



## **BIG DATA ANALYTICS CAPABILITY AND FIRM COMPETITIVE ADVANTAGE: EVIDENCE FROM QUOTED MONEY DEPOSIT BANKS IN NIGERIA**

**Rotimi Olugbohunge** ([rotimiolugbohunge@yahoo.com](mailto:rotimiolugbohunge@yahoo.com))

Computer Science Department,  
Babcock University,  
Ilisan Remo, Ogun State, Nigeria.

**Oludele Awodele** ([awodeleo@babcock.edu.ng](mailto:awodeleo@babcock.edu.ng))

Computer Science Department,  
Babcock University,  
Ilisan Remo, Ogun State, Nigeria.

### **Abstract**

The underperformance of most Nigerian companies with their key performance indicators on the downward trend and the experience of a lower return on investment unlike their counterparts in developed climes cannot be downplayed. While studies on information systems implementation on firms' financial performance are yet to consider other financial indicators like ROA and ROE, studies in Nigeria have not yet discovered the role big data analytics capability can play in scaling up their performance. The aim of this study is to determine the relationship between big data analytics capability and firm performance of money deposit banks in Nigeria. Objectively, the study attempts to ascertain the type of interactions that exist between big data analytics capability and company's performance when firms make investment in big data analytics and also in the development of capability for the use of the analytics. The methodology used was the event study methodology which is often used in information systems research. The duration of the study would be three (3) years. Using social media analytics principle, fifteen (15) banks whose shares are quoted on the Nigerian Stock Exchange will be used. The study intends to use Natural Language Processing (NLP) techniques. Linguistic tools would be used to solve the complexity and ambiguousness of natural language. The source of data will be social media networks, feeds, blogs, online forums, corporate documents website. The study will use the architecture, CROSB, as the platform for data collection, cleaning, and analyses for the study. Statistical techniques, computational linguistics and data learning techniques would also be used to find relationship between big data analytics capability and firm performance in the area of their stock price, Return on Assets (ROA), Return on Equity (ROE), Return on Investment (ROI), and other key performance indicators such as customer churn and staff turnover.

**Keywords:** Big data analytics capability (BDAC), BDA infrastructure capability, Big data management capability, Big data analytics personnel capability, Firm performance, Operational performance

---

## 1.0 INTRODUCTION

Within turbulent and highly competitive global environments, firms are compelled to adapt more rapidly, boldly, and to experiment in order to survive and thrive. They are increasingly seeking ways to identify the constraints in advancing business processes which severely hampers their ability to respond to accelerating competitive demands. According to Bagshaw and Okpakpam (2017), firm competitiveness mean having an edge over competitors and maintaining a large market share by satisfying the customers. In addressing the issue of maintaining competitive advantage. Porter (1998) proposed that the basis of competitive advantage is on the organizational resources such as human resources, knowledge resources, natural resources, capital resources, and infrastructural resources. Du Plessis (2017) posited that the human resource is one of the organization's most effective and valuable resources but it depends on another resource, information. Information is the basis for the acquisition and use of many resources. The effectiveness of a business is at the peak, when the information-processing capacities of the structure fits the information-processing requirements of the output (Griffith, Sawyer, 2010).

It has been established that the information systems (IS) and IT based resources provides firm decision making and improve their competitiveness (Mithas, Lee, Earley, Murugesan, Djavanshir, 2013). As an emerging technology, big data, characterized by numerous data that can be generated, captured, and processed quickly and cannot be categorized into relational databases, contains data that can be stored in a highly structured format to maximize its meaning. However, the translation between structured data in relational systems of database management and unstructured data for analytics can impede end-to-end processing (Chen, Mao & Liu, 2014). Big data requires database management systems with capabilities beyond those in standard SQL-based systems. In the past, data was largely from transactions and was stored as rows and columns. Today, data come from a variety of sources, in different formats, photos from a mobile phone, maps from a GPS device, video from a surveillance camera, audio from a call center, e-mails, tweets, and text messages, and is stored unstructured. All of this data can be digitized, analysed, and stored. It is the manner of this data that constitute the expression called big data (Jiang & Chai, 2016).

Now, with real-time data, we can receive data about events as they are happening, influence the outcome, and prevent bad outcomes before they happen. Additionally, using real-time decisions, organizations can take quick action and gain the needed competitive advantage. Basically, a general understanding on data indicate that data do not reveal their full value until insights are drawn from them. Thus, Big Data is mainly used to analyse insights that lead to certain strategic business moves and better decisions. It involves automating insights into a certain dataset as well as it supposes the usage of queries and data aggregation procedures (Monnappa, 2016). This requires new approaches to obtain insights from highly detailed, contextualized and rich contents that may require complex math operations, such as machine learning or clustering (Chen, Chiang & Storey, 2012). According to Pavithra, Niranjnath, Kamal, Martien (2016), the most widely used definition for big data is in term of three terms like volume, velocity, and variety also referred as 3 –Variety refers to the heterogeneous nature which comprises of structured and unstructured data.

Before an organization can make real-time decisions with big data, it must get data scientists and analytics experts embedded into decision processes. This will require a shift in power from experienced and judgmental decision makers to digital decision makers or experts who generate new insights from big data. These are the same decisions that new insights from big data can improve. But to what extent can big data add value to firms and financial institutions like bank? Big data can predict customer churn or attrition by helping banks understand customer activities that might indicate that the client is considering moving their funds elsewhere by correlating this data and determining the statistical relevance of each activity. Some of the indicators pointing to attrition could be cancellation of preauthorized payments, customer complaints, social media sentiment, and major withdrawals (Somal, 2017). Thus, if banks use customer data, they can gain insights to build effective and targeted promotions by correlating customer purchase history, customer

profile data, and customer behaviour on public social media sites that indicate areas of interest. Sentiment analysis tool helps identify influential customers and engage the right people who are critical to a successful acquisition strategy.

As a likely source of larger-scale digital, unstructured, and finer-grained data from social media, human interactions irrespective of their locations is contributing greatly to huge quantities of data, real-time data, social media analytics, and next generation data management capabilities (Matilda, 2016). However, before such scenario will occur there has to be deployment of big data capabilities and infrastructural platform in the bank. These raises another question whether such financial investment would have value and bring about competitive advantage. Hence, the crucial question being asked: “how will the gain from big data justifies its deployment in firms like banks or will it prove to just another Buzz word?” Research has shown the benefits of using data and analytics in decision making. One study of 179 large publicly traded firms found that companies that have adopted data-driven decision making have output and productivity that is 5% to 6% higher than that of other firms (Brynjolfsson, Hitt and Kim, 2011).

In 2010, the *MIT Sloan Management Review*, in collaboration with the IBM Institute for Business Value, surveyed a global sample of nearly 3,000 executives (LaValle, Lesser, Shockley, Hopkins, and Kruschwitz, 2010). Among the findings were that top performing organizations use analytics five times more than do lower performers, and that 37% of the respondents believe that analytics creates a competitive advantage. A follow-up study in 2011 found that the percentage of respondents who reported that the use of analytics was creating a competitive advantage rose to 58%, which is a 57% increase. Many studies on competitive advantage using big data abound all over the world. For instance, many companies like Google, Twitter and Facebook in processing big data within cloud-environment get the benefit from collecting and analysing data. Amazon can predict what the customer prefers to buy and introduce the products to them according to their preferences. As at 2014, companies such as Bosch and General Electric are putting sensors and microprocessors into all of their products. Bosch has created a central unit, Bosch Software Innovation, which is to lead many of the new digital initiatives (Richard, Dolly, & Fabrizio, 2014).

Competitive advantage cut across financial and non-financial measures like operational performance, marketing performance. The non-financial measures can also include service quality, customer satisfaction, loyalty, market share, productivity, innovation and retention, and employee turnover (Graybeal, Franklin, & Cooper, 2019). All these factors combine as a whole or in part to influence bank competitiveness. Like other industries in the economy, the banking sector has been faced with hyper-competition in recent years. Banks remain the most fragile institution in the financial system, which in an event of financial shocks may affect the functioning of the entire economy due to its intermediation function (Saka, Elegunde & Lawal, 2014). Due to the importance of banks in the financial system and the economy as a whole, scholars, stakeholders, and practitioners have paid serious attention to their financial health (Graybeal, Franklin, & Cooper, 2019). Banks are most active and vulnerable to shocks in the financial system that is why banks performance has remained an area with the highest number of researches in the financial institution.

The key business challenge for Banks is increasing its revenue and margin. Other challenges banks faced is in fraud. According to the Chartered Institute of Management Accountants (2008), the risks of fraud will continue to increase as globalization broadens, more competitive markets emerges, rapid increase in developments, technology, and periods of economic difficulty. Surveys embarked on by CIMA (2008) highlight that organizations may be losing as much as 7% of their annual turnover as a result of fraud. Also, corruption is estimated to cost the global economy in which Nigeria is part of about \$1.5 trillion each year. With growth in the digital realm, the banks have to face a huge challenge of fraud and scams. Staying vigilant to this digital distress is the only possible way to have peace of mind and combat losses. Big data can help in identifying the services that bank customers want and help in customising services to drive new customer demands which can lead to bank performance.

For instance, big data methodology is successfully used by HDFC bank in India, where the customer lifecycle is used to boost credit card activations. Another example is OCBC bank in Singapore leveraged a similar big data methodology and achieved. The main role of BDA is to provide firms the recognition of what is happening now, what is likely to happen next, and what should to be done to get more optimal results (Lavalle, et al. 2011). BDA presents insight of mining hidden patterns to support innovation, more appropriate and real-time decisions, value creation, and subsequently firm performance improvement (Manyika, et al. 2011). This study views BDAC as an important organizational capability leading to sustainable competitive advantage in the big data environment. This study seeks to advance the argument that capability will always lead to sustained competitive advantage among Money deposit banks.

Conversely, big data and cloud computing have some challenges that are impeding its operation. Some of them include data storage, security and privacy, quality, hacking and various attacks, interoperability and portability business, service delivery and billing, availability and reliability, performance and bandwidth cost (Balachandran & Prasad). The Nigerian banking sector had undergone a number of major changes over the last two decades. The major one being consolidation and recapitalization of 2005. These reforms were made with the understanding that it will increase banks' propensity towards risk taking through increases in leverage and off-balance sheet operations (Somoye, 2008). Several forms of reform have been executed in the Nigerian financial system as a way of positioning Nigeria as the financial hub of Africa. This move has brought about financial expansion with large and remarkable banks having an improved financial inclusion in the economy, with payment systems that are effective (Okoi, Ocheni, & Orok, 2019).

The proponents of bank consolidations are of the opinion that the banking sector reform helped banks become stronger players, and in a manner that will ensure higher returns, increased profits/revenue for a variety of reasons including: increase in size, increased product diversification, expanding the pool of potential customers, increased size allowing firms to increase the riskiness of their portfolio. However, evidences show that performance improvements of mergers are seldom realized and as such have not had a positive performance (Ugwunta, 2011). Subsequent events and outcome of previous researches, Odeleye (2014), have suggested mixed outcome of the 2005 banking consolidation exercise in Nigeria thereby calling for further investigation. Past studies have mainly focused on the reform but this study will go beyond reform rules and survey the data capability angle of the reform to determine the banks competitive advantage.

Furthermore, Sharma, Mithas, and Kankanhalli (2014) argue that while there is some evidence that investments in business analytics can create value, the claim that 'business analytics leads to value' needs deeper analysis. For instance, there are inconsistent results between bank internal controls and operational performance in the area of productivity and profitability (Sharma, 2004). Furthermore, Banks all over the world suffer from retention challenges. Faster promotions are one of the means to retain skilled staff but despite this, many organizations are still in danger of losing top employees (Holston-Okae & Mushi, 2018). Likewise, in Nigeria, retention has become one of the biggest issues for the banking industry because employees leave their current organizations due to current employment proposition which is unsatisfactory (Madueke & Emerole, 2017).

One solution to this problem lies in big data analytics that enables a much more efficient ingestion, enrichment, analysis, and visualization of large, diverse and constantly changing data sets so they can be harnessed in the fight against money laundering. Therefore, against this backdrop, this study seeks to prove that money deposit banks that have made investment and improve their big data capability in the area of data management, infrastructure (technology) and talent (personnel) will perform better in terms of competitive advantage than banks that have not made such investment (Kiron et al., 2014).

## 1.2 STATEMENT OF THE PROBLEM

Banks are losing financial assets because of their inability to know when customers are leaving and dissatisfied with them (Somal, 2017). This resulted in stagnating bank revenue and maligned the brand image of the bank. Furthermore, there has been complaint about operational manager's inability to find ways to reduce cost. Although in some cases, quality standards and internal control mechanism effectiveness deployed by banks has been found to reduce cost, and reduce the likelihood of fraudulent financial cases and operational waste. Abbott, Park, and Parker (2000), but in majority of the cases the problem still persist and researchers. Arena & Azzone (2010) are advocating for further studies in that direction. Another problem facing money deposit banks in Nigeria is that as a result of globalisation, economic growth, and shortage of skilled labour, many organizations, including banks, find it difficult to retain valued employees due to factors like improper structural plan, wrong employee engagement, poor compensation package, and imbalance work life, to mention a few (Akintayo, 2010).

In spite of having huge amount of data, the banks are unable to utilize it and get valuable customer insights. However, in order to improve the performance of banks in Nigeria, they can constantly keep collecting data from the front end (trading data), the middle (operations data) and the back-end (finance data) to be used for analytical studies. The banks sit on a huge amount of data which can be unleashed to solve its myriads of problem. Not, mining these data is a massive opportunity loss for the banks to be competitive (Hiba, Mohammed, Issa and Samah, 2017). By using big data capability, banks in Nigeria can use machine learning algorithms will detect changes in digital networks, predict customer churn or attrition, determine customers position in their product life, improve employee retention, marketing performance, operational performance, financial performance and ultimately become competitive.

This study argues that banks in Nigeria that have big data capability and uses machine learning algorithms will detect changes in digital networks, predict customer churn or attrition, determine customers position in their product life, improve employee retention, marketing performance, operational performance, financial performance and ultimately become competitive in the market space. Thus, the purpose of this study is to encourage banks to deploy data to solve the problems they are encountering and to use such insights to gain competitive advantage. By using data that assess banks analytics capability through responses and data derive from data mining in social media complaints, the researcher intends to gain insight and find underlying cause and effect relationships between variables of big data analytical capability and variables of bank competitive advantage.

However, some studies could not establish any underlying mechanisms linking IT to firm competitiveness. A promising approach might be to combine measures of BDA assets with data about data-driven decision-making practices to model the information value chain at a more fine-grained level, starting from the collection and extraction of knowledge from big data via BDA assets over the actual use of this knowledge in decision making to enhanced firm performance. It is in line with this and previously stated problem in the banking industry that this study seeks to determine the role data analytics capability play in listed money deposit bank competitiveness in Nigeria.

## 1.3 OBJECTIVE OF THE STUDY

Drawing on the resource-based view and the literature on big data analytics capability (BDAC), the general objective of this study is to evaluate the connection between big data analytics and firm competitive advantage by focusing on listed money deposit banks in Nigeria.

The specific objectives of the study are to:

- 1 determine the relationship between big data analytics capability and firm financial performance.

- 2 evaluate the relationship between big data analytics capability and firm operational performance
- 3 determine the relationship between big data analytics and firms marketing performance
- 4 explore the relationship between big data analytics capability and firm corporate social responsibility
- 5 assess the relationship between social media data and customer satisfaction
- 6 explore the relationship between social media data and firm productivity
- 7 explore the relationship between social media data and firm service delivery

#### **1.4 JUSTIFICATION OF THE STUDY**

Drawing on the emerging literature on BDA capabilities and IT capabilities, this study proposes BDA capabilities as a third-order, hierarchical model manifested in three second-order constructs – BDA infrastructure capability, management capability, and personnel capability – and eleven first-order constructs: BDA planning, investment, coordination, control, connectivity, compatibility, modularity, technical knowledge, technology management knowledge, business knowledge and relational knowledge. The study argues and justifies this on the ground that BDA capabilities will have a significant impact on money deposit bank's performance based on the principle of resource-based model. The study is justified in that studies in Nigeria have employed data analysis methods like SEM (Structural Equation modelling), regression, and secondary data but coming from a computer science perspective, this study intends to use both earlier methods and data mining technique on social media data.

#### **1.5 SIGNIFICANCE OF THE STUDY**

Big data is skyrocketing due to social networking, the internet, mobile telephony and all kinds of new technologies that create, and capture data. Indeed, banks are swimming in the vast sea of data which basically includes transaction data (e.g., structured data from retail transactions, customer profiles); clickstream data (e.g., web and social media content—tweets, blogs, Facebook wall postings, etc.); video data (e.g., retail and other stores); and voice data (e.g., data from phone calls, call centres and customer service). Gleaning meaningful information from the data will surely translate to areas where banks can gain competitive advantage. Nigeria had the second largest financial sector in Africa in terms of bank assets, market capitalization, and a number of listed companies in the stock market. It follows that improvement in the banking sector will ultimately translate to improvement in the overall economy.

Hence, this study is important to the Nigerian economy. Besides enabling more informed decisions, this study can provide value to organizations by offering new insights and automating business processes. In other words, this study will have contributed to the revolution of business opportunities in Nigeria. Given the development and adoption of AI, machine learning technology, and big data, though still in its early phase, such that there are many issues and these systems need a lot of human management for now, however, this is likely to change in the near future owing to study such as this. This study will be of great benefit to the management of financial institutions, especially in the banking sector. It will provide insight on the best path to take when applying big data analytics as a measure to drive organizational competitiveness in the areas of operational performance, productivity, service delivery, productivity, social responsibility, marketing performance, and financial performance. In addition, the findings from the study would help money deposit banks to develop a better understanding on how to install big data platforms and IT infrastructure as well as the various capability needed.

Furthermore, internal auditors will benefit from the study given the fact that they work hand in hand with the management of the banks to install and enforce internal control system within the bank. Thus, the findings of this study might expose internal data analyst personnel deficit in banks and inefficient techniques or operational mechanism put in place that hinders bank competitiveness. This study will also contribute to existing literature on big data analytics capability and bank performance. Other researchers

will also build on the findings of this study to carry out further study on the phenomenon highlighted. In the long run, this study will constitute part of periodic and continuous evaluations and review series in the advancement of computer science.

## 1.6 SCOPE OF THE STUDY

The study is based on the data analytics capability of money deposit banks listed in Nigeria stock exchange and how it influences their competitive advantage in the areas of service delivery, productivity financial performance, marketing, corporate social responsibility. The study will make use of data from social media and primary data obtained from questionnaire as well as banks financial performance. Fifteen banks listed in NSE (Nigerian Stock Exchange) will be used and data from Social Media Platform were collected, stored, cleansed, and analysed through social media analysis method or social web mining. Hadoop, a free and open-source software, although written in Java, was chosen as the Big Data framework for this. It supports other programming languages through its streaming module.

For this study, only 500MB of data will be obtained from social media owing to time, financial and technological constraint. Additionally, statistical algorithm will be developed to ensure that the data collected are free from noise and form a representative sample. Data collected through questionnaire would be limited to banks corporate offices in Lagos state. The limitation of the study is that respondents' may be reluctant to give extensive information relating to the research project. This is because some bank managers will not be pleased if their staff reveal to the public vital information. However, irrespective of these limitations, an in-depth study will be carried out.

## 2.0 LITERATURE REVIEW

Fredriksson, Mubarak, Tuohimaa, and Zhan (2017), described big data with reference to its characteristics. The study covered a limited amount of data base. The study focuses on understanding big data. However, it did not address any practical use of big data in the public sector. Müller, Fay, Vom (2018), carried a study that aimed to find the presence of a positive link between data-driven innovation and productivity growth across the economy. The study made use of a unique panel data set that contains detailed information about BDA solutions owned by 814 companies during the timeframe from 2008 to 2014. The study used the Cobb-Douglas production function framework to measure the marginal effect of BDA on firm output after accounting for various firm inputs (i.e., labour, capital, materials, IT and the results show a positive and significant relationship between BDA and firm productivity, suggesting that live BDA assets are associated with a 4.1% increase in firm productivity. However, study focused on technical BDA assets, that is, ownership of hardware and software licenses, and did not explicitly measure other types of BDA assets, such as, BDA-related human resources or management capabilities and the dataset includes only companies that adopted BDA solutions from one particular vendor was used.

Ayeisha, Francisco, and Angel (2017) analysed the effects of Big Data Analytics on the Competitive Advantage and Performance of different medium to large size businesses that use Big Data Analytics as part of their operations. A critical analysis was conducted using 20 published papers from Information Systems Journals ranging from the years 2011 to 2016. The papers analysed, examined Competitive. The result of shows that Businesses using Big Data as part of their daily operations have demonstrated moderate to significant improvement in their Competitive Advantage and Performance. Findings indicated that businesses should consider using Big data as part of their operations, as it will improve their Competitive Advantage and Performance. However, the study did not make use of a standard database.

Raguseo and Vitari (2018), examined the forms of business value that companies can create from big data analytics investments, the direct impacts it has on the financial performance of a firm, and the mediating effects of market performance and customer satisfaction. The results show that the business value of big data analytics solutions is able to explain 62.4% of the variance of customer satisfaction, 71.9% of the

variance of market. The study has had a cross-sectional research design, in which all the measurement items were collected at the same point of time. A longitudinal study could extend this research by capturing the dynamics of the business value of big data analytics solutions on different firm performances. Second, this research has employed one data collection method. Objective data from multiple sources could be used to further verify the proposed research mode.

Wamba, Gunasekaran, Akter, Ren, Dubey, and Childe (2017) investigated whether BDA capabilities measured is linked with firm performance? And to determine the effect of process-oriented dynamic capabilities (PODC) mediating role in the relationship between BDAC and FPER. The study used an online survey to collect data from 297 Chinese IT managers and business analysts with big data and business analytic experience. The findings confirmed the value of the entanglement conceptualization of the hierarchical big data analytics capability (BDAC) model, which has both direct and indirect impacts on firm performance (FPER). The results also confirm the strong mediating role of PODC in improving insights and enhancing FPER. Their study adopted perceptual performance measures, which could be replaced by objective measures to present a concrete picture of BDAC's impact on firm performance.

Nyeadi, Ibrahim, and Sare (2018) investigated empirically the impact of corporate social responsibility (CSR) on Financial performance in South African listed firms. The paper uses panel corrected standard errors to estimate the effect of CSR on firm financial performance and thus addresses contemporaneous cross correlations across the panel cross sections. The study uses a broad base measure of CSR created by the Public Investment Corporation data set and the combination of accounting and economic means of measuring firm financial performance. CSR is found to have a strong positive impact on firm financial performance in South Africa. When CSR is decomposed further into its major components, governance performance positively impacts a firm's financial performance with no evidence of any relationship between social components and firm performance and between environmental components and firm performance, the positive impact of CSR on firm performance is greater in big firms. At the industry level, CSR is noticed to impact positively on financial performance in the extractive industry via good governance and responsible environmental behaviours. It however has no impact on firm performance in the financial sector. The survey was carried out on selected firms on the Johannesburg Stock Exchange for three years spanning from 2011 to 2013).

Lee, Kweon, Kim, and Chai (2017) investigated the changes in stock price caused by announcements on big data analytics solution investment. A total of 54 investment announcements of firms publicly traded in NASDAQ and NYSE from 2010 to 2015 were collected. The results empirically demonstrate that announcement of firms' investment on big data solution leads to positive stock market reactions. In addition, they also found that investments on small vendors' solution with industry-oriented functions tend to result in higher abnormal returns than those on big vendors' solution with general functions. The limitation of our study is the small number of samples. As big data analytics is a new and emerging technology with a short period of use the research cannot collect a sufficient. Number of announcements during the past six years. To overcome this limitation, future research would focus on measuring long-term effects of adopting big data analytics solutions by taking return on asset (ROA) and return on investment (ROI) of firms into account.

Gathogo and Ragu (2014) carried out a study was to investigate the influence of ICT and value chain analysis on the competitiveness of manufacturing firms in Kenya. The survey was cross sectional in nature since it covered a sample of 24 firms in various segments of the manufacturing industry in Kenya. The research employed sampling technique and this is validated by the vast number of firms in the manufacturing industry. Findings indicate that a large number of firms sampled used some element of technology in managing their value chains and supply chains. However, there are no clear cut ICT frameworks that specifically target the impact of ICT on the value chain and how the members of the public through social media interactions influences firm performance.



Sudhahar, Veltri, and Cristianini (2015) sought to determine how Automated analysis of the US presidential elections using Big Data and network analysis can predict relationships. The study employed a text analysis method. They analyse texts by an automatic identification of 'semantic triplets', formed by key actors, objects and their relationships. First, we observe that the fundamental split between the Republican and Democrat camps can be easily detected by network partitioning, which provides a strong validation check of the approach adopted, as well as a sound way to assign actors and topics to one of the two camps. The study was employed in the political sphere and was limited to the USA presidential election and politics. Future studies can be employed in firm performance.

### **3.0 METHODOLOGY**

The chapter focused on the methodology that will be used for the study. The section presents the two approaches used in this study. On one hand, it shows how a piece of data posted on a Social Media Platform is collected, stored, cleansed, and analysed including the technologies and frameworks that would be used. In doing this social media analysis method or social web mining, a sub-set of data mining will be used. On the other hand, it includes: the research design, the population sampling technique and sample size determination, method of data collection, research instrument pilot testing, validity and reliability of instrument, method of data analysis, model specification, conceptual model, ethical considerations and limitation of the research methodology.

The two-pronged approach is chosen because in other to explore thoroughly big data technology and capability in money deposit banks in Nigeria. Using just one of the methods will not sufficiently address the gaps as they exist in Nigeria. Due to the massive work, it entails data that would be obtained from social media that would be limited to 500MB owing to time, financial and technological constraint. However, statistical algorithm will be developed to ensure that data collected are free from noise and form a representative sample.

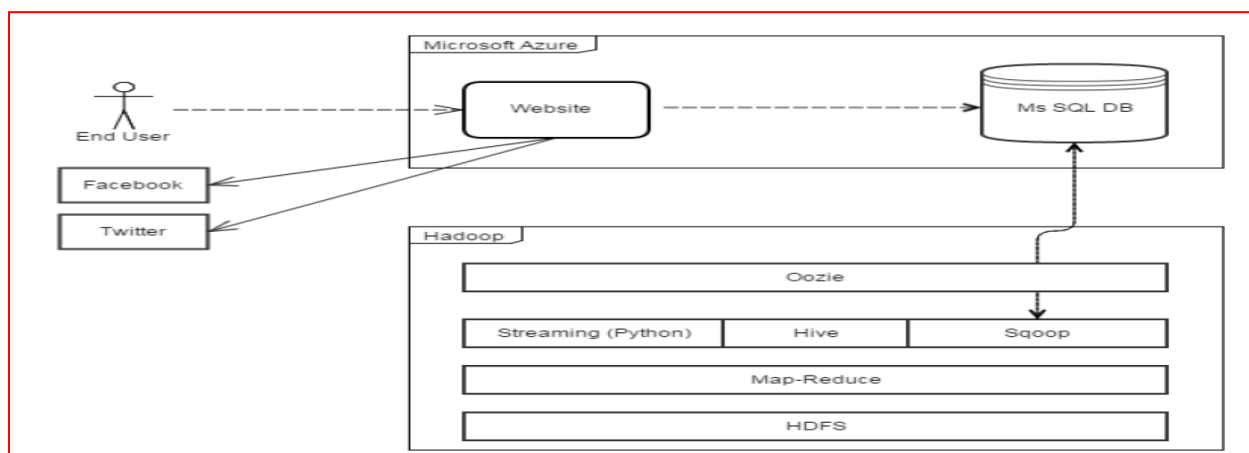
#### **3.1 RESEARCH DESIGN**

The design that will be adopted for this study is survey design and data mining based on positivist research approach. Survey research design is conducted about a research problem when there are few or no earlier studies to refer to or rely upon to predict an outcome. This study adopted the questionnaire-based survey method because it captures causal relationships between constructs and hence provides generalizable statements on the research setting. Moreover, surveys can accurately document the norm, identify extreme information, and delineate associations between variables in a sample. Survey research is used for explanatory and predictive theory in order to ensure greater confidence in the generalizability of the results.

Using the positivist research approach, the study captures the objective and social reality by survey measures to identify the BDA capabilities in order to address the research questions. As part of this approach. For the purposes of this research, social media analysis will be used to extract the human generated data on social media platforms. The platform that would be use to analyse user data is Twitter and Facebook. The bulk of the analysis will be performed using a technique called machine learning. Machine learning will make up an extensive part of this research as it will be used to analyse each piece of data and by a given set of criteria, it will try to give a score to each piece of data received. At the end of the research, a score will be awarded to each user. This score will be the average score of all their data that has been analysed. The machine learning code will apply numerous algorithms to each piece of data. Each piece of data will be broken down into sentences and other algorithms will evaluate the semantics and syntax of each sentence, before the sentence is parsed into words. Each word will then carry a particular weight depending on the length, how often it used, and how specific the word is.

The main components needed for this research are a front facing website which is exposed to an end user, external social platforms, and a Big Data framework. The architecture is Social Platform which is an external entity to the solution, the website itself to gather the data, and the big data framework to store the data. Hadoop was chosen as the Big Data framework for this research to analyse the data. As a free and open source, and written in Java, Hadoop is an ecosystem of components that work together to create a Big Data framework. The main component of Hadoop is the distributed file system called Hadoop Distributed File System (HDFS), but there are other components can store data instead like Hive; it is designed to work on commodity hardware to provide high fault tolerance. The file system stores the application data across multiple data nodes and provides high throughput to the application data. Map-Reduce is a software framework for working with large data sets across a distributed network.

Although, Python, Perl and Ruby are all supported, using the Streaming utility of Hadoop increases the time for each job to complete compared to Java. The main advantage of using a scripting language and particularly Python, is that there is a toolkit called the Natural Language Toolkit (NLTK) which provides a helpful platform to interact with human language. Another sub platform is Apache Sqoop. Apache Sqoop is a tool which can import or export from Hive or HDFS to a relational database. Sqoop has a connector for SQL Server databases and can be easily configured to connect to a SQL Server database running in Azure.



**Figure 3.1 Study Architecture Source: Burke (2015)**

### 3.2 POPULATION

The survey population comprises of five hundred (500) IT managers, marketers and financial officers of banks. Also, Respondents that possesses distinguishing academic, IT, financial understanding on banks operations and performance will be engaged.

### 3.3 SAMPLE SIZE AND SAMPLING TECHNIQUE

For the primary data, Purposive sampling technique will be adopted for this study.

To determine the sample size for this study, Krejcie and Morgan formula for calculating sample size for finite population will be adopted.

$$S = \frac{X^2 NP(1-P)}{d^2(N-1) + X^2 P(1-P)}$$

Where:

- S = Required Sample size
- X = Z value (that is, 1.96 for 95% confidence level)
- N = Population Size

P = Population proportion (expressed as decimal) (assumed to be 0.5 (50%))  
d = Degree of accuracy (5%), expressed as a proportion (.05); It is margin of error.

### **3.4 METHOD OF DATA COLLECTION**

#### **3.4.1 DATA COLLECTION USING QUESTIONNAIRE**

Questionnaire was used as the primary source of data collection method because it enabled the researcher to gather first-hand information from bank. Basically, the primary source of data provided for this study are reliable, dependable, based on direct-response and feedback, and risk-reduced for ingenuity of findings. And the rest of the needed data will be gotten from Facebook and Twitter. Furthermore, 40 copies of the questionnaire will be distributed to the two banks: Union Bank and GTBank in Lagos State. The data obtained would be used in gauging the validity and reliability of the instrument.

#### **3.4.2 DATA COLLECTION USING DATA MINING**

Before data can be collected from a Social Media Platform, an application on each platform to obtain an API key and a secret key will be created. Thereafter, a website to retrieve the data from each API will be created. Making a request with both the API key and the secret key created by the social media platform, the social media platform will provide an access token that will be sent with each request as part of the header to obtain a user's data. The social platform twitter provides public access to tweets on their site. Request for tweets are made with an access token through the platform API. Facebook on the other hand do not provide public access to status updates or posts. Special permissions are required, and a user must first authorize an app with those permissions before it is possible to retrieve them. Since this is complicated owing to the fact that Facebook app will not receive permissions for data. The study will not analyse Facebook data.

An admin section to the end user website will be developed. This will create an account on behalf of a social media user. The information needed is the users Social Platform id which will enable request of latest posts/tweets, previous posts/tweets of users, and update the users profile e.g. profile picture. The admin section makes it easy to update the ever-growing database of data while using a smart phone, instead of having to load up a laptop with Visual Studio to access the API. The data from Social Media will be stored in the MS SQL DB in Azure with an associated user id foreign key. This database is a great place to persist the data and display on the website when requested, however it is not the optimum place to analyse the data. Hadoop Big Data Framework would be used to analyse the data.

For the purposes of this research, a pre-configured single-node cluster would be used from Cloudera. The single-node cluster works with VirtualBox from Oracle and runs on the RedHat operating system. The Sqoop component will be used to import or export data in Azure into HDFS in the VM database from an external source into Hadoop, or from Hadoop to an external source. The MS SQL connector available in Microsoft website called Microsoft JDBC Driver 4.0 for SQL Server would be used. Each column would be enclosed in quotes in the sqoop to handle escape characters and enclose strings. Sqoop also would be used to handle social media data that contains quotes.

The text posted on social media platforms do not always contain words found in a proper English dictionary. For this reason, data cleansing must happen before the data is analysed. The collected data will be cleaned to remove any meaningless or noisy data from each input. Meaningless data may refer to any piece of information which does not provide any purposeful information to the text. Noisy or meaningless data may also be data which cannot be interpreted or understood.

Since the data is stored in HDFS in the Hadoop cluster, Python is the programming language of choice to cleanse and process the data. A Map-Reduce program will be used for this. The data cleansing process will be executed at the beginning of the mapper.

A code will be written that will simply try to remove all links, hashtags and mentions using regular expressions. After that, the modified text then tries to remove most popular emoticons. The next step in the cleansing process is to replace certain characters found. When the data is imported in HDFS, the ampersand '&' character is commonly seen as the character reference value or "&#amp;" This value occurs very often so it is replaced with the word 'and'. Once the text has been cleansed, it is decoded to 'utf-8' and tokenised into sentences using NLTK library.

### 3.7 DATA ANALYSIS

#### 3.3.1 ANALYSING SOCIAL MEDIA DATA

In analysing the data, the social media data will be split into two processes: the mapper and the reducer program.

**Mapper:** The mapper is a python script that runs across the entire data set. The output of this script would be passed into the reducer. The mapper contains most of the logic for the data analysis step. The mapper would read each line in HDFS which is essentially a CSV format of the MS SQL database in Azure. The line will then be parsed to extract the social media text column, the user id and the line id or primary key. After the data has been parsed, the data would be cleansed. Furthermore, a check would be made if the data is an original post or not. A Twitter retweet will not be analysed because it was posted by another user so it cannot be awarded a score to the current user. The program does a quick check to find "RT". If this is present, the programs just move onto the next line, otherwise this line is analysed. The text would be tokenised into sentences using the NLTK framework. Each sentence will be analysed separately and awarded a score. The score for each sentence is summed at the end to award a score to the text.

After this stage a mapper would iterate over each sentence. The sentence is then tokenised into words using the NLTK framework. Each word is checked against a CSV (Comma Separated Values) file containing a list of acronyms and their full text meaning. If an acronym is found, it is replaced with its full text meaning. This process is used to provide a better-quality outcome, as the NLTK framework does not understand acronyms. Once the acronyms have been replaced, the NLTK framework would be used again to assign each word a part-of-speech tag (POS tagging). This process may also be known as lexical categories or grammatical tagging. A context free parser is used and each word is awarded a tag based on its definition. The tags signify whether a word is a noun, adjective, verb or any English part of speech type. A score is awarded to each work depending on which tag it has. The code written for the mapper would be used. The mapper code would output the user's id, the primary key of the line, the number of words in the text, the score awarded to the text and a counter. These values are tab separated, and will be fed into the reducer.

**Reducer:** The role of the reducer is to take the input from the mapper and perform an action on the data, and output the result. There are two reducers written in Python used within this research, which means there are two separate Map-Reduce jobs. The mapper stays the same for both, but the reducer is slightly different. The first reducer very simply outputs the primary key of each row along with the score value. This is picked up by Sqoop to perform an export which will update the MS SQL database running in Azure. The second reducer gets the average of every unique user's score by the user id value sent on by the mapper and summing up all the score values. The reducer also sums up the counter value from the mapper, and once the reducer has finished reading each line, the score value for each user is divided by the counter. This will provide the average score for the number of social media data the mapper has processed. The output of the reducer would print out the user id, the average score and the number of online posts analysed.

#### 3.3.2 ANALYSING THE DATA FROM SURVEY

A Confirmatory factor analysis using PLS-SEM would be used. In order to assess the higher-order BDA capabilities model, the study would apply partial least squares based on structural equation modelling (PLS-SEM). This is used because it estimates hierarchical models by removing the uncertainty of inadmissible solutions using its flexible assumptions (Joe, Sarstedt, Ringle & Mena, 2011). PLS-SEM is employed because it ensures greater theoretical parsimony and less model complexity to estimate the hierarchical model (Wetzels, Odekerken, & Oppen, 2009).

#### **4.0 EXPECTED OUTCOMES**

It is expected that banks with big data analytics capability would have competitive advantage over their peers in terms of their stock price, Return On Assets (ROA), Return on Equity (ROE), Return On Investment (ROI) and other key performance indicators. They will also have less customer churning and less staff turnover which will translate to better performance for the institution. In developing and hosting the website a Microsoft .NET MVC website will be used along with Visual Studio 2013 on the .NET 4.5.1 framework and the C# programming language. The website will be split into separate projects following the SOLID principles for best practice. The website will be deployed to Microsoft Azure, a cloud platform that can host a variety of programming languages, frameworks and databases. Along with the website, the Microsoft SQL database is also hosted in Azure.

This project will be built using Entity Framework and is separated into a different project because multiple projects in this solution utilise it. The SocialIQ Clients projects contains a client for each Social Media Platform used. To make it easier to interact with Twitter, a .Net library called CoreTweet will be used. This library is a wrapper around the Twitter API to make it much faster to get an application up and running. It is licensed using the MIT License. The Social platforms used here is a website and application which allows a user to hold an account and interact with. Common interactions may be posting messages, which is commonly known as status updates on Facebook or tweets on Twitter as the two of the best examples of social platforms are Facebook and Twitter and would be used in this study. They are popular, have more content and have API to utilise. An API or Application Programming Interface is a set of functions or methods made available by an application to allow external applications interact with it. In this case, both Facebook and Twitter provide an API which will allow an application to request data.

#### **5.0 CONCLUSIONS AND RECOMMENDATIONS**

The paper proposed to determine the relationship between big data analytics capability and performance of money deposit banks in Nigeria. Specifically, the study attempted to determine the type of interactions between investment in big data analytics and the capability to use it by a bank on one hand and the performance of the bank in comparison to its peers. It is expected that banks with big data analytics capability would have competitive advantage over their peers in terms of their stock price, Return On Assets (ROA), Return on Equity (ROE), Return On Investment (ROI) and other key performance indicators. They will also have less customer churning, less staff turnover, and positive image perception by the public which will translate to better performance for the institution.

Future studies or parallel studies can be conducted by focusing on the analysis of everyday spoken language acquired through mobile phones and emerging technologies such as Google Glass. Current social media captures a part of people's everyday thoughts and feelings; phone and mobile data can capture even more. Mobile devices have the additional advantage of providing context through motion and location sensors. There is need for studies on real data. Real Time Big Data Analytics attempts to analyse constantly changing data – streaming data. Events/information is continuously fed to the system and is analysed based on insights from already collected data. There is also a need to build real time real data to assess the effect of social media negative comment on firm decisions.

## References

- Abbott, J. L., Park, Y., and Parker, S. (2000). The effects of audit committee and independence on corporate fraud. *Managerial Finance* 26(11):55-68. DOI:10.1108/03074350010766990
- Akintayo, D. (2010). Influence of Emotional Intelligence on Work-Family Role Conflict Management and Reduction in Withdrawal Intentions of Workers in Private Organizations. *International Business & Economics Research Journal (IBER)*, 9(12). <https://doi.org/10.19030/iber.v9i12.354>
- Arena, M., Arnaboldi, M. and Azzone, G. (2010) The Organizational Dynamics of Enterprise Risk Management. *Accounting, Organizations and Society*, 35, 659-675.
- Augusto (2007). Banking Industry Report. Lagos state.
- Ayeisha, C Francisco J. R. and Angel, O. (2017). Big data analytics, competitive advantage and firm performance. *International Journal Of Information Research And Review*, 04(2), 3599-3603
- Bagshaw, K. B., & Okpakpam, J. A. (2017). Workforce Big Data Analytics And Production Efficiency: A Manager's Guide. *Archives of Business Research*, 5(7), 121-131.
- Balachandran, B. M. and Prasad, S. (2017). Challenges and Benefits of Deploying Big Data Analytics in the Cloud for Business Intelligence. *International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, Marseille, France.*
- Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: how does data-driven decision-making affect firm performance? Cambridge: Working Paper, Sloan School of Management, MIT.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *Management Information Systems Quarterly*, 36(4), 1165-1188.
- Chen, M., Mao, S., and Liu, Y. (2014). Big data: a survey. *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171-209.
- Du Plessis, M. (2017). The role of knowledge management in innovation. *Journal of Knowledge Management*, 11(4), 20-29.
- Edwards, R. J. (2001). Multidimensional Constructs in Organizational Behavior Research: An Integrative Analytical Framework.
- Fredriksson, C., Mubarak, F., Tuohimaa, M., and Zhan, M. (2017) Big Data in the Public Sector: A Systematic Literature Review. *Scandinavian Journal of Public Administration*. Vol 21, No 3.
- Gathogo, G. and Ragu, M. (2014). Capital Structure of Kenyan Firms: What determines it? *International Knowledge Sharing Platform*. Vol 5, No 5. Retrieved from: <https://www.iiste.org/Journals/index.php/RJFA/article/view/11418>.
- Graybeal, P., Franklin, M., and Cooper, D. (2019). Describe the Balanced Scorecard and Explain How It Is Used. *Principles of Accounting, Volume 2: Managerial Accounting*, Ch. 72. Retrieved from: <https://opentextbc.ca/principlesofaccountingv2openstax/chapter/describe-the-balanced-scorecard-and-explain-how-it-is-used/>

- Griffith, T. L. and Sawyer, J. E. (2010). Multilevel knowledge and team performance. *Journal of Organizational Behaviour*. Wiley Online Library.
- Hiba, A., Mohammed, A., Issa and Samah A. (2017). The importance of big data analytics in business: A case study. *American Journal of Software Engineering and Applications*, 6(4), 111 -115
- Holston-Okae, B. L. and Mushi, R. (2018). Employee Turnover in the Hospitality Industry using Herzbergs Two-Factor Motivation-Hygiene Theory. DOI:10.6007/IJARBSS/v8-i1/3805. Retrieved from: [https://www.researchgate.net/publication/325121955\\_Employee\\_Turnover\\_in\\_the\\_Hospitality\\_Industry\\_using\\_Herzbergas\\_Two-Factor\\_Motivation-Hygiene\\_Theory](https://www.researchgate.net/publication/325121955_Employee_Turnover_in_the_Hospitality_Industry_using_Herzbergas_Two-Factor_Motivation-Hygiene_Theory)
- Jiang, W. and Chai, H., (2016) July. Research on big data in business model innovation based on GA-BP model. In *Service Operations and Logistics, and Informatics (SOLI)*, 2016 IEEE International Conference on (pp. 174-177). IEEE.
- Joe, F. H., Sarstedt, M., Ringle, C. M., and Mena, J. A. (2011). An assessment of the use of partial least squares structural equation modelling in marketing research. *Journal of the Academy of Marketing Science*. Volume 40, Pages 414–433
- Kiron, D., Prentice, P.K., & Ferguson, R.B. (2014). The analytics mandate. *The 2014 Data & Analytics Global Executive Study and Research Report*.
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*, 52(2).
- Lee, H. Kweon, E., Kim, M., and Chai, S. (2017). Does Implementation of Big Data Analytics Improve Firms' Market Value? Investors' Reaction in Stock Market. *Sustainability* 9(6):978. DOI:10.3390/su9060978
- Madueke, C. V. and Emerole, I. C. (2017). Organizational Culture and Employee Retention of Selected Commercial Banks in Anambra State. 2.3.16 244. ISSN 2415-6663 (Print) Scholars Middle East Publishers ISSN 2415-6671 (Online) Dubai, United Arab Emirates.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, J., Roxburgh, C., and Byers, A. H. (2011) *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute.
- Matilda, S. (2016). Big Data in Social Media Environment - A Business Perspective. Retrieved from: [https://www.researchgate.net/publication/316665818\\_Big\\_Data\\_in\\_Social\\_Media\\_Environment\\_A\\_Business\\_Perspective/link/5b8b68b592851c1e1241ee8b/download](https://www.researchgate.net/publication/316665818_Big_Data_in_Social_Media_Environment_A_Business_Perspective/link/5b8b68b592851c1e1241ee8b/download)
- Mithas, S., Lee, M. R., Earley, S., Murugesan, S., & Djavanshir, R. (2013). Leveraging big data and business analytics. *IT Professional*, 15, 18–20.
- Monnappa, A. (2016). Data Science vs. Big Data vs. Data Analytics. Retrieved March 1, 2017, from Simplilearn: <https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article>.
- Mosavi and Vaezipour (2013). *Developing Effective Tools for Predictive Analytics and Informed Decisions*, Technical Report, University of Tallinn.
- Müller, O., Fay, M., Vom Brocke, J. (2018)'. The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of Management Information Systems*, 1(10), 1 – 40.

- Nyeadi, J. D, Ibrahim, M., Awudu Sare, Y. A. (2018). Corporate social responsibility and financial performance nexus: Empirical evidence from South African listed firms. *Journal of Global Responsibility* 9(4). DOI:10.1108/JGR-01-2018-0004
- Odeleye, A. T. (2014). Pre-Consolidation and Post-Consolidation of Nigerian Banking Sector: A Dynamic Comparison. *International Journal of Economics and Financial Issues, Econjournals*, vol. 4(1), pages 27-34.
- Okoi, I. O., Ocheni, S. I., and Orok, A. B. (2019). Impact of Banking Sector Reforms on Economic Growth in Nigeria. *European Journal of Scientific Research*. ISSN 1450-216X / 1450-202X Vol. 154 No 2. pp. 230-240. Retrieved from: <http://www.europeanjournalofscientificresearch.com>
- Pavithra B, Dr. Niranjnmurthy M, Kamal Shaker J, Martien Sylvester Mani F. (2016). The Study of Big Data Analytics in E-Commerce. *International Journal of Advanced Research in Computer and Communication Engineering*, 5, 2, 126 – 131
- Porter, M. E. (1990). The competitive advantage of nations. *Harvard Business Review*. p. 73–93.
- Raguseo, E., & Vitari, C. (2018). Investments in big data analytics and firm performance: an empirical investigation of direct and mediating effects. *International Journal of Production Research*, 1-16.
- Richard, M. B., Dolly, M. and Fabrizio, S. (2014). *Journal of Organization Design*. JOD, 3(1): 1 - 10
- Saka, E. and Lawal, (2014). Effects of Customer relationship Marketing on Bank Performance in Nigeria. An Empirical Approach *European Journal of Business and Management*. Vol 6, No. 17.
- Sanni H. O. ( 2015 ). How to Improve Efficiency of Banking System with Big Data (A Case Study of Nigeria Banks). *International Journal of Science and Research*, 78(96), 2319-7064
- Sharma, P. (2004). An Overview of the Field of Family Business Studies: Current Status and Directions for the Future. Retrieved from: <https://journals.sagepub.com/doi/10.1111/j.1741-6248.2004.00001.x>
- Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*, 23, 433–441 }
- Sharma, R., Mithas, S., and Kankanhalli, A. (2014). Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*.
- Somal HK (2017). Big Data & Analytics: Tackling Business Challenges in Banking Industry. *Bus Eco J* 8: 305.- 345
- Somoye, R. (2008). The Performance of Commercial Banks in Post Consolidation Period in Nigeria. An Empirical Review. *European Journal of Economics, Finance and Administrative Science*, 14, 62-73.
- Sudhahar, S., Veltri, G. A., and Cristianini, N. (2015). Automated analysis of the US presidential elections using Big Data and network analysis. Retrieved from: <https://journals.sagepub.com/doi/full/10.1177/2053951715572916>



- Ugwunta, D. O. (2011). The Effect of Bank Consolidation on Bank Performance: A Case Study of the 2005 Concluded Nigerian Bank Consolidation Exercise. University of Nigeria, Nsukka.
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. Ji-fan., Dubey, R. & Childe, S. J. (2017). Big data analytics and firm performance: effects of dynamic capabilities. *Journal of Business Research*, 70 356-365
- Wetzels, M., Odekerken. G., and Van Oppen, C. V. (2009). Using PLS Path Modeling for Assessing Hierarchical Construct Models: Guidelines and Impirical Illustration. *MIS Quarterly* 33:177-195.

© GSJ