential gold mine of data piles up, the percentage of data the business can process is going down fast (Clohessy & Acton, 2019).

Quite simply, the Big Data era is in full force today because the world is changing. Through instrumentation, we're able to sense more things, and if we can sense it, we tend to try and store it (or at least some of it). Through advances in communications technology, people and things are becoming increasingly interconnected—and not just some of the time, but all of the time. This interconnectivity rate is a runaway train. Generally referred to as machine-to-machine (M2M), interconnectivity is responsible for double-digit year over year data growth rates. Finally, because small integrated circuits are now so inexpensive, we're able to add intelligence to almost everything (Sonkar, 2020).

The volume associated with the Big Data phenomena brings along new challenges for data centers trying to deal with it: its variety (Mikalef, Boura, Lekakos & Krogstie, 2019).

With the explosion of sensors, and smart devices, as well as social collaboration technologies, data in an enterprise has become complex, because it includes not only traditional relational data, but also raw, semi structured, and unstructured data from web pages, web log files (including click-stream data), search indexes, social media forums, e-mail, documents, sensor data from active and passive systems, and so on (Sonkar, 2020). What is more, traditional systems can struggle to store and perform the required analytics to gain understanding from the contents of these logs be-cause much of the information being generated doesn't lend itself to traditional database technologies. However, an organization's success will rely on its ability to draw insights from the various kinds of data available to it, which includes both traditional and nontraditional (Mikalef, 2019).

3. Research Methodology

The research methodology comprises of the research design, target population of the study, study area, sample and sampling techniques, data collection instruments, data collection procedure and method of data analysis.

3.1 Research Design

The methodology adopted for this study was survey and descriptive research design. The appropriate design for the study is by the use of structured questionnaire for key members of the sampled telecommunication services in Nigeria. The survey research design is aimed at collecting primary data about issues surrounding big data analytics in telecommunication industry; such as various error, data failure, data interruption, data damage, data mining, data coding, data and information security and safeguarding, etc. While the descriptive research design in this regards has to do with design big data analytics system on the aforementioned issues or some of the mentioned issues above.

3.2 Population of the Study

The population of this study comprises all telecommunication service providers in Nigerian telecommunication industry; this specifically involves the MTN, Airtel, GLO Nigeria and 9Mobile. Moreover, the study all so involve telecommunication regulatory agencies that is Nigerian Communication Commission (NCC) and National Information Technology Development Agency (NITDA).

3.3 Sample and Sampling Technique

In order to select the sample size of the participants for this research work, a purposive sampling technique was used because it gives the researcher freedom of selecting the most appropriate samples for the study and also ensure the greater representation and guarantee the efficiency of the study. This study employ census study hence the populations of the study are very minimal and they can be studied entirely.

3.4 Method of Data Collection

The method of data collection used in the research work was both the primary and secondary sources of data collection. The primary data was obtained by administering structured questionnaire to the staff of the telecoms sector. The secondary data was collected from textbooks, journals, magazines, and other related literature that are relevant to this research work. After collecting the data, the study employed coding procedure and subject the responses for various statistical analysis.

3.5 Data Analysis Techniques

The method of data analysis employed in this research is Analysis of Variance (ANOVA) and Cronbach's Alpha Coefficient using Smart–PLS analysis software v16.0. The data collected from the respondents was first coded using a Likert scale of: Strongly agree, Agree, Disagree, Strongly disagree.

And the coded data was uploaded into the Smart–PLS software for calculation.

A one way analysis of variance was conducted using K-means cluster of each network provider. The telecoms provides 4 cluster hopes for the data, each cluster was calculated separately and later converted into simple method for interpreting the result from technical form to simpler form. The R–Square and Cronbach's Alpha coefficients was calculated using Smart–PLS software and the result and interpretation were presented in later section.

This model manifest each network provider cluster and their sub-clusters as shown in Figure 4.1. Big data experience and issues to influence the industrial big data. However, from the analysis, it was shows that coefficient determination (R-square) of Big Data with 0.651 coefficient value is showing 65%; meaning that, the four latent variables (GLO and 9MOBILE) independent variables are able to generate higher data because the inner model path values are higher than 0.1 with 0.322 for GLO and 0.666 for 9MOBILE. On the other hand, MTN and AIRTEL generate lower data traffic with the inner model path less than the 0.1 scale i.e. MTN with -0.023 and -0.191 respectively. To reduce that data generated from these companies cluster there is need for them to implement the same algorithm can be used to calculate generated data from each network antenna so as to reduce redundancy. The algorithm is set to minimize generated data by these companies in order to cut down the expenses of storage system and customer generated data for future use. From my views, when K – means Partial Least Square (PLS) algorithms are implemented in each network cluster, generated data will be minimized.

R-square, also called the coefficient of determination, it is the overall effect size measure for the structural model, as in regression, indicating below that 0.651 which approximately 65% of the variance in the big data variable is explained by the model. No R-square is shown for MTN, AIRTEL, GLO and 9mobile as these are exogenous latent factors of 0.95 for MTN, 0.90 for both GLO and AIRTEL, 0.78 for 9MOBILE and 0.55 for the cumulative Big Data generated as shown in Figure 1 below. Nakagawa, Johnson & Schielzeth (2017) describes results above the cut offs 0.67, 0.33 and 0.19 to be "substantial", "moderate" and "weak" respectively. The R-square here would be considered to be substantial and excellent strength or effect.

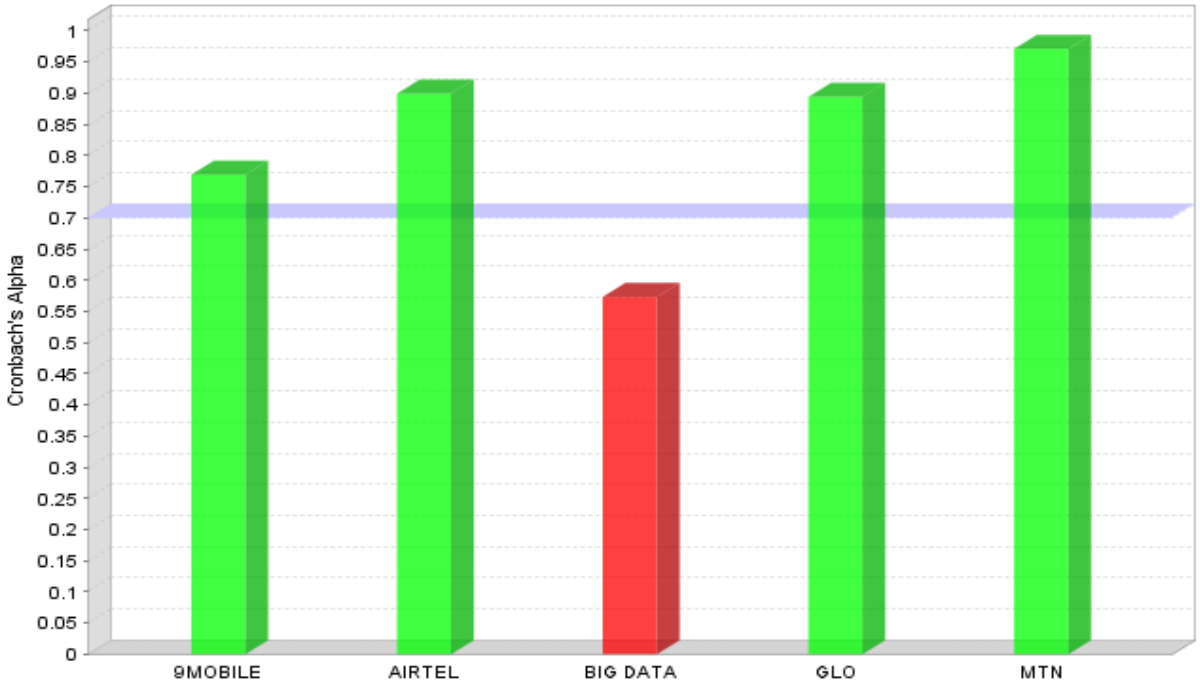

Figure 1: R – Square Bar chart of each network provider and their big data

3.6 Cronbach's Alpha

The Cronbach's alpha also addresses the question of whether the indicators for latent variables display convergent validity and hence display reliability. By convention, the same cutoffs apply: greater or equal to .80 for a good scale, .70 for an acceptable scale, and .60 for a scale for exploratory purposes. Note, however, Cronbach's alpha is a conservative measure which tends to underestimate reliability as shown in Figure 2 below. For this study, MTN, AIRTEL, GLO, 9MOBILE and overall performance factors are measured at an acceptable level for confirmatory research as such MTN has the highest value of 0.95, GLO and AIRTEL with 0.90 each, 9MOBILE with 0.78 and the cumulative reliability as 0.77. As Cronbach's alpha is biased against short scales of two or three variable as in this study, all the variables are within the acceptable scale of 70%, except the big data which is at 55%.

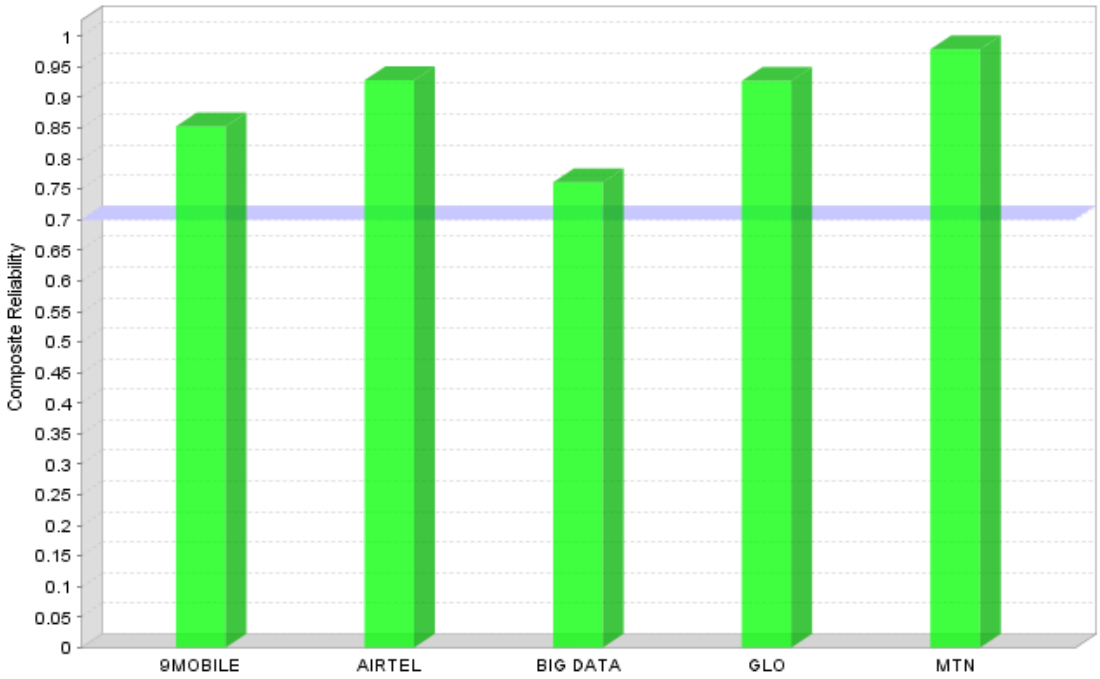

Figure 2: Cronbach's Alpha Coefficient of each network provider and their big data

4. Discussion

The study primary data reviled that, individual telecommunication service provider in Nigeria, like MTN Nigeria's big data have significant impact on the overall big data issues in the entire telecomm industry in Nigeria. Moreover, this study discovered that, the entire 4 leading telecommunication service providers are impact full on the overall Nigeria's big data in the telecommunication industry; with R-square stood at 65%.

Furthermore, very high composite reliability (> .90) may indicate that the multiple indicators are minor wording variants of each other rather than being truly representative measures of the construct the factor represents. However, in this study, composite reliability stood.

This shows that, all the independent variables (MTN, Airtel, Glo, 9mobile) are very much adequate in measuring the effect of big data issue in Nigerian Telecommunication industry, because they manifest value higher than the acceptable validity rate of adequate i.e. 70% and good at 80%. Furthermore, overall effect of the three variables stood at 96% which shows excellent effect. From the forgoing therefore, all the variables use in this study are valid and reliable considering good measure of performance as such findings of the study are free from validity and reliability bias.

5. Conclusion

The primary objective of this thesis was to get insights of spatial temporal telecom networks' data to detect meaningful traffic patterns. The traffic pattern is a result of various user activities like voice, text, and internet, which happened at a specific time or within a particular duration at a specified location. To accomplish the desired goal and end objectives, The DBSCAN (Eps, MinPts) was identified as the base algorithm after conducting a literature review of many scientific research papers based on machine learning and clustering analysis. The next major challenge was to select the input parameters and optimization of data sets in order to achieve the thesis's objective and desired results. In density-based, spatial-clustering algorithms, the three dimensional data has been provided and another parameters Eps and MinPts given to the algorithm. The algorithm has been run and results verified for various values of input Eps and MinPts in order to receive better and refined results and fulfil the thesis' objective. Here better and refined results means clusters results, which reveal the facts of different clusters, the optimal number of clusters required, minimum standard deviation of traffic volume across all clusters, convex shaped cluster and these results stands true for average 1000m coverage range of small cells. The DBSCAN algorithm is sensitive to its input parameters, hence to use this algorithm the dataset and domain knowledge is required so that inputs parameters can be selected precisely to get the desired results. During experimentation, 16 input parameter combinations were executed with the algorithm, and the results were further verified with the other algorithm for all input data. Also, manually and based on the domain knowledge, I have verified all the results. In this analysis, it is evident that this input parameter with the algorithm covers all the geographical area in the form of 16 clusters from the Katsina dataset.

Whereas other approaches like bin-packing algorithms and grid-based algorithms are giving correct results theoretically but practically are not suitable for the analysis of spatial data. The main advantages of this algorithm are the ability to extract all meaningful and hidden facts from the datasets based on the input parameters. The input parameters have a decisive impact on the cluster result. The proposed solution can extract spatial- and semantic-based clusters that allow service providers to plan the network planning effectively. This approach can be useful for similar types of dataset, but certain extra preprocessing might be required based on the nature of the data. During the execution of the clustering algorithm, many problems occurred. For example the precise selection of input parameters to overcome the noise or to minimize the noise, data cleaning to filter out the null and unused values from datasets, visualizing the data in order to represent the results which can be understand were among the challenging tasks. The objective of this thesis has been achieved, but there are still a few aspects that can be considered further. As mentioned earlier, the proposed solutions can be considered for Directional Small Cells in Millimeter Wave Cellular Networks.

6. Recommendations

This study made the following recommendations:

1. Big data should be given due protection and be used in making national budget and policy formulation.
2. Tertiary institution should create big data analytics training to create platform for training and development of manpower mainly on big data analytics as is obtainable in developing and Information Technology (IT) driven countries like Singapore etc.
3. Regulatory commissions such as Nigeria Communications Commission (NCC) and National Information Technology Development Agency (NITDA) should lay much emphasis on big data analytics as it helps in addressing most challenges in the telecommunication industry of the country.
4. Government should leverage on big data at telecommunication services providers in fighting insurgency and other insecurity challenges in the country, hence most of this criminals are communicating with their mobile phones before, during and after the crime.

References

Bao, L., Liu, X., & Chen, W. (2018, December). Learning-based automatic parameter tuning for big data analytics frameworks. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 181-190). IEEE.

Bharti, V. (2013) NS-DBSCAN: A Density-Based Clustering Algorithm in Network Space. *International Journal of Geo-Information.* 8, 218; DOI: 10.3390/ijgi8050218

Clohessy, T., & Acton, T. (2019). Investigating the influence of organizational factors on blockchain adoption: An innovation theory perspective. *Industrial Management & Data Systems*, *119*(7), 1457-1491.

Curry, E. (2016). The big data value chain: definitions, concepts, and theoretical approaches. *New horizons for a data-driven economy: A roadmap for usage and exploitation of big data in Europe*, 29-37.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, *35*(2), 137-144.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191-209.

Jensen, K., Nguyen, H. T., Do, T. V., & Årnes, A. (2017). A big data analytics approach to combat telecommunication vulnerabilities. *Cluster Computing*, *20*, 2363-2374.

Khan, M. W., Khan, M. A., Alam, M., & Ali, W. (2018). Impact of big data over telecom industry. *Pakistan Journal of Engineering, Technology & Science*, *6*(2).

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business horizons*, *60*(3), 293-303.

Mercier, M. (2019). *Contribution to High Performance Computing and Big Data Infrastructure Convergence* (Doctoral dissertation, Université Grenoble Alpes).

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, *98*, 261-276.

Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R 2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), 20170213.

Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, *30*(4), 431-448.

Raban, D. R., & Gordon, A. (2020). The evolution of data science and big data research: A bibliometric analysis. *Scientometrics*, *122*(3), 1563-1581.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of business research*, *70*, 263-286.

Sonkar, S. (2020). What is Big Data? A Quick Introduction for Analytics and Data Engineering Beginners. *Analytics Vidhya*.

Wilmer, H. H., Sherman, L. E., & Chein, J. M. (2017). Smartphones and cognition: A review of research exploring the links between mobile technology habits and cognitive functioning. *Frontiers in psychology*, *8*, 605.