

# Application of Big Data Analytics to Curb Maternal and Child Mortality in Nigeria

Michael, O Daniel, Agbakwuru A.O, Ekwonwune Emmanuel,  
Amanze Chibuke bethran

*Department of ICT, School of GS Alvan ikoku fed collage of education Owerri ,Imo state, Nigeria is currently pursuing Doctorate degree program in computer science in imo state University, Owerri Nigeria.*

*Currently the present Head Of Department Of Computer Science Faculty Of Science Imo State University, Nigeria.*

*Department Of Computer Science Imo State University, Faculty Science Of Owerri, Nigeria.*

*Department Of Computer Science Imo State University, Nigeria*

Submitted: 01-12-2021

Revised: 12-12-2021

Accepted: 15-12-2021

## ABSTRACT

Big data analytics is modern day trend of data implementation. The paper handled the application of data of big data analytics to curb maternal and child mortality In Nigeria. Based on the following of the old system revealed; 1. lack of predictive mechanism in place to curb the menace. 2. Lack of skilled attendants at delivery due to lack of predictive mechanism. The aim for this study, is to design and implement a big data analytics system model to curb maternal and child mortality in Nigeria. Methodology, Agile methodology was deployed for architectural framework, and testing of the different phases of the program development model. Technique Employed was Data mining. Result; The implementation of the system It is expected that at the end of the research, a predictive decision support system would have been designed, implemented and delivered for use by Nigerian medical personnel and their clients. The system will run analysis in the pool of data harnessed through Big Data Analytics. After running the analytics the system then sends its updates result findings, in which Doctors and health organizations will rely on for decision making. In fact, the system will dictate upward or downward trends which will prompt the needed decision. Good programming package such as Python, Java Script, and Knox programming language then Mongo DB and Mysql for the database development.

### Keywords:

Agile, Big data analytics, Curbing, Data crawling, Data Mining, Mongo DB, Mortality, Maternal, Predictive System, web crawling

## INTRODUCTION:

Big data comprises of millions and uncount-

able data analysis being generated daily and on this data mining technique is the AI that will be monitoring the big data platform for different specific analysis that are centered on maternal and child mortality, and then signal any change in data to the system at any point in time. The people who are expected to use this system comprises, the health care workers, Medical doctors, health ministers, Health commissioners, Health admin and so many other with health affiliations. Big data is where the world is moving to now and its taking over in almost all the fabric of IT development. Big Data is the avalanche of data pool which has been pouring out different data sources through our day to day data processing and information generation.

According to wikipedia, "Big data" is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software. Data with many cases (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source. Big data was originally associated with three key concepts: volume, variety, and velocity. Other concepts later attributed to big data are veracity (i.e., how much noise is in the data) and value. On the need to improve services aimed at mitigating impacts of maternal death on children, BIG data analytics is introduced such as enhancing systemic linkages between health and social welfare programs, and strengthening social protection schemes to better address the needs of vulnerable children and families, our findings

support the need to redouble efforts to combat maternal deaths, which are the root cause of the impacts on families and children documented here.

On that, the discovery of the Big data solves this problem that connects millions of its users across geographic divides irrespective of distance, location or demographic. Just a few years ago, people carried their documents around with disks, memory sticks and USB devices, these have become history as evolution in computing research has given rise to big data analytics, making it easy, flexible and cost effective using data mining technique and its application into maternal and mortality rate.

On this, Data mining is aspect of Big data analytics which is a machine learning algorithm that deals with prediction, just as we have phishing and botnecks techniques in machine learning algorithms. So data mining is the algorithm meant to serve as the technique being introduced into the big data system in order to produce a model that will predict and guide against the of child maternal and mortality rate. In fact, data Mining has formed the bedrock of research interest and will play a major role in achieving the best result as seen by the trend of research works in the past. The consequences of the poor state of pregnant women in Nigeria are numerous and affect ma-

ternal as well as child mortality. The under-five mortality ratio in Nigeria is 201 per 1000 live births meaning that one in five Nigerian children never reach the age of 5. Infant deaths, which account for half of child mortality, have increased from what they were in 1990. With a 13% immunization rate for children between 12-23 months, Nigeria is the African country with the lowest vaccination rate. Substantial presence of Acute Respiratory Infections and diarrhoea also contribute to the elevated mortality rates for children [1].

### I.The Concept Of Big Data Analytics

Big data analytics is the science of working on complex large amount of data both structured and unstructured to uncover information. The term big data was first used to refer to increasing data volumes in the mid-1990s. In 2001, Doug Laney, then an analyst at consultancy Meta Group Inc., expanded the notion of big data to also include increases in the variety of data being generated by organizations and the velocity at which that data was being created and updated. Those three factors -- volume, velocity and variety -- became known as the 3Vs of big data, a concept Gartner popularized after acquiring Meta Group and hiring Laney in 2005.



Fig 1:The four classic dimensions of Big Data (4 Vs)

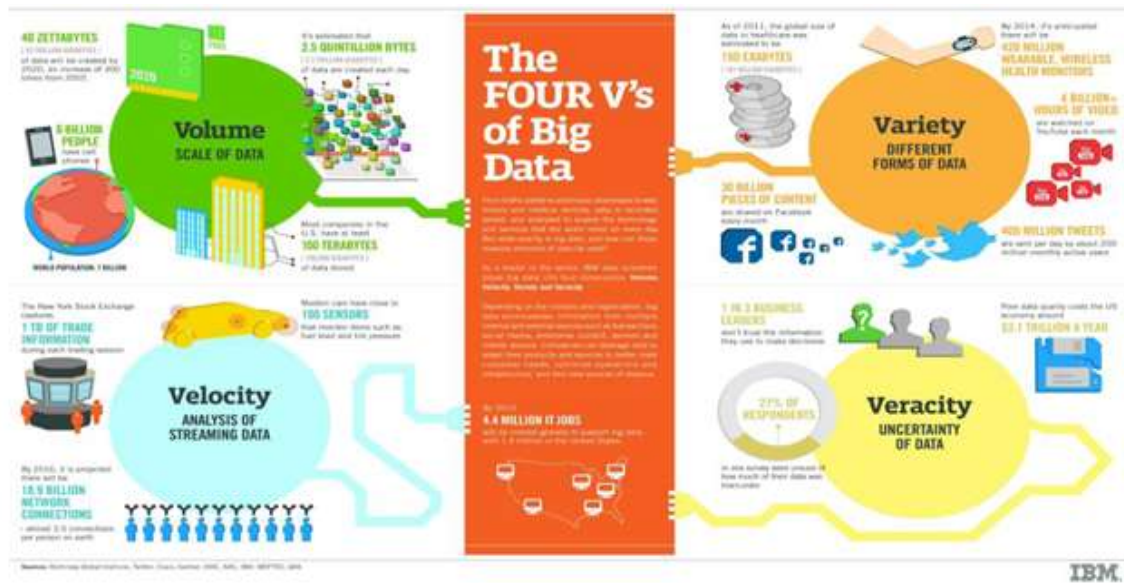


Fig 2: The 4 Vs of Big Data - IBM Infographic

Source :IBM skill academy( introduction to big Data Ecosystem )

**II. The-4-Vs-Of-Big-Data-For-Yielding -Gems Of-Information**

**a. Volume**

As the name suggests, the main characteristic of big data is its huge volume collected through various sources. We are used to measuring data in Gigabytes or Terabytes. However, according to various studies, big data volume created so far is In Zettabytes which is equivalent to a trillion gigabytes. 1 zettabyte is equivalent to approximately 3 million galaxies of stars. This will give you an idea of colossal volume of data being available for business research and analysis. Take any sector and you can comprehend that it is flooded with loads of data. Travel, education, entertainment, health, banking, shopping - each and every sector can benefit immensely from the Big data advantage. Data is collected from diverse sources which include business transactions, social media, sensors, surfing history etc. With every passing day, data is growing exponentially. Thousands of TBs worth data is created every minute worldwide via Facebook, tweets, instant messages, email, internet usage, mobile usage, product reviews etc. Every minute, hundreds of twitter accounts are created, thousands of applications are downloaded, and thousands of new posts and ads are posted. According to experts, the amount of big data in the world is likely to get doubled every two years. This will definitely provide immense data in coming years and also calls for smarter data management. As the volume of the data is growing at the speed of light, traditional database technology will not suffice the need of efficient data management i.e. storage and analysis.

The need of the hour will be a large scale adoption of new age tools like Hadoop and MongoDB. These use distributed systems to facilitate storage and analysis of this enormous big data across various databases. This information explosion has opened new doors of opportunities in the modern age.[6]

**b. Variety**

Big data is collected and created in various formats and sources. It includes structured data as well as unstructured data like text, multimedia, social media, business reports etc. Structured data such as bank records, demographic data, inventory databases, business data, product data feeds have a defined structure and can be stored and analyzed using traditional data management and analysis methods. Unstructured data includes captured like images, tweets or Facebook status updates, instant messenger conversations, blogs, videos uploads, voice recordings, sensor data. These types of data do not have any defined pattern. Unstructured data is most of the time reflection of human thoughts, emotions and feelings which sometimes would be difficult to be expressed using exact words. As the saying goes “A picture paints a thousand words”, one image or video which is shared on social networking sites and applauded by millions of users can help in deriving some crucial inferences. Hence, it is the need of the hour to understand this non-verbal language to unlock some secrets of market trends. One of the main objectives of big data is to collect all this unstructured data and analyze it using the appropriate technology. Data crawling,

also known as web crawling, is a popular technology used for systematically browsing the web pages. There are algorithms designed to reach the maximum depth of a page and extract useful data worth analyzing. Variety of data definitely helps to get insights from different set of samples, users and demographics. It helps to bring different perspective to same information. It also allows analyzing and understanding the impact of different form and sources of data collection from a 'larger picture' point of view. For instance, in order to understand the performance of a brand, traditional surveys are one of the forms of data collection. This is done by selecting a sample, mostly from panels. The advantage of this approach is that you get direct answers to the questions. However, we can obtain real time feedback through various other forms like Facebook activity, product review blogs, and updates posted by customers on merchant websites like Flipkart, Amazon, and Snap deal. A combination of these two forms of data definitely gives a data-backed, clearer perspective to your business decision making process.[6]

**c. Velocity**

In today's fast paced world, speed is one of the key drivers for success in your business as time is equivalent to money. Fast turn-around is one of the pre-requisites to stay alive in this fierce competition. Expectations of quick results and quick deliverables are pressing to a great extent. In such scenarios, it becomes vital to collect and analyze vast amount of disparate data swiftly, in order to make well-informed decisions in real time. Low velocity of even high quality of data may hinder the decision making of a business. The general definition of Velocity is 'speed in a specific direction'. In big data, Velocity is the speed or frequency at which data is collected in various forms and from different sources for processing. The frequency of specific data collected via various sources defines the velocity of that data. In other terms, it is data in motion to be captured and explored. It ranges from batch updates, to periodic to real-time flow of the data. The frequency of Facebook status updates shared, and messages tweeted every second, videos uploaded and/or downloaded every minute, or the online/offline bank transactions recorded every hour, determine the velocity of the data. You can relate velocity with the amount of trade information captured during each trading session in a stock exchange. Imagine a video or an image going viral at the blink of an eye to reach millions of users across the world. Big data technology allows you to process the real time data, sometimes without even capturing in a database. Streams of data are processed and databases are updated in real-time, using parallel processing of live streams of data. Data streaming helps extract valuable

insights from incessant and rapid flow of data records. A streaming application like Amazon Web Services (AWS) Kinesis is an example of an application that handles the velocity of data. The higher the frequency of data collection into your big data platform in a stipulated time period, the more likely it will enable you to make accurate decision at the right time.[6]

**d. Veracity**

The fascinating trio of volume, variety, and velocity of data brings along a mixed bag of information. It is quite possible that such huge data may have some uncertainty associated with it. You will need to filter out clean and relevant data from big data, to provide insights that power up your business. In order to make accurate decisions, the data you have used as an input should be appropriately compiled, conformed, validated, and made uniform. There are various reasons of data contamination like data entry errors or typos (mostly in structured data), wrong references or links, junk data, pseudo data etc. The enormous volume, wide variety, and high velocity in conjunction with high-end technology, holds no significance if the data collected or reported is incorrect. Hence, data trustworthiness (in other words, quality of data) holds the highest importance in the big data world. In automated data collection, analysis, report generation, and decision making process, it is inevitable to have a foolproof system in place to avoid any lapses. Even the most minor of slippage at any stage in the big data extraction process can cause immense blunder. Any reports generated based on a certain type of data from a certain source must be validated for accuracy and reliability. It is always advisable that you have 2 different methods and sources to validate credibility and consistency of the data, to avoid any bias. It is not only about accuracy post data collection, but also about determining right source and form of the data, required amount or size of the data, and the right method of analysis, play a vital role in procuring impeccable results. Integrity in any field of business life or personal life holds highest significance and hence, proper measures must be put in place to take care of this crucial aspect. It will definitely allow you to position yourself in the market as a reliable authority and help to you to attain greater heights of success.[6]

**e. Parting thoughts**

These 4v's are like 4 pillars lending stability to the giant structure of big data and adding a precious 5th "V" - **value** - to the information procured leads to the whole purpose of big data, smart decision making. Separately, the Hadoop distributed processing framework was launched as an Apache open source project in 2006, planting the seeds for a clustered platform built on top of commodity hardware and geared to run big data applications. By 2011, big data analyt-

ics began to take a firm hold in organizations and the public eye, along with Hadoop and various related big data technologies that had sprung up around it. Initially, as the Hadoop ecosystem took shape and started to mature, big data applications were primarily the province of large internet and e-commerce companies such as Yahoo, Google and Facebook, as well as analytics and marketing services providers. In the ensuing years, though, big data analytics has increasingly been embraced by retailers, financial services firms, insurers, healthcare organizations, manufacturers, energy companies and other enterprises.[2].

[2] looked at big data analytics on a broad scale. They opined that data analytics technologies and techniques provide a means to analyze data sets and draw conclusions about them which help organizations make informed business decisions. Business intelligence (BI) queries answer basic questions about business opera-

tions and performance. Big data analytics is a form of advanced analytics, which involves complex applications with elements such as predictive models, statistical algorithms and what-if analysis powered by high-performance analytics systems.

### III. The Importance Of Big Data Analytics

Driven by specialized analytics systems and software, as well as high-powered computing systems, big data analytics offers various business benefits, including:

- New revenue opportunities
- More effective marketing
- Better customer service
- Improved operational efficiency
- Competitive advantages over rival.

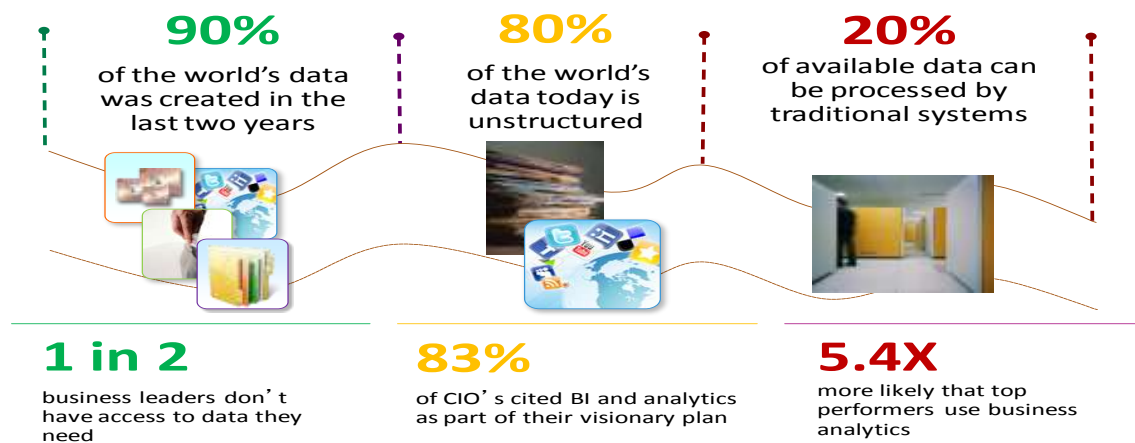


Fig 3: Below diagram shows Data has an intrinsic property.

Source :IBM skill academy( introduction to big Data Ecosystem )

Big data analytics applications enable big data analysts, data scientists, predictive modelers, statisticians and other analytics professionals to analyze growing volumes of structured transaction data, plus other forms of data that are often left untapped by conventional BI and analytics programs. This encompasses a mix of semi-structured and unstructured da-

ta -- for example, internet click stream data, web server logs, social media content, text from customer emails and survey responses, mobile phone records, and machine data captured by sensors connected to the internet of things (IoT). [6]

#### IV. A look at Types of Big Data

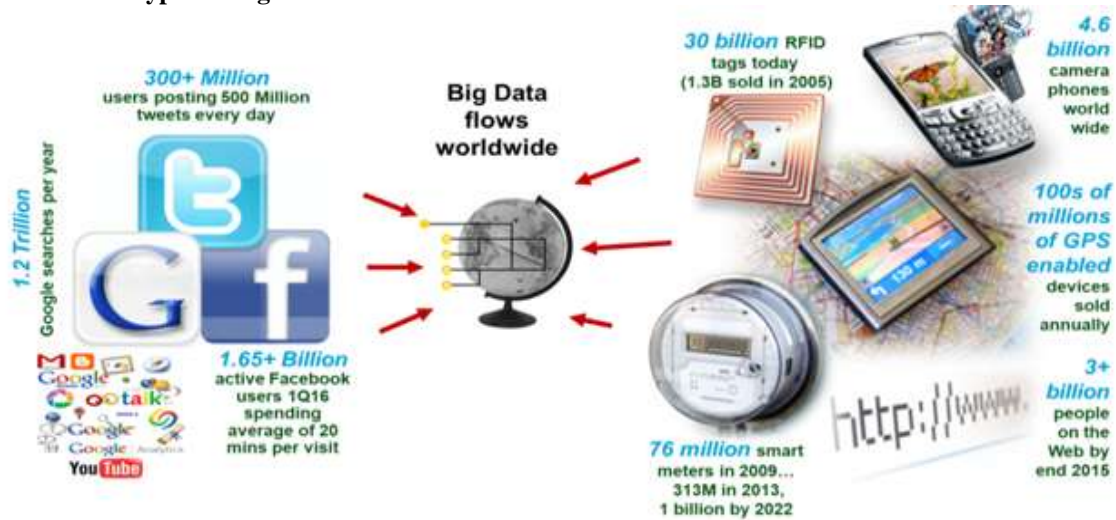
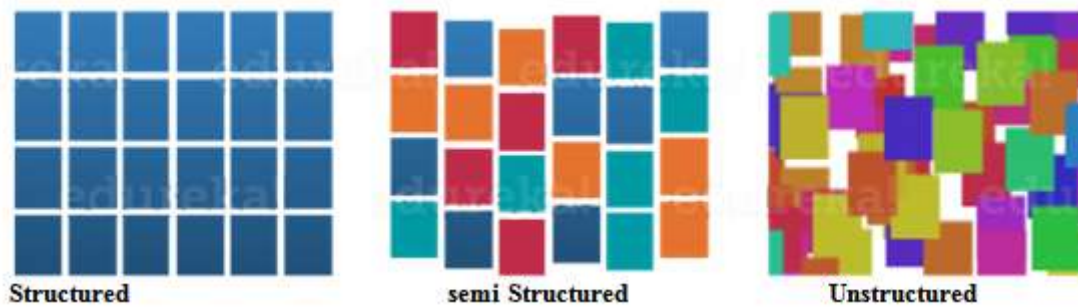


Fig 4: Growing interconnected & instrumented world of Big Data  
 Source :IBM skill academy( introduction to big Data Ecosystem )



- a. **Structured:** The data that can be stored and processed in a fixed format is called as Structured Data. Data stored in a relational database management system (RDBMS) is one example of 'structured' data. It is easy to process structured data as it has a fixed schema. Structured Query Language (SQL) is often used to manage such kind of Data.[6]
- b. **Semi-structured:** Semi-structured Data is a type of data which does not have a formal structure of a data model, i.e. a table definition in a relational DBMS, but nevertheless it has some organizational properties like tags and other markers to separate semantic elements that makes it easier to analyze. XML files or JSON documents are examples of semi-structured data.[6]
- c. **Unstructured:** The data which have unknown form and cannot be stored in RDBMS and cannot

be analyzed unless it is transformed into a structured format is called as unstructured data. Text Files and multimedia contents like images, audios, and videos are examples of unstructured data. The unstructured data is growing quicker than others, experts say that 80 percent of the data in an organization are unstructured.[6]

#### V. Big Data Analytics Technologies And Tools

Unstructured and semi-structured data types typically don't fit well in traditional data warehouses that are based on relational databases oriented to structured data sets. Further, data warehouses may not be able to handle the processing demands posed by sets of big data that need to be updated frequently or even continually, as in the case of real-time data on stock trading, the online activities of website visitors or the performance of mobile applications. As a result,

many of the organizations that collect, process and analyze big data turn to NoSQL databases, as well as Hadoop and its companion data analytics tools, including:

- **YARN:** a cluster management technology and one of the key features in second-generation Hadoop.
- **MapReduce:** a software framework that allows developers to write programs that process massive amounts of unstructured data in parallel across a distributed cluster of processors or stand-alone computers.
- **Spark:** an open source, parallel processing framework that enables users to run large-scale data analytics applications across clustered systems.
- **HBase:** a column-oriented key/value data store built to run on top of the Hadoop Distributed File System (HDFS).
- **Hive:** an open source data warehouse system for querying and analyzing large data sets stored in Hadoop files.
- **Kafka:** a distributed publish/subscribe messaging system designed to replace traditional message brokers.
- **Pig:** an open source technology that offers a high-level mechanism for the parallel programming of MapReduce jobs executed on Hadoop clusters.

## VI. How Big Data Analytics Works

In some cases, Hadoop clusters and NoSQL systems are used primarily as landing pads and staging areas for data before it gets loaded into a data warehouse or analytical database for analysis -- usually in a summarized form that is more conducive to relational structures. More frequently, however, big data analytics users are adopting the concept of a Hadoop data lake that serves as the primary repository for incoming streams of raw data. In such architectures, data can be analyzed directly in a Hadoop cluster or run through a processing engine like Spark. As in data warehousing, sound data management is a crucial first step in the big data analytics process. Data being stored in the HDFS must be organized, configured and partitioned properly to get good performance out of both extract, transform and load (ETL) integration jobs and analytical queries.

Once the data is ready, it can be analyzed with the software commonly used for advanced analytics processes. That includes tools for:

- **data mining**, which sift through data sets in search of patterns and relationships;
- **predictive analytics**, which build models to forecast customer behavior and other future developments;

- **machine learning**, which taps algorithms to analyze large data sets; and
- **deep learning**, a more advanced offshoot of machine learning.

**Text mining** and statistical analysis software can also play a role in the big data analytics process, as can mainstream business intelligence software and data visualization tools. For both ETL and analytics applications, queries can be written in MapReduce, with programming languages such as R, Python, Scala, and SQL, the standard languages for relational databases that are supported via SQL-on-Hadoop technologies.

## VII. Big Data Analytics Uses And Challenges

Big data analytics applications often include data from both internal systems and external sources, such as weather data or demographic data on consumers compiled by third-party information services providers. In addition, streaming analytics applications are becoming common in big data environments as users look to perform real-time analytics on data fed into Hadoop systems through stream processing engines, such as Spark, Flink and Storm.

Early big data systems were mostly deployed on premises, particularly in large organizations that collected, organized and analyzed massive amounts of data. But cloud platform vendors, such as Amazon Web Services (AWS) and Microsoft, have made it easier to set up and manage Hadoop clusters in the cloud, as have Hadoop suppliers such as Cloudera-Hortonworks, which supports the distribution of the big data framework on the AWS and Microsoft Azure clouds. Users can now spin up clusters in the cloud, run them for as long as they need and then take them offline with usage-based pricing that doesn't require ongoing software licenses.

Big data has become increasingly beneficial in supply chain analytics. Big supply chain analytics utilizes big data and quantitative methods to enhance decision making processes across the supply chain. Specifically, big supply chain analytics expands datasets for increased analysis that goes beyond the traditional internal data found on enterprise resource planning (ERP) and supply chain management (SCM) systems. Also, big supply chain analytics implements highly effective statistical methods on new and existing data sources. The insights gathered facilitate better informed and more effective decisions that benefit and improve the supply chain.

Potential pitfalls of big data analytics initiatives include a lack of internal analytics skills and the high cost of hiring experienced data scientists and data engineers to fill the gaps.

### VIII. Some Examples Of Big Data Sources

- Science • Astronomy • Atmospheric science • Ge-  
nomics • Biogeochemical • Biological • Other complex  
/ interdisciplinary scientific research • Social • Social  
networks • Social data
- Person to person (P2P, C2C):
- Wish Lists on Amazon.com • Craig’s List • Person to  
world (P2W, C2W): • Twitter • Facebook • LinkedIn
- Medical records • Commercial • Web / event / data-

base logs

- "Digital exhaust" - result of human interaction with  
the Internet • Sensor networks
- RFID • Internet text and documents • Internet search  
indexing • Call detail records (CDR) • Photographic  
archives • Video / audio archives • Large scale  
eCommerce
- Regular government business and commerce needs •  
Military and homeland security.

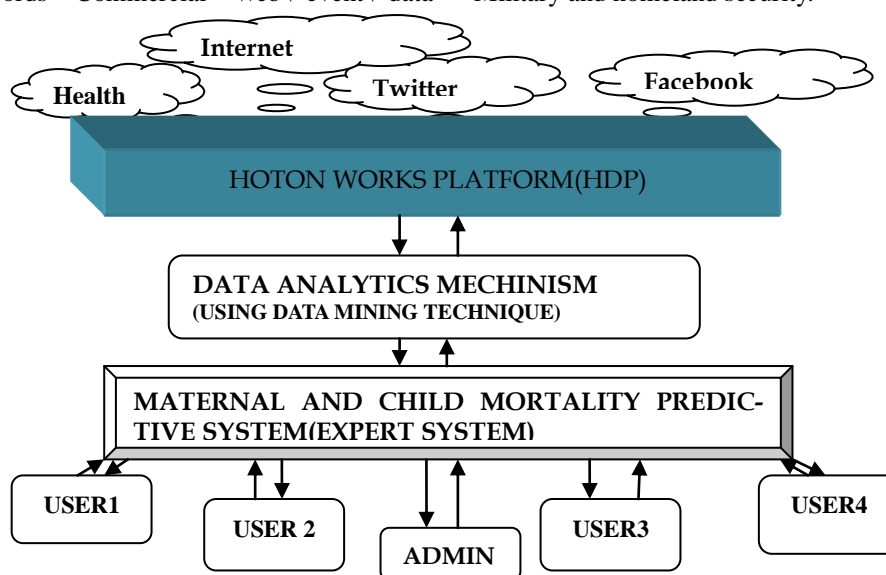


Fig:5 Maternal And Child Mortality Decision Support Making Predictive System

### IX. The big data analytics application system to curb, maternal and child mortality.

Maternal and child mortality in Nigeria is an epidemic that is really ravaging the health of our pregnant mother as well as our newborn . Big data is one element that has consistently helped to achieve gaint strides, through its ability to deliver to practitioners a volume and variety of structured or unstructured data in solving organizational problems . Big data has enabled more widespread and specific research and trials of stratifying and segmenting populations at risk for a variety of health problems. Examples of success using big data are surveyed in surveillance and signal detection, predicting future risk, targeted interventions, and understanding disease. Using novel big data or big data approaches has risks that remain to be resolved. The continued growth in volume and variety of available data, decreased costs of data capture, and emerging computational methods mean big data success will likely be a required pillar in solving maternal and child mortality in Nigeria. The use of Big Data in curbing maternal and child mortality problem in Nigeria, aims to identify the maternal and child mortality use cases where big data has added value, identify

classes of value that big data may bring, and outline the inherent dimensions in using big data in developing the system .

[3]The term big data refers to the emerging use of rapidly collected, complex data in such unprecedented quantities that terabytes ( $10^{12}$  bytes), petabytes ( $10^{15}$  bytes) or even zettabytes ( $10^{21}$  bytes) of storage may be required. The unique properties of big data are defined by four dimensions: volume, velocity, variety and veracity. As more information is accruing at an accelerating pace, both volume and velocity are increasing. Use of a variety dimension marks a shift from data as information that is collected directly to information that is assimilated from multiple sources. Big data outputs tend to increase in value as sources become more diverse. At a population level, traditional health data included information from vital statistics registries and hospital admission statistics. In the last few decades, however, more health data have been assimilated from electronic medical records, mobile phone and purchase records, geographical positioning systems, social media and beyond. The veracity dimension refers to the uncertainty around data and their collection, standardization and validation. As the



quantification and articulation of the uncertainty in reported data have been a part of health-care research and practice for many years, health practitioners may be more familiar with data veracity than many other users of big data. Uncertainty and confidence intervals are now commonly reported in projects that use these data. Together, these dimensions enable a big data approach to health, in which health priorities and poli-

cies are driven by analytics of large data sets.

### X. Big Data Analytic Techniques

An Insight into Big Data Analytic Techniques, I discovered that big data has 26 analytical techniques but the technique which am centered on is data mining technique.

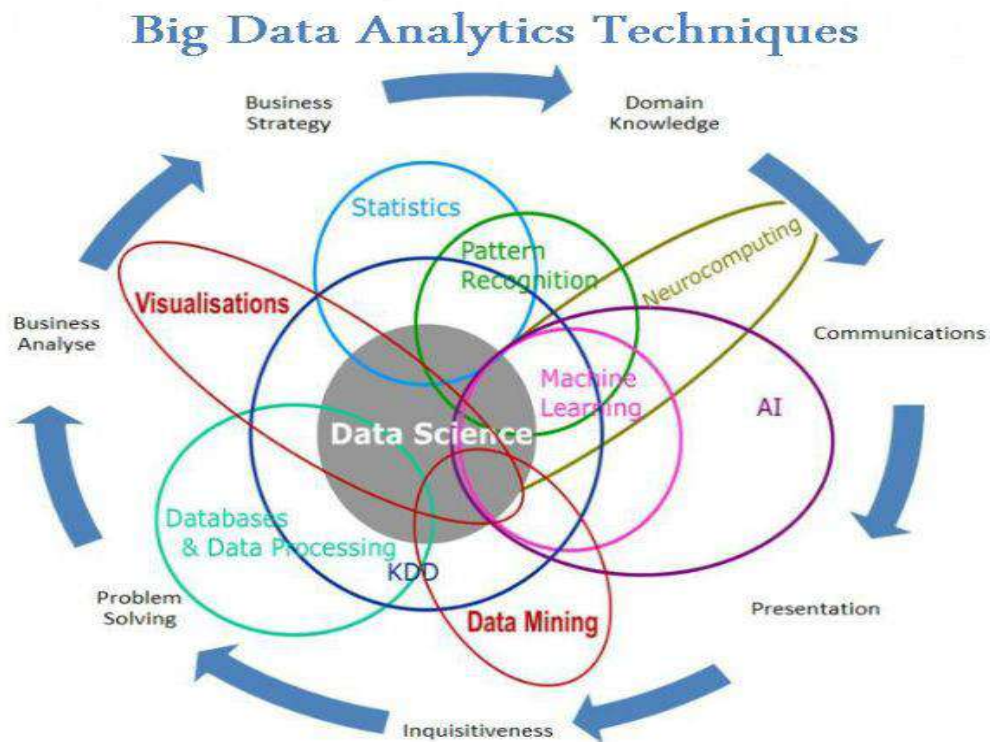


Fig 6; Source :IBM skill academy( introduction to big Data Ecosystem )

Data mining techniques are used in many research areas, including mathematics, cybernetics, genetics and marketing. While data mining techniques are a means to drive efficiencies and predict customer behavior, if used correctly, a business can set itself apart from its competition through the use of predictive analysis. Web mining, a type of data mining used in customer relationship management, integrates information gathered by traditional data mining methods and techniques over the web. Web mining aims to understand customer behavior and to evaluate how effective a particular website is. Other data mining techniques include network approaches based on multitask learning for classifying patterns, ensuring parallel and scalable execution of data mining algorithms,

the mining of large databases, the handling of relational and complex data types, and machine learning. Machine learning is a type of data mining tool that designs specific algorithms from which to learn and predict.

#### 1. Data mining

Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Data mining tools allow enterprises to predict future trends.[2].[4] sees Data Mining as a logical process of finding useful information to find out useful data. Once you discover the information and patterns, Data Mining is used for making decisions for developing

the business. Data mining, also called knowledge discovery in databases, in computer science, the process of discovering interesting and useful patterns and relationships in large volumes of data. The field combines tools from statistics and artificial intelligence (such as neural networks and machine learning) with database management to analyze large digital collections, known as data sets. Data mining is widely used in business (insurance, banking, retail), science research (astronomy, medicine), and government security (detection of criminals and terrorists [5]).

### I. Data mining parameters

In data mining, association rules are created by analyzing data for frequent if/then patterns, then using the support and confidence criteria to locate the most important relationships within the data. Support is how frequently the items appear in the database, while confidence is the number of times if/then statements are accurate. Other data mining parameters include Sequence or Path Analysis, Classification, Clustering and Forecasting. Sequence or Path Analysis parameters look for patterns where one event leads to another later event. A Sequence is an ordered list of sets of items, and it is a common type of data structure found in many databases. A Classification parameter looks for new patterns, and might result in a change in the way the data is organized. Classification algorithms predict variables based on other factors within the database. Clustering parameters find and visually document groups of facts that were previously unknown. Clustering groups a set of objects and aggregates them based on how similar they are to each other. There are different ways a user can implement the cluster, which differentiate between each clustering model. Fostering parameters within data mining can discover patterns in data that can lead to reasonable predictions about the future, also known as predictive analysis.[2]

### II. Benefits of data mining

In general, the benefits of data mining come from the ability to uncover hidden patterns and relationships in data that can be used to make predictions that impact businesses.

Specific data mining benefits vary depending on the goal and the industry. Sales and marketing departments can mine customer data to improve lead conversion rates or to create one-to-one marketing campaigns. Data mining information on historical sales patterns and customer behaviors can be used to build prediction models for future sales, new products and services.

Companies in the financial industry use data mining tools to build risk models and detect fraud. The manu-

facturing industry uses data mining tools to improve product safety, identify quality issues, manage the supply chain and improve operations.[2].

### III. Data-mining techniques

There are many types of data mining, typically divided by the kind of information (attributes) known and the type of knowledge sought from the data-mining model:

#### - Predictive modeling

Predictive modeling is used when the goal is to estimate the value of a particular target attribute and there exist sample training data for which values of that attribute are known. An example is classification, which takes a set of data already divided into predefined groups and searches for patterns in the data that differentiate those groups. These discovered patterns then can be used to classify other data where the right group designation for the target attribute is unknown (though other attributes may be known). For instance, a manufacturer could develop a predictive model that distinguishes parts that fail under extreme heat, extreme cold, or other conditions based on their manufacturing environment, and this model may then be used to determine appropriate applications for each part. Another technique employed in predictive modeling is regression analysis, which can be used when the target attribute is a numeric value and the goal is to predict that value for new data.

#### - Descriptive modeling

Descriptive modeling, or clustering, also divides data into groups. With clustering, however, the proper groups are not known in advance; the patterns discovered by analyzing the data are used to determine the groups. For example, an advertiser could analyze a general population in order to classify potential customers into different clusters and then develop separate advertising campaigns targeted to each group. Fraud detection also makes use of clustering to identify groups of individuals with similar purchasing patterns..

### 3. Methodology

The research adopted agile research methodology approach. This will provide periodical testing prototypes to continue building the system until all user requirements are completely satisfied. Good programming package such as Python, Java Script, and Knox programming language then Mongo DB for the database development.

The implementation platform for the proposed system is to be in all the health service institutions. The proposed system will be designed as a mobile application, deployed to health centers especially in rural communities where access to analytical data is at its lowest ebb.

Python programming language is the tool selected for

the implementation of the system. Tensor flow open source library of python has been chosen for the implementation of the suite.

**4. Design And System Implementation**

Application of Big data analytics on maternal and child mortality is design with Knox, Python, Java and HTML programming language using Mysql server architecture. The system can be hosted online and once this is done, all FORMS, MODULES and FUNCTIONS that make up the system are activated ready for use. The application is window based, menu driven and highly interactive. The system has been built around a large body of Big data platform known as Horton works platform. It becomes therefore necessary that a sensitive platform of this kind should be adequately protected to grant confidence in its applications.

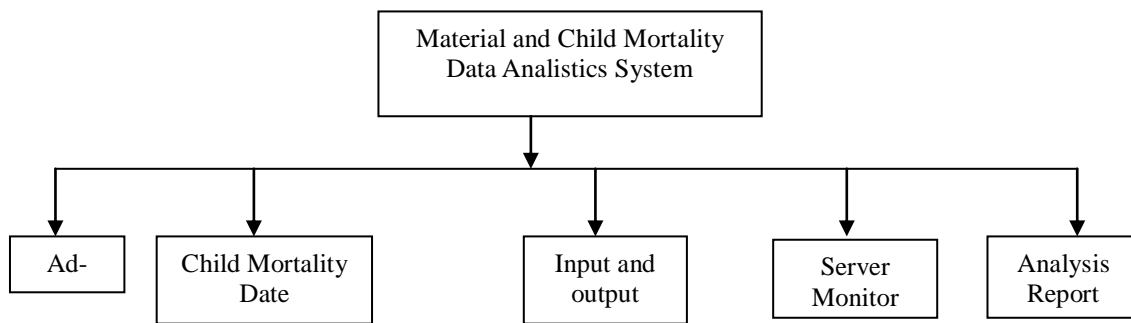
**I. Objective of the design :** Design is the first step into the development phase for any engineered product or system. Design is a creative process. A good design is the key to effective system. The term design is defined as “the process of applying various techniques and principles for the purpose of defining a process or system in sufficient details to permit its physical realization. More so, The design considerations for the design of the new system are as follows: This is to say that the application of big data analytics to curb maternal and child mortality is structured to effectively achieve these following objectives.

- a. To develop a compatible system that will effectively run a prediction and analysis that will curb the maternal and child mortality.
- b. To design a user-friendly and interactive system,
- c. To design a high performance application or link that will help capture the periodic maternal and child mortality .
- d. To build a robust mortality detection system that will enable doctors and nurses to intermittently get steady update.
- e. To develop application that will help the health care examiners capture both mortality activities as well as supervise the trend.
- f. To design an application that will generate fast upload and update of maternal and child mortality information and as well as search and retrieve data’s for decision making.
- g. To develop a high secured big data warehouse that will keep all the analytical records.

**5. Control centre and Main Menu**

The Control Center/Main Menu of the system is the main menus to which the system workings of the deployment of the expert system to curb maternal and child mortality flows.

Thus, the Control Center identifies and drives the following “actors”, the Admin (Administrator), the Input/Output directory ,Child Mortalit Data, Server monitor and Analysis Report, which is where the deployment of expert system using big data analytics to curb maternal and child mortality is implemented through hadoop file system technique.



**Fig 7:** The Control Center

**I. The Submenus / Subsystems**

This work makes the design specification quite precise without unnecessarily restricting the selection of the best available means for developing the

final product. This establishes the fact that the relationship between information structure design and hardware/software is similar to that of architectural design and building materials.

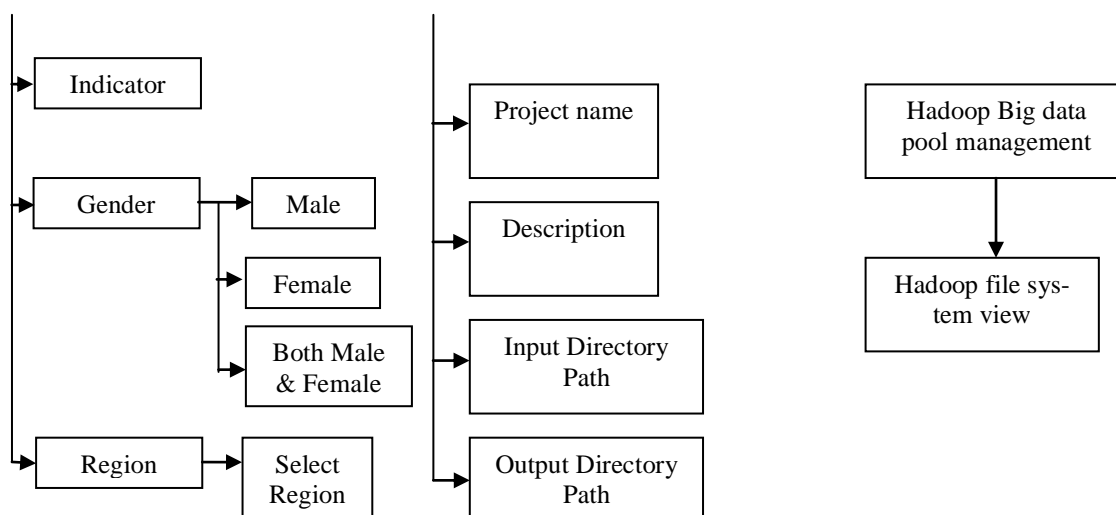


Fig 8 : Subsystems

- a. login as admin and the portal opens. In this way, it take you to the main portal which is the welcome page.
- b. The next will take you to create project. This is were you create a project name and description . This technique ensures a secured area of which you will want to select for analysis. For instance you may wish to select IMO State as a place to analyse, so when you select a project it enbles you select state or are of interest to analyse for every healthcare service as it concerns maternal and child mortality.
- c. The next one is menu is the child mortality data. That is where you select the indicator, Gender and Region.
  - i. The indicator ,indicates area the area of analysis.
  - ii. The Gender requires you select the gender you want to analyse.
  - iii. The Region shows the coverage area you want to analys e.g you may decide to select and run the analysis of Eastern region.
- d. The next is the input and output area . In this input and output area is where you will select the directory at which your input directory will operate from and also the directory at which the output will root from.
- e. The next is the hadoop big data pool. This is where after creation of the input and output directories then you select expected document from the document directory and then move it into the hadoop big data pool to be worked on , where it will run or perfrom the big data analytics through the hadoop file system and

hadoop file system server output to produce the expected maternal and child mortality result.

## 6. Findings and Results

The existing system did not proffer solution to curb the menace of maternal and child mortality. It is expected that at the end of the research. The result of this system will be a predictive model and support system designed to be delivered for use by Nigerian medical personnel and their clients.

## 7. Benefit of the System

The benefits of this research when completed includes a provision of a big data connection platform model that will help provide unlimited data to forecasts and warns against impending maternal and child mortality cases with a section of it itemizing ways of curbing and reducing this menace in our society .

## 8. Knowledge contribution

The knowledge contribution in this work, is the introduction of Big data analytics to bridge the challenge in expert system knowledge acquisition which is to help bring about accurate predictive system to curb maternal and child mortality.

## CONCLUSION

The function of big data in the present time can not be over emphasized, Therefore, this research work has been able to describe in details, the design and implementation of a big data analytics model that will curb the menace of maternal and child mortality issues in Nigeria. This analytics mechanism was made possible through the deployment of a Big data analytics, using data mining technique. As part of its non functional requirements, it will also serve as a com-

prehensive mobile application for e-health management system institutions. This model, when implemented, is expected to eliminate the uncertainties being experienced through maternal and child mortality in Nigeria.

### REFERENCES

- [1] Usman Gulumbe, Olatunji Alabi, Olusola A. Omisakin and Semeeh Omoleke, (2018). "Maternal mortality ratio in selected rural communities in Kebbi State, Northwest Nigeria". BMC Pregnancy and Childbirth. Available @ <https://bmcpregnancychildbirth.biomedcentral.com/articles/10.1186/s12884-018-2125-2> World Health Organization. "TRENDS". Available at <https://www.who.int/pmnch/activities/countries/nigeria/en/index1.html>.
- [2] Mark L, Lisa M and Craig S (2019) Guide to big data analytics tools, trends and best practices .
- [3] Rosemary W , Samuel V, William P , Priya M , Temitope F & Leo Anthony Celi (WHO) Big data in global health: improving health in low- and middle-income countries
- [4] Bonani Bose 2019 What is Data Mining: Definition, Purpose, and Techniques.
- [5] Christopher Clifton (2019) Data mining COMPUTER SCIENCE <https://www.britannica.com/data-mining>. IBM skill academy, (2019). introduction to big Data Ecosystem.