Bayero Journal of Pure and Applied Sciences, 11(2): 252 - 259
Received: July, 2018
Accepted: December, 2018
ISSN 2006 – 6996

BIG DATA-DRIVEN E-GOVERNMENT FRAMEWORK IN NIGERIA

[1]Ogbuju, E., [2]Taiwo, K., [3]Ejiofor, V. and [4]Onyesolu, M.
[*1,2]Department of Computer Science, Federal University Lokoja, Kogi State, Nigeria
[3,4]Department of Computer Science, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria
[1]Correspondence Author: emeka.ogbuju@fulokoja.edu.ng

ABSTRACT
Many nations of the world are employing the use of big data in e-government for integrated policy making. However, in Nigeria, data-driven decision making is still a challenge due to lack of a centralized framework that can integrate different silo data centers for insightful and evidence-based governance. The paper proposes a big data-driven e-government framework for e-governance operations. The framework was designed in the Nigerian context using big data mechanisms. It was tested with an experiment that performed a sentiment analysis to uncover the opinion of the masses on selected agencies in order to help policy making. This was done with datasets collected from the social media handles of two (2) government agencies – the Nigerian police force and the Nigerian army. The results uncovered the actual opinions of the populace concerning both agencies as it regards national security. The framework promises a host of benefits such as centralized data entry point for governance, data harvesting system, seamless integration and collaboration of government operations and an easy application of open data policy. The paper concludes with four (4) recommendations which the government of any developing nation can imbibe for results-oriented e-governance and increase the quality of service to citizens and businesses.
Keywords: big data, data, e-government, e-governance, framework, hadoop, sentiment analysis

INTRODUCTION

There is need for an automated solution that will enable governments in the developing nations to embrace movements toward policy integration and integrated service delivery with the help of big data analytics (UNDESA, 2016). The aim of this paper is to propose a solution that will meet this need in the Nigerian context. Big data analytics is the current technology that many nations of the world are adopting for improved governance. Though progress had been made in Nigeria in e-government initiatives, big data perspectives had not been employed (Hilbert and Lopez, 2011). There have been efforts by the government to establish unified framework for adoption of ICT in governance (NITDA, 2010). These efforts started since 2007 with the creation of the National Information Technology Development Agency (NITDA) by the federal government, specifically to plan, develop and promote the use of information technology in Nigeria. The agency developed a national e-Government framework in 2010 with a vision to be among the top 20 ranking nations in the UN e-Government survey come year 2020. They went ahead to establish the National e-Government Strategies Limited (NeGST) which specifically sees to facilitate drive, and implement the Nigerian e-government programme under a public-private partnership model. Some of the e-governance initiatives projected by NeGST include e-NYSC, e-Passport, e-readiness, e-LGA and e-parliament. NITDA also oversees the National ICT Policy which holds standards for e-government practices in Nigeria and serves as a unifying policy for Information Policy of the National Media Commission and Telecommunication Policy of the National Communications Commission (Fraser-Moleketi and Senghor, 2011).

Despite these progresses made so far in the framework proposition, there are situations that demonstrate lack of evidence-based governance in Nigeria. For instance, there is a local community with an old native stream flowing with natural and fresh water but without pipe borne water. The government proceeded to provide them with a central borehole. Few years later, during a census exercise, the government officials sent to the community were astonished that the community still walk a long distance to the native stream for their water supply without using the central borehole. They tried to find out why they could abandon such a 'huge' government provision to the community. They found out that the community do not actually need the borehole because there is a concealed culture in that community for husbands to send away their children to the streams in order to allow them private time to meet with their wives. Moreover, they prefer the fresh water from their native stream to the water from the borehole. This is the fate of some government projects in Nigeria. They are projects executed without proper assessment of the actual needs of the people; projects without background data collection, analysis and smart recommendation. Government decisions affecting the lives of citizens need to be made with facts and evidences. This is referred to as evidence-based governance or evidence-based decision (Wilma *et. al*., 2010). This form of decision must be data-driven. Government projects should not be motivated by personal aggrandisement, political underplays or self-seeking glory else they will exist as monumental wastes which do not really address the actual needs of the people. There is a high volume of structured data available in the coffers of all government establishments.

In addition, there is a high variety of datasets which are unstructured in nature and are generated in high velocity by government operations. These datasets come from government websites and its associated blogs, social media accounts, emails, videos and photos. So, governance cannot be disassociated from data. These data move in and out of government channels from both private and public domains. In fact, e-governance can be defined in terms of these data movement as the automated movement of relevant data within e-government channels. E-government is essentially data-driven. Data is the primary input into any e-government platform. These data are processed and information is returned as output, which sometimes also serves as an input for other e-government channels. So, there exists a revolving movement of data in and out of any e-government platform. One of the major challenges the e-government platforms in Nigeria are facing is the lack of technical analysis of the huge available datasets in their various platforms. For example, most government datasets are text-based and commonly available to the public through their various online platforms. Technical skills in Machine Learning are required to analyze these texts. The availability of these texts is an advantage in itself for insights generation as they could be mined, extracted and analyzed for informed decisions. Government can be said to be data-driven when vital decisions are made from sentiments expressed by these texts.

The population of Nigeria is over 190 million (Worldometers, 2018). Apart from data generated by government institutions and agencies, this populace generates personal data on a daily basis. This has given rise to a data deluge in Nigeria thereby leading the government to establish designated agencies for data collection and management; which according to Hilbert and Lopez (2011) includes: *(i).* Federal Road Safety Commission (FRSC), responsible for drivers' license and vehicle number plates; *(ii).* Independent National Electoral Commission (INEC), responsible for voters registration exercise; *(iii).* National Bureau of Statistic (NBS), responsible for the production of national official statistics; *(iv).* National Identity Management Commission (NIMC) which is responsible for the national identity database; *(v).* National Population Commission (NPC), in charge of national demographic data; *(vi).* Other organizations including the banks in the financial sector and telecommunication companies in the telecommunication sector such as MTN, Globacom, Airtel, etc. Most of the data collected by these agencies are structured in nature. These structured data are very suitable for e-government services (Onwudebelu *et. al.*, 2016). However, knowing that over 80% of the world's data are currently unstructured (Iyilade, 2015), there is a large gap in the collection and management of unstructured datasets for e-government services. So, there are many government agencies that collect same form of data in silos. Citizens are made to provide same information at different times to different agencies causing lots of time wastages. For example, the same personal details a citizen is required to fill in NIMC will be repeated at NBS, NPC, MTN, etc. This duplication of data in silos is common. There is no seamless interface of data sharing among governmental agencies. Every agency collects and manages its own data separately. In confirmation, Kaka (2015) noted that almost all e-government applications in Nigeria operate in silos and in custody of their respective government institutions with limited or no sharing capabilities.

It is against these backdrops that this work seeks to propose a solution for an integrated service delivery. Instead of making decisions without data or reacting to situations as commonly seen in governance, government can concentrate on preventive measures using predictions or forecasts delivered by big data analytics. A similar project to the proposed framework had been applied in Turkey to integrate some sectors like transportation, health, public safety and others in a single data centre (Government of Turkey, 2015). Integrating government functions in this manner will enhance data collection, analyses and smart recommendation for improved governance. Also, Ogbuju et al. (2017) had proposed a centralized real-time data collection framework that is capable of aggregating all forms of tweets from social media channels with a predefined set of keywords into a data lake. Other nations applying analytics in governance include China (Hongli and Zhangxi, 2003; Greengard, 2014), India (Yadav and Singh, 2012; Thukral, 2014) and Australia (Pearce, 2015).

MATERIALS AND METHODS

The methodology adopted for this work is the big data mechanism. This mechanism follows standard data analytics methodologies like CRISP-DM (Chapman *et al.,* 2000), KDD (Fayyad *et al.* (1996) and SEMMA (Umair and Haseeb, 2014). Figure 1.0 is a schematic diagram of a typical big data mechanism. The mechanism encompasses six (6) machineries identified as resource management, data organization and management, analytics and discovery, and decision support and visualization (Layne and Lee, 2001; Kaka, 2015).

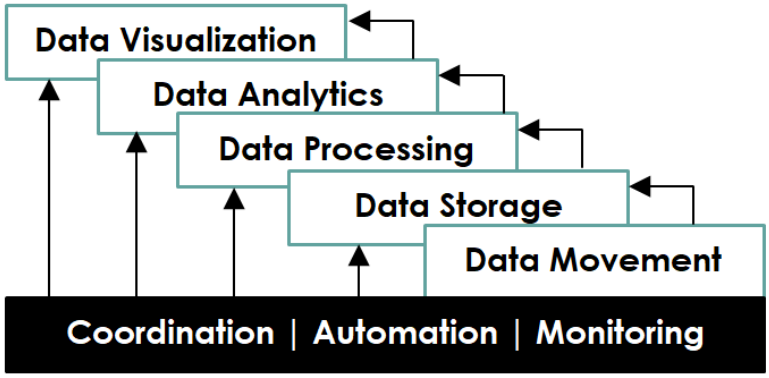

Figure 1: Big data mechanism

The holistic nature of the six-stage mechanism shown in Figure 1.0 enables the management of data in an all-encompassing environment made of different tools. Data is first moved into the system. The data movement mechanism can import and export relational data from RDBMS systems (data at rest) through the use of Sqoop, a bulk data transfer tool. Likewise, it can import or export streaming data flows from social networks (data in motion). The second mechanism ensures that the transferred dataset in their different formats are stored in the system using Hadoop Distributed File System (HDFS) for unstructured files and tools like HBase, MongoDB and Cassandra for NoSQL datasets (White, 2015). The third mechanism allows data processing, graph processing, scripting and querying to be performed on the dataset. Tools like Giraph, HCatalog, Pig, Hive and Drill are used for processing. Big data goes beyond data input, storage and processing to analytics and result visualizations; hence, the availability of the fourth mechanism for data analytics. Machine learning applications are performed to derive meaningful insights from the processed data using artificial intelligence algorithms. Tools like Mahout, Python and R can be used. The analytics mechanism is the backbone of the data-driven insight generation. It has the ability to perform several functions such as predictive analytics, prescriptive analytics, sentiment and behavioral analysis, fraud detection and analysis, risk analysis, forecasting, etc. (White, 2015). These tools can also be used in the fifth mechanism to provide functionalities for visualizations. Visualizations enable the system to show results in graphical ways. The last mechanism in every big data platform is the coordination, automation and monitoring (CAM). In this mechanism, tools like Zookeeper helps in cluster coordination, Oozie in workflow automation while Ambari in cluster monitoring. The applications of these mechanisms in e-government will uncover the potentials of big data analytics to increase service delivery. The materials used in the research are textual datasets from both websites and social media channels of the selected government agencies presented in the Experimental Design section.

RESULTS AND DISCUSSION

A. Conceptual Framework: Figure 2 depicts the proposed big data e-government framework. The diagram is a conceptual framework made up of e-government channels or platforms.

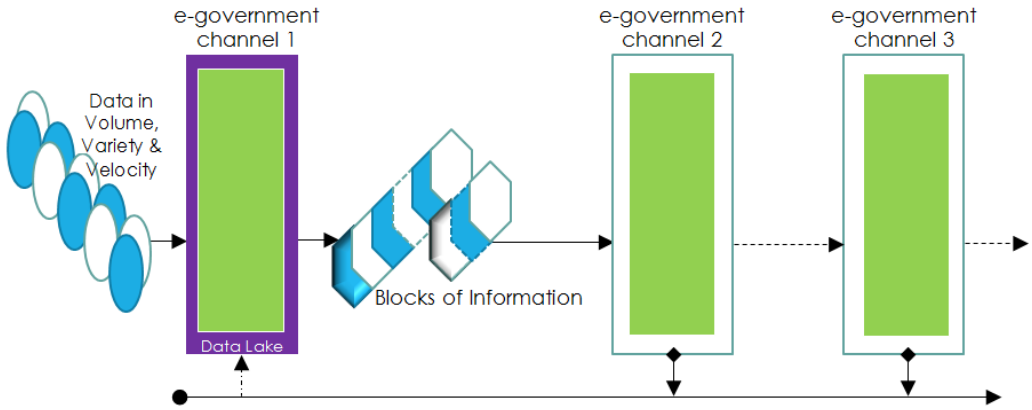

Figure 2: Big data e-government framework

These platforms can be seen as data centres for the agencies concerned. The features of the framework are discussed as follows:

1. **One point of data entry:** The framework advocates a single point of data entry, a centralized data collection point through which all forms of data may be delivered into the system. Other agencies will definitely be in need of these data but a designated agency is made to collect them for further distribution to others. This will significantly reduce the tedious processes citizens face in repeating their data details at various

platforms. In Figure 2.0, the *e-government channel 1* serves as the single data collection point with capabilities to store all variety of datasets coming in their streams. This system represents a typical big data technology platform that can collect, aggregate and manages structured and unstructured datasets. It is a typical data lake which holds all forms of data. Using Hadoop to implement this framework will offer an advanced platform for distributed computing thereby ensuring no loss of data due to its replication ability (White, 2015).

2. **Data Sharing:** Data will be freely allowed to revolve around the framework. For instance, say *e-government channel 2* is the Ministry of Health; the actual data needed by this ministry would be health information of the citizen. Hence, they would request from the *e-government channel 1* the needed information (represented by *blocks of information*). Likewise, say *e-government channel 3* is the Ministry of Finance; the actual data needed by them would be the financial records of the citizen. This record will be requested from the *e-government channel 1* again and the blocks of information are released to them. If, for instance, the Ministry of Finance needs health record also, they can directly and easily request that from the Ministry of Health who will release to them their already processed block of information relating to the citizen. This guarantees a free flow of information within the e-government framework.

3. **Citizen engagement:** With the advent of social networks, citizens, as agents of data, generate both structured and unstructured data on a daily basis. The big data e-government framework offers facility for engaging citizens and harnessing their user-generated data into the system. Most government establishments in Nigeria already have social network handles on their e-government channels. Citizens utilise these handles to post opinions directly. The central big data collection channel may not have these data but they can be transferred into it for storage and subsequent analysis. In some cases, machine learning algorithms may be deployed in a near real-time basis to provide faster response to these opinions. This will make citizens' engagement very high with the framework as issues are resolved promptly. This also fulfils one of the advantages of e-governance with respect to quick and easy access to information and fast resolution of citizen's agitations. Also, government to citizen as well as government to employee models of e-governance are effectively exploited when citizens engage with e-government channels

4. **Data harvesting:** Data can both be farmed and harvested. The framework provides a central platform where data may be farmed and harvested. Data farming is our new concept that entails cultivating (i.e. data collection and aggregation) and growing (i.e. data storage) data from scratch in a central location that can hold the nation's various forms of data. In Figure 2.0, data is primarily farmed on the *e-government channel 1*. New sets of data keep being added to this channel thereby growing it until it's due for harvesting (procession or analytics). The other e-government channels that request their needed data from channel 1 are simply harvesting the data for their use.

5. **Open data policy:** Similar to data sharing, open data is a more advanced form of data availability whereby the government puts to the public domain national datasets free from privacy concerns. These datasets are principally meant for professional analytic activities geared towards national growth and development. A seamless access to each agent's data is made possible.

B. Experimental Design: One of the major application areas of big data that applies to e-government is sentiment analysis. Sentiment analysis is an automated extraction of public opinions expressed in volumes of texts. Public opinions are available in many online e-government platforms (e.g. social networks). There is a need to extract these texts, analyze them and uncover the overall feelings of the people. This will advance a lot of public projects in the right direction. However, applying sentiment analysis and indeed other big data analytic applications will not be realistic without a unified framework. Having proposed and conceptualized the framework in the preceding section, an experiment with sentiment analysis is performed in this section.

a) **Data Collection:** To operationalize the proposed framework, sample data from two government websites were collected – **www.npf.gov.ng** and **www.army.mil.ng** and their associated social media accounts **@PoliceNG** and **@HQNigerianArmy** respectively. The social media data which were in textual format were loaded onto MongoDB, a NoSQL document store while the data collected over the websites (multiple format datasets) were loaded onto Hadoop. These tools are resident on the central data lake, a pseudo-distributed system for the experiment, from where the agencies may request their results. The interest here is to perform opinion mining and analyse the internal and public sentiments of these agencies. Data from the websites may be used to perform their internal sentiment (their perceptions about themselves) while data from the social networks were used to perform their public sentiment (people's perceptions about them).

b) **Implementation:** For the brevity of the demonstration, the experiment was restricted to the analysis of public sentiments using the social network handles identified above. R programming was used to deploy the system and obtained interesting insights into the activities of the agencies. Figure 3.0 illustrates the process of the work carried out.

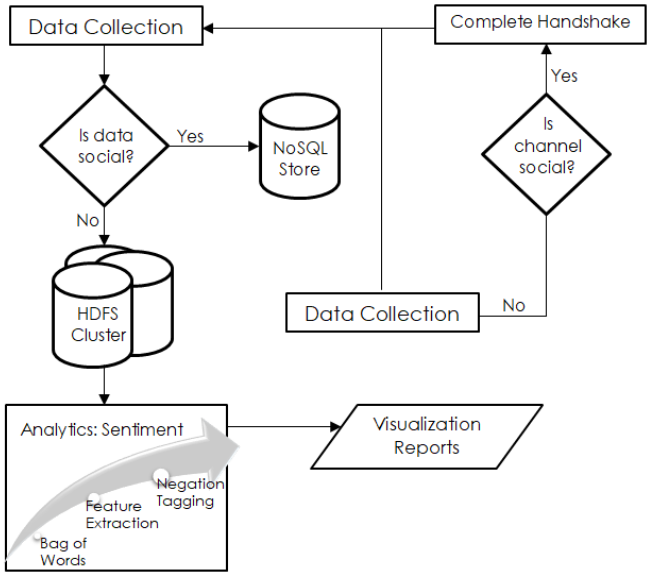

Figure 3.0: Experimental design

Through a technique known as complete handshake, a total of 10,000 tweets from the Police handle and 6,695 tweets from the Army handle were collected. The system worked with the help of word cloud, sentiment analysis and graphics/visualization libraries. The tweets were cleaned, processed and developed into a corpus. Bing's Lexicon approach (Liu & Hu, 2015) was used to classify the public sentiments into positive, neutral and negative. The NRC Lexicon approach (Mohammad & Turney, 2013) was used to further express the sentiments into eight (8) emotions. Both Figures 4.0 and 5.0 show a high negative sentiment expressed by the public towards the services of the agencies. This is also seen in the word clouds in Figures 6.0 and 7.0 showing negative words like kill, abduct, etc.

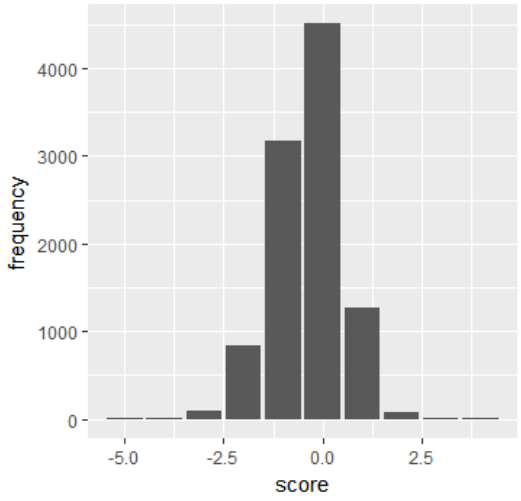

Figure 4.0: Sentiment Score from @PoliceNG

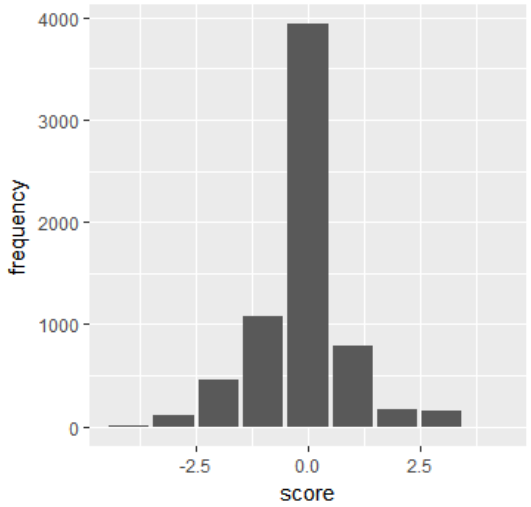

Figure 5.0: Sentiment Score from @HQNigerianArmy

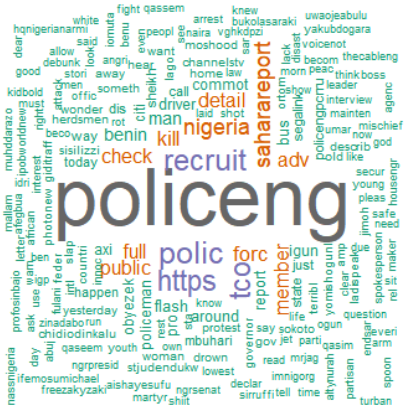

Figure 6.0: Word Cloud from @PoliceNG

Figure 7.0: Word Cloud from @HQNigerianArmy

There is however an observed high value of the neutral sentiments in both agencies. To understand the extent of this neutrality, the tweets were further analysed to classify the emotions of the sentiments. Figures 8.0 and 9.0 show the various classifications.

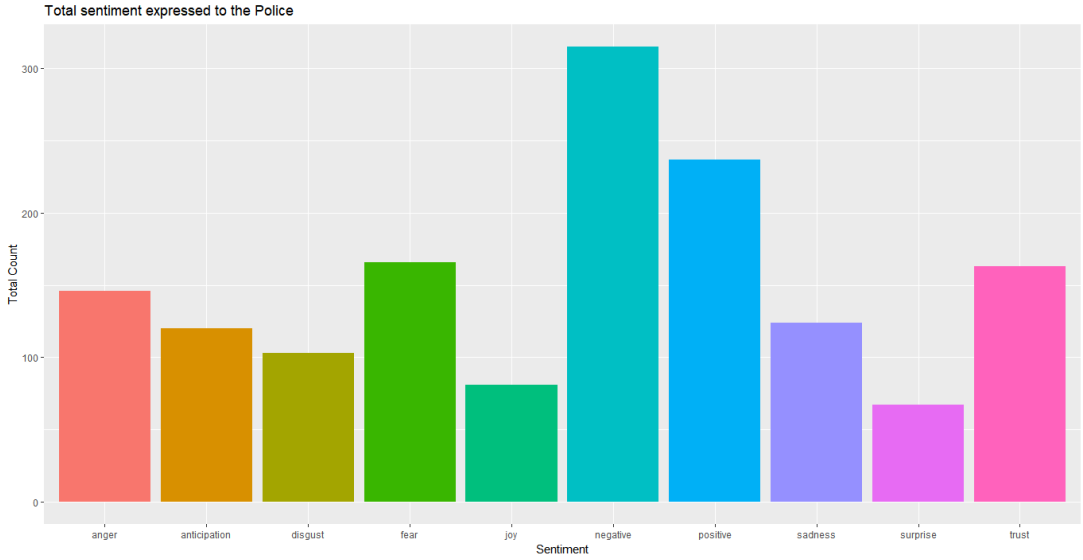

Figure 8.0: Emotional classification of sentiments from @PoliceNG

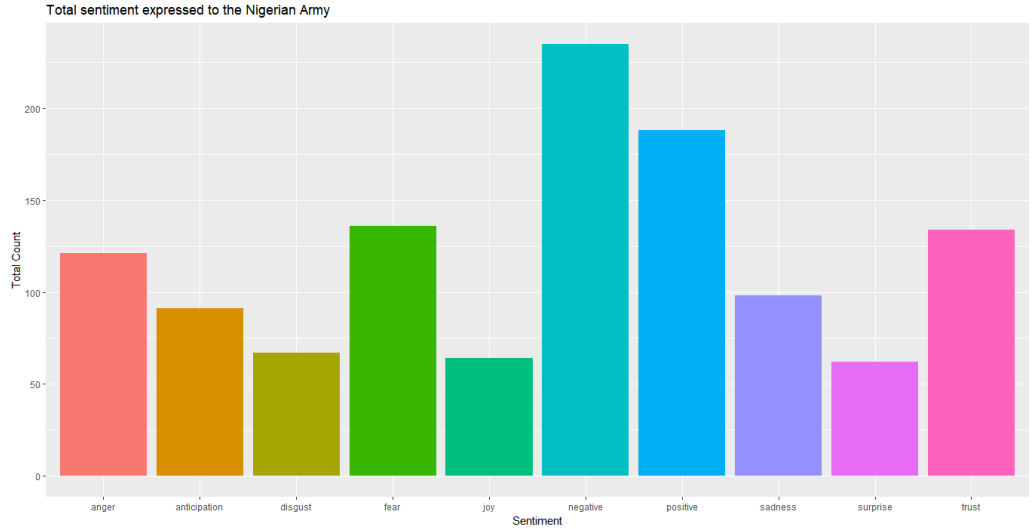

Figure 9.0: Emotional classification of sentiments from @HQNigerianArmy

The experiment had uncovered the actual opinions of the populace concerning both agencies in question. The expression of fear, disgust, anger and sad emotions may be an indication of failure in service delivery to the citizens or a show of conveyed feelings while reporting a case to the security agencies using the online channel. However, there is a show of trust, anticipation and joy towards the agencies in the discharge of their duties.

The aim of this work which is to propose a big data-driven e-government framework for e-governance operations had been met through presenting a conceptualized context and implementing same with a big data experiment. The work had shown that the feedback received through tweets to the Nigerian Police and the Army contains mixed feelings of both positive and negative sentiments. It also shows a higher negative sentiment in essence. This display of evidenced-based feedback is necessary for the government. It may steer policy makers in these security agencies into creation of schemes that would serve the people in meeting their actual security needs.

RECOMMENDATIONS

1. The Nigerian government needs to establish a specialized agency "Big Data Management Authority" saddled with the responsibility of implementing the framework discussed in this paper.
2. A policy direction needs to be made as it regards open data access across government agencies and for individual professionals for data analytics as it is common in US, Canada, Britain, Australia, France, Japan - that has national initiatives on Big Data.
3. The government need to review the Cybercrimes Act 2015 to include the concerns on privacy matters which may arise as a result of implementing an open data policy in the country.
4. There should be a national chief data officer whose role will include implementing the e-government policies and ensuring the integrating of services and citizen feedbacks across all government websites right into the big data e-government proposed framework.

CONCLUSION

This work is a practical demonstration of the data-driven decision making. It has been shown in this paper that big data holds the answer to the question of data explosion in Nigeria. It holds promises for great performance in the e-governance of the country. It has been demonstrated that apart from having data processed and getting information as done with conventional database environments, government needs insights for decision making; and insights cannot be possible without proper and precise analytics of both raw datasets and information. This can only be made possible through the application of big data mechanism across all government channels deploying e-government initiatives.

Author's Contributions: Ogbuju collected the datasets and conducted the technical parts of the research while Taiwo validated the algorithms and visualizations. Ejiofor and Onyesolu supervised the overall execution of the project. All authors contributed in writing, editing and proofreading the manuscript.

Conflict Of Interest: The authors declare that there was no conflict of interest among the authors in whatever form.

REFERENCES

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000): CRISP-DM 1.0 step-by-step data mining guide. Retrieved from https://www.the-modeling-agency.com/crisp-dm.pdf

Fayyad, U. M., Piatetsky-Shapiro, G. and Smyth, P. (1996): From data mining to knowledge discovery: An overview, AAAI/MIT Press.

Fraser-Moleketi, G. and Senghor, D. (2011): E-governance and citizen participation in West Africa: Challenges and opportunities. Panos Institute West Africa (PIWA) and the United Nations Development Programme (UNDP).

Government of Turkey (2015): Turkey's integrated social assistance services system: Recent effort, impressive virtual integration. Workshop on Integrated Data and Information Management for Social Protection: Bridging the Gaps between Theory and Practice, Jakarta, 11-12 March. Retrieved on February 8, 2018 from http://www.opml.co.uk/

Greengard, S. (2014): Analytics tools help China deal with air pollution. Available online at http://www.baselinemag.com/

Hongli, H., L.Q., and Zhangxi, L. (2003): From concept towards implementation: E-Government in China. AMCIS 2003 Proceedings of the AMCIS 2013, 136. Available at: https://aisel.aisnet.org/amcis2003/136

Hilbert, M. and Lopez, P. (2011): The world's technological capacity to store. Communicate and Compute Information Science, **33**(2): 60-65.

Iyilade, J.S. (2015): Big data & analytics opportunities and challenges. Paper presented at the 12th NCS International Conference on Information. Retrieved from http://www.ncs.org.ng/

Kaka, S. (2015): E-government adoption and framework for big data analytics in Nigeria. National Information Technology Development Agency (NITDA). Retrieved on 6 January, 2018 from http://eprints.covenantuniversity.edu.ng

Layne, K., and Lee, J. (2001): Developing fully functional e-government: A four stage model. Government Information Quarterly, **18**(2): 122-136

Liu, B. and Hu, M (2005): Retrieved on 18 November, 2017 from https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon

Mohammad, S. and Turney, P. (2013): Sentiment and emotion lexicons. Retrieved on 16 February, 2019 from http://saifmohammad.com/WebPages/meetsaif.htm

NITDA (2010): The e-government framework for Nigeria. National Information Technology Development Agency (NITDA). Retrieved on 4 February, 2018 from http://nitda.gov.ng/standard-guidelines-and-regulations/

Onwudebelu, U., Fasola, S. and Joseph, O.A. (2016): Big Data in big giant in big continent. Proceedings on Big Data Analytics & Innovation Conference (Peer-Reviewed), **1**(1): 8-20.

Ogbuju, E., Aminu, I., and Musa, A. (2017): Towards a Data-driven smart governance in Nigeria. Zenodo. http://doi.org/10.5281/zenodo.884103

Pearce, R. (2015): DTO works on Whole-of-Government Service Analytics. Computer World. Available online: http://www.computerworld.com.au/.

Thukral, S. (2014): Can BI & Analytics eradicate poverty from India?, iGovernment. Retrieved on 19 February 2019 from http://www.igovernment.in

UNDESA (2016): United Nations E-Government Survey 2016: E-government in support of sustainable development. Economic and Social Affairs. United Nations: New York.

Umair, S., and Haseeb, Q. (2014): A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). International Journal of Innovation and Scientific Research, **12**(1): 217-222

Worldometers (2018): Population of Nigeria (2018 and historical). Available at http://www.worldometers.info/world-population/nigeria-population/

White, T. (2015). Hadoop: The Definitive Guide. 4th edition. O'Reilly Publishers, USA

Wilma, F. S., Nikki, F., Shanna, N., Karen, N., and Maronel, S. (2010): Evidence-based policymaking: a review. South African Journal of Science, 106(5-6).

Yadav, N., and Singh, V. B. (2012): E-Governance: Past, present and future in India. International Journal of Computer Applications, **53**(7): 36 – 48.