Information and Knowledge Management www.iiste.org
ISSN 2224-5758 (Paper) ISSN 2224-896X (Online)
Vol.7, No.12, 2017

Application of Statistical and Mathematical Algorithms to Data Analytics and Job Creation in Nigeria[1]

David E. Omoregie[1*] Moses K. Tule[2]
1.Principal Economist, Monetary Policy Department, Central Bank of Nigeria
2.Director, Monetary Policy Department, Central Bank of Nigeria

Abstract
In this paper, we examine the use of statistical and mathematical algorithms in data analytics and their application in business intelligence, insights and collective intelligence, for enhanced job creation interventions in Nigeria. The paper argues that the demand-driven job creation, involving developing skills for existing vacancies or opportunities is no longer sustainable in the current challenging economic conditions. Rather it makes a case for supply-driven job creation, where skills are developed in technology and data analytics (with strong reliance on statistics and mathematics), with a view to solving business and corporate problems, thereby enhancing job creation in those businesses and corporations, which hitherto had no vacancies. The paper surveys statistical and mathematical algorithms, categorized as supervised and unsupervised learning techniques, applied in data analytics, and discusses the emerging requirements for data analytics in modern business and corporations. It further discusses modern application of data analytics in a number of business areas such as marketing, customer management, finances, data mining, web and learning, highlighting a number of metrics specific to each sector. The paper also identifies the specialized skills required to create job opportunities in key sectors in Nigeria. Drawing extensively from the lead author's experience in the UK, the paper presents how skills in modern data analytics can lead in creating job opportunities, a major lesson for Nigeria.
Keywords: Job Creation, Data Analytics, Data Science, Business intelligence, Insights, Algorithms

1. Introduction

Job creation is a major macroeconomic objective of modern governments, in the face of rising unemployment. In today's world, to create jobs and reduce unemployment would require unconventional methods. One major method that holds out great promise for job creation, especially in the corporate sector, is the application of statistical and mathematical algorithms in data analytics. Data analytics is the application of analytical techniques (including statistics and mathematics) to data with a view to understanding patterns, causes and effects of phenomena, and drawing meaningful inferences and insights to aid decision-making. It's basically a data crunching and interpretation exercise aimed at revealing patterns and structures, to derive intelligence from the data. *Data Analytics* is "*a range of techniques and processes for the collection, classification, analysis and interpretation of data to reveal patterns, anomalies, key variables and relationships, leading ultimately to new insights and better answers faster* (Sabherwal, and Becerra-Fernandez, 2010)." The key point in that definition is generation of "insights and better answers faster".

The last two decades have witnessed increased prominence of data analytics and data science as a job sector in many economies, especially amongst the industrial nations. This is primarily due to a number of developments, including the success stories of major companies such as Google, eBay, Amazon and Netflix. Google's popular page rank software, which uses over 200 metrics in search results composition, earning billions of dollars annually, is based on a sophisticated system of algorithms used in optimizing user searches. Similarly, Netflix, eBay and Amazon are reputed to use their cutting-edge recommendation engines, based on a complex system of statistical and mathematical algorithms, to earn billions of dollars in profits from repeat business, cross-selling and upselling.

There is now a high demand for professional data analysts (with statistical and mathematical backgrounds) in many businesses and Government Departments. Huge budgetary expenditure is made yearly on data analytics services. Businesses and government departments devote huge amounts of investment to get leading insights and intelligence for better decisions. The global market for financial services data analytics alone is worth over $20bn (£12.4bn) annually while the market for health analytics is as much as $300 billion annually in the United States alone. Data analytics, therefore, hold out the answers to what leaders of government and businesses are asking. Nigeria, like other developing countries, can benefit from the rapidly bourgeoning data economy, and help to apply cutting-edge data analytic algorithms to solve problems for businesses and governments to create job opportunities.

This paper discusses and argues that modern application of statistical and mathematical algorithms in data analytics can lead to new job opportunities. Following the introduction, Section two presents emerging requirements for data analytic services and professionals. In Section three, the paper surveys statistical and mathematical algorithms used in modern data analytics. Section four presents the background to demand-driven

[1] Initial draft of this paper was presented at the 3rd Biennial National Conference on Modern Mathematical Statistics, organized by Statistical Research Group (SRG), University of Benin, Benin City, May 11-13th, 2016. The authors are grateful to participants for their useful comments.

and supply-driven job opportunities, while Section five discusses areas of potential application of data analytics in Nigeria, with major lessons from the UK experience. Section six concludes the paper.

1.0 Emerging Requirements[1] for Data Analytics and Scientists

The demand for data analytic services and professionals has assumed exponential proportions in the last decades, especially in the developed economies. A number of factors are responsible for this development. First, is the huge data from call centres and other data repositories. The increased emphasis on data capture applications and softwares prior to the 2000s led to the development of several databases, especially the relational database management system (RBDMS), which enabled businesses, government agencies and institutions to capture all kinds of data, ranging from text, images, audios, videos as well as social network data.

Over those years, databases were designed, developed and populated with huge amount of data from business activities, call centres, logs and other data intensive-activities. The resulting data-overload is due, largely, to a multiplicity of online and offline data capture tools and technologies that are now available at home, on the road and in the offices. Many mobile devices (e.g. iPad, PDA's, Mobile Phones etc.) are now used to capture significant amount of data on the go. Chips in credit cards, microwave, refrigerators and other household and in-car devices are also generating data for households, government institutions and businesses. With internet-enabled sensors increasingly embedded in smart meters, in cars and in our electrical appliances, the amount of data being produced is increasing rapidly. Farber (2010) observes that "*experts predict that in just four years from now, the amount of digital information in the world will double every 11 hours*". Almost all media types and contents about anything and users can now be stored in databases. Call Centres, for instance, can now store conversations with customers in text as well as audio and video records. Annually, data is now worth a $300 billion industry, and employing over 3 million people in the U.S. alone (Jason Morris and Ed Lavandera, 2016)[2].

The emergence of web 2.0 technologies led to the proliferation of social networks, leading to an era of user generated contents (UGC). Social networks and discussion forums coupled with their associated content generated by users have again added to the information and data available to businesses and governments.

Governments have invested massively in databases and data repositories at all levels. For instance, the UK Government has tagged its attempt as the "open data revolution", involving huge investments by government and private organisations in databases and data analytics services. The databases are, however, useless unless the data can be analysed and mined for insights and business intelligence. In the post-2000s period, there has been increased requirement for application of alternative statistical and mathematical algorithms in the mining and analysis of the huge databases that were developed over the years.

These developments led to huge interest among businesses and governments, desiring to get more out of existing databases. As someone said *"the combination of an increasingly complex world, the vast proliferation of data, and the pressing need to stay one step ahead of the competition has sharpened focus on using analytics within organizations."* Leaders, both in governments and businesses, now need more insights in making better decisions. Data analytics current provide that insights and business intelligence.

The recent advances in computer technology, computer programming and data mining have opened up new opportunities for getting more out of existing databases. Age-long statistical and mathematical algorithms have now found new application in data analytics, collective and business intelligence, big data and insights as computing technologies advance.

The big data phenomenon has added another dimension to the exponential increase in the demand for data analytic services and professionals. The big data refers to large datasets in businesses and government data repositories. The increased volume, detail and frequency of information captured by companies, the rise of social networks, multimedia, and the wide-spread use of Internet continue to contribute to exponential growth in data. According to Manyika et al (2011), big data would be the next frontier for innovation, competition and productivity. Big data is, however, difficult to analyze using traditional database management system and existing analytical and software techniques. A new set of skills and technologies are emerging for the storage and analysis of big data. Consequently, there is a significant shortfall in skills necessary for organizations to take advantage of big data. It has been estimated that by 2018, the United States alone faces a shortfall of over 140,000 people with relevant analytical skills, and additional dearth of 1.5 million managers and analysts with the requisite skillset for analyzing big data to make sound decisions.

These developments have expanded the scope, degree and intensity of application of data analytics in deriving business intelligence, collective intelligence and insights. Business intelligence (BI) derives from transformation of large amounts of structured or unstructured data into meaningful and useful action-points for decision-making purposes. There are several BI tools (e.g. SAS, SPSS, IBM Modeler, etc.), for analyzing large amounts of raw data

[1] **"**Data has swept into every industry and business function and is now an important factor of production, alongside labor and capital." (Manyika, et al, 2011).

[2] "*Acxiom is one of the largest data-brokering firms in the world. It is just one of hundreds of companies who are peering into your personal life, collecting data that is generated from everything you do online, and much of what you do in the real world. The company recorded $1.1 billion in sales last year offering "analytical services" on 144 million households" (Jason Morris and Ed Lavandera, 2016)*

to provide necessary insights for generating new business opportunities. The primary goal of BI is easy interpretation of large volumes of data. At other times, what is required is to identify new opportunities in the existing database of customers. BI tools can also generate actionable insights for businesses to take market advantage and achieve long-term stability. Insights are generated from an existing database after subjecting it to rigorous data mining and other statistical analyses. Still, data analytics can also be used in developing collective intelligence for better decision-making. Collective intelligence derives from collaboration, collective efforts and competition of users/customers in consensus decision making. Typical online collective intelligence involves online voting or rating of a product or service, with a view to creating the general impression of users.

2.0 A Survey of Statistical and Mathematical Algorithms for Data Mining

There are numerous statistical and mathematical algorithms that are currently applied in data analytics by businesses, government agencies and other institutions. Although most of the algorithms are not new, their application is new in many sectors of the economy. For instance, Netflix, a major online cd/video renting company in the United States in 2009, announced a prize of $1million for anyone or group of persons or institutions, who would help improve their in-house recommendation software, *Cinematch*. Their objective was to improve their existing online recommendation engine. Many individuals and groups enlisted in the entry, competing for the prize. On the 21st September 2009, the team "BellKor's Pragmatic Chaos" emerged the winner, and was awarded the $1M Grand Prize**.** On July 26, 2009 at 18:18:28 UTC, the prize was awarded to that team for their verified submission, which achieved the winning RMSE of 0.8567 on the test subset. This score represents a 10.06% improvement over and above the Cinematch's score on the test subset prior to the contest.

The winning team's major key to achieving a 10.06% improvement on the quiz set was the complex blending scheme that combined many results from the individual teams that joined to form the winning team. The work involved a complex array of algorithms including *matrix factorisation, logistic regression and neural networks*, to name a few of the analytical techniques used by the winning team. The winning recommendation engine is entirely the product of intensive application of statistical and mathematical algorithms in data analytics, again underpinning the relevance of data analytics to modern business success.

Why was Netflix willing to give away $1.0 million to a project involving a rigorous mathematical and statistical exercise? The answer is simple: The company uses the recommendation engine to make better decisions to improve its profitability. Thus Netflix was willing to pay a whopping prize for the development of a piece of software that is based on a complex system of statistical and mathematical algorithms in order to enhance the effectiveness of their recommendation engine.

Broadly, data analytics techniques can be sub-divided into two, namely supervised and unsupervised learning techniques.

2.1 Supervised Learning

In supervised learning, the output datasets provided are used to train the machine and get the desired outputs. Supervised learning is the machine learning task that can generate a function from training data. It is supervised because a set of training dataset is provided for the purpose of enabling the algorithm to learn about the sample datasets before the actual production datasets are fed to the algorithm. The training data consists of a set of training examples. As the learning has been done on the training datasets, the algorithm is now able to recognize patterns and structure similar to the training datasets in a production environment.

2.2 Unsupervised Learning

Unsupervised learning is a machine learning task which derives a function that describes hidden structure from unlabeled data. In unsupervised learning, no datasets are provided instead the existing data is clustered into different classes or clusters. Since the examples given to the learner are unlabeled, it is referred to as unsupervised. Clustering techniques such as KNN, K-Means, TwoStep, etc. are a popular group of unsupervised learning platforms.

Most other statistical and mathematical algorithms for DA are supervised learning techniques such as: Econometrics (multiple, ridge regression, polynomial, etc), Neural Networks, Matrix factorization, Bayesian Filter Techniques, Mathematical Optimisation techniques e.g. Mathematical programming, Data Envelopment Analysis (DEA); IBM ILOG, Measures of similarities and dissimilarities (Cosine, Pearson, etc.), and Translog cost function.

3.0 Demand-Driven and Supply-Driven Job Creation

In what follows, we make a difference between demand-driven job opportunities and supply-driven job opportunities in the labour market. On the one hand, we define demand-driven job opportunities as the existing vacancies which are perceived to be available by the firm, government or non-governmental agencies. Therefore, the job market would need to develop skills and manpower to fill those vacancies. This group of job opportunities is demand-driven as it relies on the firm's (or labour market's) demand for skills to fill existing roles, and may no longer be sustainable under the current challenging economic conditions. In this case, it is the firm's demand for skills that leads to job opportunities.

On the other hand, supply-driven job opportunities occur when available skills and manpower in the labour market help to add value to business and corporate processes and activities, leading to job openings. This group of jobs is created by the supply of specialized skills, which were hitherto non-existent in the organization, hence it is supply-driven job opportunities. It is this latter approach to job creation that we emphasise, discuss and recommend in this paper. This method leads to job creation in those businesses and corporations, which hitherto had no vacancies. The underlying principle is that businesses and government agencies want to improve in the way they do business.

Traditionally, the job market and its wages respond to supply and demand functions. The higher the supply of skills/labor in the market the lower the wages/salaries received by labour and vice versa. The supply of jobs, however, is determined to a large extent, by the available job opportunities in the private and public sectors of the economy. The available job opportunities in a firm are determined by what the firm perceives as activities or jobs to be carried out at the time. The demand-driven jobs are, therefore, a simple summation of the job opportunities in both the private and public sectors. This is what is popularly quoted and reported as the total job market for the economy. It is the demand-driven job opportunities that are used to compute various job indicators such as unemployment rates, average wages and their trends.

It is also possible that the firm may have what can be referred to as "hidden job opportunities" that can only be unlocked by use of specialized skills such as sophisticated and complex system of algorithms and other machine learning techniques. In such cases, the firm does not need to know what can be done or what is possible at the time. The skill-holder would have to offer his services either as an individual or as a consultant (e.g. private partnership) to the firm to help unlock its opportunities for business success. The job opportunities created due to this type of intervention from a skilled person/entity is what is referred to as supply-driven job creation. It is a supply-driven job market as the job opportunities were unlocked using specialized skills supplied by the labour market.

With regard to information technology, there have been dynamic changes in the areas of machine learning and data analytics disciplines since the year 2000. Prior to the 2000s, there was increased emphasis on data capture applications and softwares, which enabled businesses, government departments and institutions to capture all kinds of data, ranging from text, images, audios and videos. Over those years, huge databases were designed, developed and populated with terabytes of data. Databases with all kinds of contents are, however, useless unless the data can be mined and analysed for insights and business intelligence. Thus, there is the need for development and application of alternative statistical and mathematical algorithms in the mining and analysis of the huge databases that have been developed over the years. In particular, the post-2000s period has witnessed massive increase in the demand for skills in data analytics.

As many sectors of business and governments are still to be fully unlocked in terms of their application of modern statistical and mathematical techniques, professionals with requisite skills should be ready to help businesses and government agencies to mine their data and apply robust techniques for data analysis for the purposes of generating insights and intelligence. In doing so, the professionals would be helping to create job opportunities in those businesses and government agencies, which hitherto did not have any job vacancies. This is what supply-driven job creation entails.

The UK Example

In the UK, there is a copious application of statistical and mathematical algorithms in data analytics for business insights and intelligence. In a number of UK data analytics agencies (consultancies), their business model has changed from bidding for existing job opportunities or projects. Rather they submit proposals to their clients to add value and be rewarded for the added value. It has become increasingly popular as the agencies do not receive upfront fees. The new generation data analytics agencies mine and analyse their clients' (businesses, governments and institutions) databases, with a view to deriving market intelligence and insights. The results of the analysis and data mining are passed on to their clients to enhance their activities and daily operations such as marketing, customers' recruitment and products pricing. The agencies are only paid for additional profits generated by the new insights. The agency fees are paid from the added value to the business.

The business model is as follows:

a) The agency takes the client's database;
b) The agency mines the database and applies a number of analyses using statistical and mathematical algorithms to derive insights;
c) The insights and intelligence are used by the client in its business activities such as marketing, customer recruitment, pricing, etc.
d) The agency and its client monitor the returns to investment in the above activities against past returns. Additional profits are calculated for each activity; and
e) The agency is paid a proportion of the additional profits generated due to the application of business intelligence.

It is this job market model that we promote in this paper. Essentially, we are advocating that the Nigerian job market needs to diversify, to allow skilled (especially specialized data analytic) professionals to add value to activities and operations of businesses and governments. In doing so, they help to create job opportunities which would not naturally have been available.

4.0 Potential Application of Data Analytics in Nigeria

4.1 Marketing Analytics

Marketing Analytics (MA) is the application of quantitative and analytical techniques to marketing data in order to help business to make insight-based decisions and achieve higher returns to their marketing campaigns and investment (Omoregie, 2013). It involves the analysis of marketing data in order to help business reap better returns from marketing investments. Although many areas of data analytics have emerged in recent decades, nowhere has data analytic been so applied as much as marketing. This is increasingly so because the cost of marketing for most businesses is on the increase. There are now very many channels of marketing a product or service promotion as never seen before. In the past, businesses promoted their products and services through above the line (ATL) channels such as radio, TV, Bill Boards and press. But today, computing technology is at the heart of advertising, offering non-traditional channels such as direct mailing, direct email campaigns, digital displays and online channels such as organic and inorganic online campaigns. As a result there has been an upsurge in search engine optimisation (SEO) and search engine marketing such as pay per click (PPC) and social network. Customers are increasingly active in social networks such as *Twitter*, *Facebook*, *Youtube* and businesses want to advertise to them where they are. All of this has led to huge marketing costs. Given limited marketing budget, businesses now want to streamline their marketing efforts and campaigns in order to reap the best returns to their investments. MA is, therefore, increasingly being used by very many companies in today's competitive business world, to achieve better competitive advantage and secure sizeable market share.

The main issues which MA aims to address include:

- Determining the best channel of advertising to achieve the best returns (ROI) on campaign investment;
- What level of marketing investment would result in the best returns to the business;
- The levels of marketing investment would result in better competitiveness of the business;
- If the current level of marketing investment is optimal for the attainment of business goals (short or long term);
- What levels of marketing investment or campaign would achieve a required level of business KPI such as number of products sold, number of customers in the database, or number of applications made online, etc;
- Measures of brand equity – The company also wants to ascertain its brand equity in relation to competitors. For example, what price the consumers are willing to pay over and above other brands; and
- Brand loyalty as measured by NPS (Net Promotion Score) and its consideration among its customers and non-customers.

Case Study 1: The UK Experience in Marketing Analytics

The lead author led several marketing analytics projects for several UK organisations, financial and non-financial. Typical projects involved KPI's such as recruitment of commercial customers, number of loans application, volume of savings, and credit card application, to name a few. The company (a financial institution) invested huge amount of money in alternative marketing channels, including above-the-line and below-the-line channels. The financial institution's objective was to determine the most cost-effective channel of advertisement as well as the channel which has the highest returns on investment (RoI). It was also interested in knowing the returns on a pound investment in advertisement. To meet these objectives, the lead author built a ***marketing mix model***, a multiple regression technique, which uses the above named KPI's as the objective function, regressed on a number of variables. The model is referred to as marketing mix as it includes a mixed-grill of variables, including the company's internal variables, competitors' variables and macroeconomic variables. Specifically, the variables included competitors data (their spends across a number of advertising channels, account and credit conditions) as well as company's internal variables such as advertisement spends across all channels, lending and saving rates, as well as macroeconomic variables.

Marketing mix modeling technique involves the following steps:

- Definition of KPI (e.g. number of account opened, credit cards applications, number of loans applications, etc.)
- Identification of independent variables
- Specification of appropriate equations
- Deciding of decay of ads
- Data collection and collation

The model was setup and executed in SAS codes.

$$Z = \alpha + Ias_{\tau} + Ios_{\tau} + Com_{\tau} + Ma_{\tau}$$

Where variables are defined below:

Z = KPI (e.g. number of account opened, credit card applications, loan account, etc.)

Ias_{τ} = Vector of firm-specific Ad spends (e.g. TV, Press, Online, Radio, Digital and other Outdoor advertising)

Ios_{τ} = Vector of other firm-specific independent variables (e.g. Bank NPS, Product Pricing such as Current Account, Savings interest rates, survey data, PR, branch numbers, etc).

Com_{τ} = Competitors independent variables (e.g. Competitors Ad Spends across all channels, product pricing, NPS, branch numbers, PR, etc.)

Ma_{τ} = economic factors e.g. consumer confidence, interest rates, inflation, FTSE, Seasonality, etc.

Outcomes of Marketing Mix Modelling:

- Different marketing channels that significantly contributed to the KPI, for each campaign;
- The ROI on each investment for each marketing channel and each campaign; and
- The least cost marketing channel;

4.2 Customer Analytics

Customer analytics can be defined as a process by which data from customer behavior is analyzed to help make key business decisions, and may involve segmentation and predictive analytics. Customer analytics (CA) is the application of quantitative analytical techniques to customer data in order to derive insights and intelligence to help make better business decisions. This field of analytic involves varied analysis of customer data (including loyalty, purchases and preferences history) so as to derive business insights and intelligence to help in minimizing churn or attrition rate while enhancing life time value, satisfaction and retention rate of customers. Therefore, businesses are not just interested in the physical numbers of customers in their databases. They want to know the demographics and segmentation of those customers. They now want to know the rate of change of those numbers, along with other specific characteristics of those customers. They want to know what the customers are saying and asking, and how they can meet those needs.

Business is about customers. Customers have always been the major determinants of the success of businesses. If a business loses its customers or cannot offer them what they want, a certain future for that business is extinction. It's right to say therefore, that businesses exist because of customers, and they are sustained because of customers. In today's business world, the power of customers in determining the success or failure of a business is even greater than ever before. In fact, the major competitive challenge of today's businesses is how to attract new customers and retain existing ones. That also implies competing with other businesses for the available customer base. Competition is even now further worsened by the increased power of the customers. Ours is now an era of customers. This is the era of increased awareness, interaction and knowledge of the customers themselves. So they are using available media to interact with themselves and with businesses, in order to influence what products and services are offered and bought in the market. Customers are talking and sharing information about products and services. In the process they are working to influence product prices and quality.

As a result, customer databases now contain huge information about customers, including buying patterns, product choice and preferences, demographics and even what the customers are saying and doing in social media. Customer databases have now become a huge repository of information and resources that help businesses make better business decisions. Someone said that customer database is the oil-field of the future. It's truly speaking the goldfield of the future for many businesses. That's why business leaders are keen on managing their customer databases effectively. In the last two decades, several management systems and IT infrastructure have been developed and offered as CRM (customer relationship management) tools. In the 21st century, a major part of managing customer database is data analytics that are specifically aimed at building and enhancing the relationship between businesses and their customers in order to achieve customer satisfaction, resulting in repeat business, upselling, cross-selling and other customer added value. When fully accomplished, it can help to predict a customer's buying pattern, reveal information about a customer's taste and preferences and make offerings to meet them. It can help determine the rate of retention or churn of customers while predicting the life time value of customers.

The main aspects of customer analytics are store cards, loyalty cards, store location, customer segmentation, pricing, demand analysis, forecasting, product assortment, store location plans). Others are: cost of acquisition, product assortment, promotion, merchandising, store layout, store site selection, targeted marketing, churn and retention rates, lifetime value, product mix analysis (basket analysis), Cross selling and upselling, Customer database clustering, Micro segmentation, value segmentation and multi-market segmentation analyses.

Case Study II: The UK Experience in Customer Analytics

As part of the lead author's experience in the UK, he worked in projects and teams, involving the use of unsupervised learning technique in the clustering of customers' databases. Major client databases clustered were fast-moving consumer goods (FMCG), departmental stores, clothing departments, plumbing company and a UK leading political party. The major objectives of the above named clients include clustering of their databases into distinct clusters, identification of profit clusters and products by clusters. For each cluster, the number of metrics computed included basket analysis, churn and retention rates, attrition rates, target marketing, etc.

The steps involved in mining customers' database include the following:

- Identification of database characteristics (field names, types and size);
- Adopting appropriate clustering techniques (e.g. KNN, K-Means, etc.);
- Identification of the number of clusters appropriate for the customer database;
- Cluster the database using data mining tools such as IBM SPSS Modeler;
- Examine and describe the characteristics of each cluster, each in turn and compare the clusters horizontally and vertically; and
- Identify and compute metrics such as recency and frequency (RF), attrition rates, as listed above for each cluster.

4.3 Social Network Analytics

Social networks have become a veritable platform for millions of web users for chatting, discussions, friendship and interactions. The advent of Web 2.0 technologies has further enhanced the effectiveness of social networks. Major networks are facebook, youtube, linkedIn, mySpace, Instagram, etc., to name a few. The high level of user generated content (UGC) in social network space has resulted in significant amount of data being generated daily.

There is, therefore, the need for use of sophisticated techniques and algorithms to identify patterns and analyse social network data in order to enhance effectiveness of the network.

Case Study III: The UK Experience in Social Network Analytics
The lead author led a team to design and build a recommendation engine for a social network environment, used by an educational institution in the UK to recruit students and create an effective learning environment for students and academics. The network has many business objects, including discussions forum, groups, path, profile, events, reviews, user subject matter and other pedagogical objects. The lead author designed and developed, along with web developers and programmers, the above objects using a number of statistical and mathematical algorithms, namely: cosine similarity, data mining and clustering algorithms such as KNN and K-Means, logistic regression, text analysis, including stemming, tokenization, and use of ontological databases.

- The K-Means was used to cluster the database of students, web visitors and lecturers into clusters.
- Cosine similarity algorithm was used to identify the similarity in the use of alternative keywords entered by social network users and visitors, such that contents were recommended to users of the network with similar coefficients.
- It was also used to classify users in their use of subject keywords and assign users to groups and discussions forum.
- Cosine similarity algorithm was used to examine similarities between Paths based on their usage of keywords, steps and other activities in the Paths, with a view to recommend association of contents, groups and events to Path clusters.

4.4 Web Analytics

Web analytic involves the measurement, collection and analysis as well as reporting of web data for purposes of understanding and improving web users' activities. Web analytics is an emerging field of analytics. There's certainly a huge prospect for this field, due to the increasing level of web data, a function of the increasing number of web data capture tools in several web applications such as ecommerce site, discussion fora, social networks and search engines.

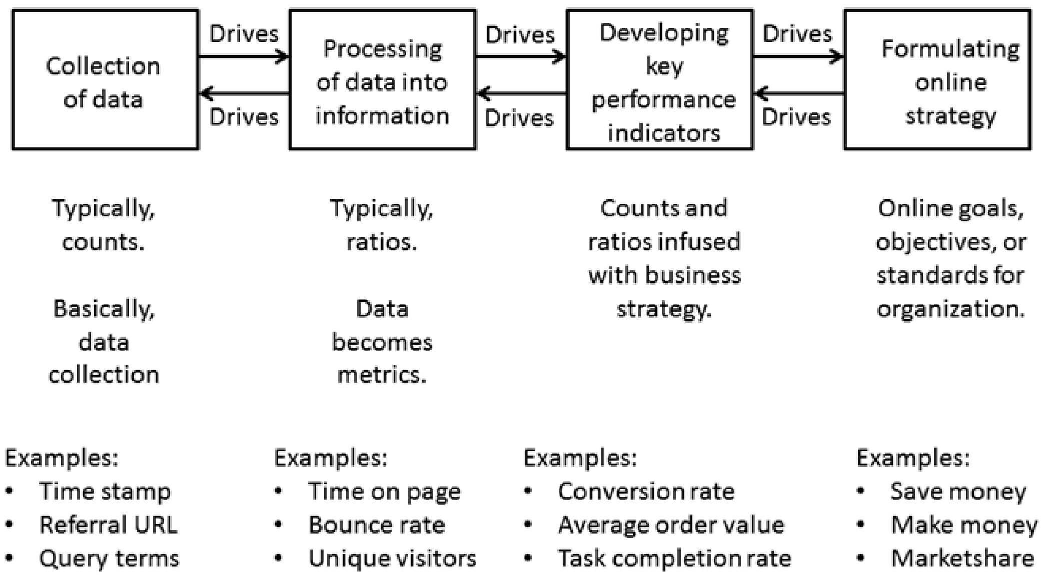

Figure 1: Typical Web Analytic Steps

Web analytics involves collecting and analyzing web users data with a view to draw inferences, insights for effective campaign and marketing purposes. The process may involve search engine optimization (SEO) and search engine marketing (SEM). A wide selection of metrics are usually computed during web analytics, including click path, bounce rate, average page view duration, repeat visitor, return visitor, new visitor, first visit, first session, hit and event. Other metrics are customer segments, referrals, traffic, conversion metrics, funnels & path analysis. Many vendors and application service providers have developed Web (Online) data analytics tools and they include:

- Tealeaf , Coremetrics or Webtrends
- Google analytics
- Adobe Sitecatalyst
- Omniture, ClickTracks, WebTrends,
- Coremetrics, HBX
- Social network listening applications e.g. Radian6, SySomos

4.5 Financial Analytics

In today's highly competitive business environment, companies need forward-looking, predictive insights that can help shape tomorrow's business strategy and improve day-to-day decision-making in real time. As a result, the application of data analytics in the financial sector has gained popular acceptance in many countries. Financial analytics involve the analysis of financial data to meet specific business questions and forecast potential financial scenarios. A major aspect of financial analytics is credit risk analytics. The current crises in the bank industry in particular and in the financial services sector in general, have underpinned the need for robust, comprehensive and science-based risk management strategies. Strong credit risk analytics should diagnose loop-holes and break-points in their existing risk strategies and then proffer robust and all-weather risk modelling techniques that ensure higher loan repayment and minimize loan-default parameters. In the UK, EU and US, credit risks measures and barometers are now built from existing client data using cutting-edge modelling and algorithmic techniques. Based on basels 1, 11 & 111, retail and corporate credit lending portfolios, with particular emphasis on calculation and modelling of economic capital, PD, LGD, EAD and their drivers, stress testing, impairment and capital forecasting.

Major areas of financial analytics are:

- Credit risk analysis
- Risk Management Analytics
- Actuarial analytics
- Portfolio management
- Insurance analytics
- Scorecards analytics
- Propensity modeling

Case Study IV: The UK Experience in Financial Data Analytics

The lead author has also applied statistical and mathematical algorithms in financial analytics for leading UK financial institutions. Some of the questions the managers of financial institutions are asking are:

- How to score applicants for available financial products such as loans, credit cards, overdrafts, etc.?
- What is the propensity of each loan applicant defaulting?
- What is the probability (PD) of a customer defaulting on his/her financial obligations?
- What are the economic capital, LGD and EAD levels?

The lead author has helped to design and build a credit scorecard for a UK leading financial organization. The scorecard was used by the financial institution to score loan and credit card applicants, to ascertain their suitability for the products. Applicants that score below a threshold is rejected and others are granted the products applied for. The Scorecards were built using historic data, and used to predict future behavior of customers.

The steps adopted in building the scorecard were:

- Collate, clean and prepare data for use (fine and coarse classing).
- Segment the data into two unequal datasets; the larger dataset used to build the model, and the smaller dataset to validate the model once built.
- Execute rigorous logistic regression procedures.
- Estimate the scorecard's scoring accuracy.
- Validate the scores against the sample that was set aside.
- Update of the scorecard as customer situation and behaviour changes from time to time.

4.6 The Big Data Analytics

Big data refers to huge structured and unstructured datasets that are so large it is difficult to process using traditional database and software tools. Many businesses and government departments now have big data, which continues to present major challenges to leaders as traditional databases SQL and other data manipulation techniques are certainly limited in their ability to handle the big data. Big data analytics involve manipulating and analyzing large, varied, structured and unstructured data sets (big data) to reveal hidden patterns and relationships, customer preferences, trends and other key information to help organizations take well-informed business decisions. Unstructured data often includes text and multimedia content such as word processing documents, presentations, webpages, e-mail messages, blog entries, photos, videos, audio files and other types of documents. Structured data, on the other hand, can be stored in fields in a record or file as is usually the case with spreadsheets and relational databases. While structured data can be managed by using Structured Query Language (SQL), unstructured data cannot, hence the need for a new set of tools for managing and querying big data. In general, big data is distinctly characterized by high volume, high velocity (the speed of data inflow and outflow makes it difficult to analyze); and a wide variety (data types are too many). Consequently, big data analytics require specialized software applications and tools. The increasing need for professionals (data analysts, data scientists and miners) with robust skillsets in the big data tools has led to expansion in job opportunities in the sector.

Table 1: List of available big data jobs in a week in selected countries

No.	Country	No. of Vacancies
1.	United States of America	10829
2.	United Kingdom	3956
3.	Germany	240
4.	Switzerland	118
5.	Ireland	69
6.	Australia	61
7.	Netherlands	60
8.	Belgium	43

Source: JobServe.Com (searched on 29th April 2016).

The United States of America churns out over 10,000 job opportunities in Big Data analytics every week, followed by the UK at about 4000 job openings, indicating the rapid expansion in the bid data sector. Consequently, a new set of tools and skillsets are now evolving for storing and analyzing big data for businesses and governments. As we write the paper, some of the specialized skills and technologies for big data scientists are as follows:

- Hadoop ecosystem e.g. IBM, Hortonworks, Cloudera, etc.
  - Hadoop (set up, administration, tuning and security
  - Kafka, Flume, Spark
  - Data ingestion
  - SQL on Hadoop tools
- Pivotal Big Data Suite e.g. Greenplum and Hawq
- ETL tools e.g. Informatica BDE, Pentaho, Talend
- NoSQL and other databases e.g. Cassandra, MongoDB, Neo4J

These skills are significantly different from the skills and tools used by data scientists working in a non-big data environment, implying that the data economy holds out great opportunities to Nigeria, both in terms of job opportunities and enhancing the competitiveness of local businesses so they can take advantage of better decisions.

5.0 Conclusion and Policy Implications

In the last two decades, we have seen increased prominence of data analytics and data science as a job sector in many economies, especially among the advanced economies. This is primarily due to a number of developments, including success stories of major companies such as Google, eBay, Amazon and Netflix, in their application of statistical and mathematical algorithms in data analytics, in response to the bourgeoning global data economy. These developments have enhanced the degree and intensity of application of data analytics to deriving business intelligence, collective intelligence and insights required by businesses and governments.

As demand-driven job opportunities are shrinking in many jurisdictions due to the challenging global economic conditions, we advocate a supply-driven job creation, where skills are developed in technology and data analytics (with strong reliance on statistics and mathematics), with a view to solving business and corporate problems.

If Nigeria is to benefit from the rapidly expanding global data economy, the country needs to focus on:

- Development of new skills and technologies that support the rapidly expanding data economy, and meet the needs of business and governments for better intelligence and insights;
- Realignment of curricula and training programmes of educational institutions to focus on teaching modern applied statistics, mathematics and other quantitative skills that would prepare our university graduates for data analytics services and career;
- The statistical and mathematical bodies/agencies should form a formidable pressure group to lobby the

ISSN 2224-5758 (Paper) ISSN 2224-896X (Online)
Vol.7, No.12, 2017

government to use their highly invaluable skills and expertise to solve national and sub-national socio-economic challenges.

References

Morris, Jason and Ed Lavandera (2012) Why big companies buy, sell your data CNN, Updated 2052 GMT (0352 HKT) August 23, 2012.

Netflix 2009 $1 million Prize Competition, 2009, http://www.netflixprize.com/

Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers (2011): Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute website.

Omoregie David (2013) Application of Data Analytics in Businesses, An unpublished draft book.

Sabherwal, Rajiv and Irma Becerra-Fernandez (2010) Business Intelligence: Practices, Technologies and Management. John Wiley and Sons, Inc.

Farber, Dan (2010): Data doubling every 11 hours Martin Lamonica of news.com reports on IBM's release of FileNet P8 4, February 13, 2007 -- 16:12 GMT (08:12 PST)

Authors

David E Omoregie, is Principal Economist, Monetary Policy Department, Central Bank of Nigeria (CBN). He holds a PhD in Applied Economics, University of Aberdeen, United Kingdom, MSc and BSc degrees (Agricultural Economics), University of Ibadan, Nigeria. He also holds a Post-Graduate Diploma in Quantitative Techniques from the University of Aberdeen. Before joining the CBN, he worked for 20 years in various capacities as economist, manager and data scientist for UK financial institutions, including Lloyds TSB Bank, Egg Plc, and GE Life (Pensions and Life).

Moses Kpughur Tule is the Director, Monetary Policy, Central Bank of Nigeria. He holds a B.Sc degree in Economics, University of Sokoto (1986), M.Sc Finance, University of Nigeria (2004), M.Sc Economics, University of Nigeria (2006), Post Graduate Diploma, Banking and Finance, Enugu State University of Science and Technology (1997), Post Graduate Diploma, Accountancy, University of Nigeria (1998), LL.B, University of Nigeria (2004), LL.M, University of Liverpool (2016). He is currently a Doctoral student at the University of Nigeria.