BIG DATA ANALYTICS OF BOKO HARAM INSURGENCY ATTACKS MENACE IN NIGERIA USING DynamicK-reference CLUSTERING ALGORITHM

C. P. Oleji[1], E. C. Nwokorie[1], G. A. Chukwudebe[1]

[1]Department of Computer Science, Federal University of Technology Owerri, Imo, State, Nigeria.

---***---

Abstract - *Nigerian security challenges have attracted attention in recent times due to several criminal activities such as kidnappings, armed robbery, hired assassins and lately Boko Haram attacks. This work aimed to analyze Boko Haram insurgency attacks menace in Nigeria using big data analytics model. DynamicK-reference Clustering Algorithm an improved Big data analytic model developed in another work was adapted for the analysis. It extracted Big Dataset of Boko Haram attacks from twitter massages to examine the insurgency activities, area of attacks, reports, period of attacks, Death rolls, and attack strategies. The analysis was carried out from the time Boko Haram Insurgency attacks started in Nigeria from the year 2008 to June 2019. The results show constant attacks of Boko Haram insurgency in Northern Nigeria, which had led to millions of people currently displaced and killed. The repeated terror attacks in Nigeria at different time intervals on the same day shows that the military defense frontline is weak. The security lapses were because the security agencies are not adequately equipped to counter attacks at the frontline. We, therefore, recommend intelligent automated systems like crime-mapping platforms, gunshot-detection systems, video analytics, smart lights, and robotic detection systems to be implemented at the border frontline for efficient monitoring and communication of attacks. Future applications should focus on Boko Haram member's profiles within the society to identify their hidden members in society. The results of this work will enable Nigeria military intelligence to devise efficient strategies to strengthen security control of insurgency attacks.*

***Key Words*: DynamicK-reference Algorithm, Insurgency, Big Data, Swarm Intelligence, Clustering, Analytics**

1. INTRODUCTION

Nigerian security challenges have attracted attention in recent time due to a number of criminal activities such as kidnappings, armed robbery, hired assassins and lately Boko-Haram Insurgency. “Also recent attacks in Pakistan (150 school children died), Canada (the government was targeted in Ottawa and in Quebec), as well as Australia (lindt Caf�siege in Sydney) illustrated that government, business and individuals are impacted by global terrorism and thus it would be useful to know what and where the risks are [1]”. Terrorist attack is a classical big data analytics task due to the fact that factors of terrorist group behaviors are complex to understand [2], the data are high velocity [3], and the volume is large since the evidence is scattered around the world. Terrorist groups tend to have behavioral patterns in their frequency, target country/region, ideology and attack type.

Sufficient knowledge of information about Boko Haram insurgency activity, external threats, and better understanding of the patterns of their operations in the attacked areas will help in devising efficient strategies to strengthen security control of insurgency attacks. This work is aimed at developing a mechanism for scraping attacks of Boko Haram insurgency historical data from social media sites at real time and none-real time processes. And it will apply artificial intelligence algorithms on the extracted dataset to unveil numerous mechanisms to mitigate ongoing insurgency and terrorism in Nigeria. Ultimately, this gives hints of potential threat to the integrity, achievement, developmental goals of the organization and present security problems in Nigeria. The Boko Haram conflict has caused about 2.1 million people to abscond their homes, with about 1.9 million currently displaced and 200,000 of these spread over Cameroon, Chad and Niger Republic. This has created a humanitarian disaster with these internally displaced persons in need of medical and food; last year, UNICEF warned that as many as 50,000 children were facing death from starvation. This humanitarian crisis is far from abetting and the recent rise in attacks by Boko Haram has only served to exacerbate the situation. At this point, over 80% of Borno State is considered high risk or very high risk for international humanitarian organizations. This constrains access to desperately vulnerable communities in need of aid. The recent increases in attacks worsen the security situation and further prevent aid organizations and donors from venturing outside areas that are relatively safe, such as Maiduguri and Biu.

The inefficient commitments of Nigeria defense agencies on the border's frontline and helpless situations of the citizens on the hands of insurgency attacks requires advanced analytical technology to mine the large historical data of terrorist attack strategies, periods of attacks, mechanism of their operation, areas of attacks and condition of the victims to obtain useful insights and measures to mitigate the situation exigently. The application of big data analytics and machine learning on the historical dataset of Boko Haram insurgency attacks in northern Nigeria could provide useful suggestions that will enable Nigeria military intelligences to devise efficient strategies to strengthen security control of insurgency attacks. The vast techniques of the terrorist attacks in the northern Nigeria needs deep analysis on the information accumulated over years of the insurgency's attacks. With such analysis, security heads and agencies can know when there is a deviation from the master plan in place and gets directives on decision control. Moreover, "[4] posed that Big Data tools have a significant contribution to optimized security management it had been integrated into the system as a solution to terrorism problem as it will be more efficient, effective, reliable and productive".

In recent time most especially in the developed world, organizations have realized the importance of mining very large historical dataset known as Big Data to extract useful information for decision making. Big data consists of complex or very large data that the traditional data mining applications cannot efficiently analyze. It has specific characteristics (like: volume, velocity, variety, variability, veracity, validity, volatility and visualization) that can help portray both the challenges and advantages of big data initiatives. The massive amount of data being generated per second through various social media platforms, online marketing platforms, and business websites among others generally defines Big Data [5]. The accuracy of the mining techniques is very important because its result gives directives for investment, control, and insights on security control challenges. It is high time, developing countries like Nigeria embraced information discovery from data mining analysis for decision making.

From research review, strong interest shown by executives in the use of big data analytics to secure information systems appears to be driven by a desire to better understand behavior of their systems. Data is the active gear that power artificial intelligence, and large datasets enhance machine learning applications to learn independently and effectively thereby identify differences, increase pattern recognition capabilities, and unveil the details within the pattern. Big data has the unique features of "massive, high dimensional, heterogeneous, complex, unstructured, incomplete, noisy, and erroneous," which affect the accuracy and time constraints of the traditional data mining methods. The hybridization of traditional data mining methods with efficient swarm intelligence optimization mechanisms to exploit features of big data would therefore provide accurate analytic techniques of solving big data problems [5]. Hence, Swarm Intelligence (SI) systems respond well to the rapidly changing environment of the dynamics of big data manipulations, making use of their inherent auto-configuration and self-organization capability. Various research works had proved the efficient analytic accuracy of the application of swarm intelligence and unsupervised machine learning algorithms can handle the analytical challenges posed by complex characteristics of big data on traditional algorithms. This work adapts DynamicK-reference clustering algorithm, a hybrid of Dynamic Multi-Swarm Optimization (swam intelligence algorithm) and K-reference clustering algorithms (an unsupervised machine learning algorithm) to unveil the hidden patterns of unstructured big dataset of Boko Haram insurgency attacks for military decision support system. "The hybridized clustering algorithms are robust and highly analytics to analyze the extracted social media unstructured datasets of Boko Haram Attacks. It possesses the potential to create clusters of similar internal structure of unstructured (mixed) datasets and produced accurate clustering interpretation for decision support system [5]".

2. LITERATURE REVIEW

Rebollo et al. [6] proposed detection of Jihadism in social networks using big data techniques supported by Graphs and Fuzzy Clustering algorithm. They focused on the analysis of twitter messages to detect how the Jihadism leaders orchestrate terrorist networks and its followers. Their proposed big data architecture was used to analyze messages in real time in order to classify users according to different parameters like level of activity, the ability to influence other users, and the contents of their messages. Graphs were used to analyze how the messages propagate through the network, and this involves the study of the followers based on retweet and general impact on other users. In addition, Fuzzy clustering techniques were used to classify users' profiles and the algorithm was tested with public database from Kaggle and other Twitter extraction techniques. Their results provided efficient decision support system for security agencies, human resources, immigration and other agencies to detect undesirable clients. Abiodum et al. [4] proposed Big Data: an approach for detecting terrorist activities with people's profiling. They reviewed advance knowledge discovery approaches to address terrorism threats which include: Terrorism knowledge portal, Terrorism Expert finder, and Dark web (consisting of internet-based terrorist multilingual resources). They used Linear Regression Analysis (Least squares method) model as data fitting techniques and develop mathematical model to detect person's involvement in terrorism. Their research objective was obtained. They recommended that "people's profiling analysis had a significant contribution to terrorist detecting if integrated into the system as a solution". Kolajo and Daramola [7] presented big data to combat terrorism in developing countries. They proposed a model that harnesses data from multiple social media source in order to detect terrorist activities using Apache Spark Technology. The work described the Social Media Analysis for Combating Terrorism (SMACT) model as a plausible approach that leverages big data analytics to address terrorism problems in developing nations. They considered Nigeria context as a case for illustration for the proposed model to depict its viability as a potential panacea for handling terrorism. Strang and Sun [2] analyzed the relationships in terrorism big data using Hadoop and statistics. They used big data software Hadoop in Google News to collect complex high velocity, high volume terrorism information. They proposed a hypothesis that there is a significant relationship between terrorist group ideology and terrorist attack type. In addition to the test hypothesis, they developed a symmetric model to visualize the hidden relationships between terrorist ideology and attack type. They concluded that the finding of a significant relationship between terrorist ideology and attack type may generalize to supply chain operations and national security planning.

Tour and Gangopadhyay [8] proposed real time big data analytics for predicting terrorist incidents. They developed a real time terrorist incidents data collection system to gather terrorist incidents data from reliable source. The collected data was modeled to calculate the terrorism risk level of different locations and predicted future terrorist incidents. The prediction method provided high precision value of 96.30% and high recall values of 100%. The results of their work will guide terrorism analysts to improve counter-terrorism measures to prevent future attacks. Einstein [9] addressed the cybersecurity skills gap through cooperation, education and emerging technologies. The work sets out examples of existing efforts and initiatives to tackle the gap and demonstrated that new technologies, such as machine learning and AI, can provide efficient mechanism and measures to mitigate cybercrime threats. Maya [10] proposed application of machine learning and deep learning in cybercrime prevention. He stated that Threat of cyber security has drastically transformed over the past few years due to large amount of data produced every second and stored in various forms and in different platforms, which makes it necessary to develop techniques to curbed attacks and theft of critical information. He discussed techniques for applying machine learning and deep learning to control cybercrime and compared the performance of various machine learning techniques such as Logistic Regression (LR), Classification and Regression Trees (CART), Bayesian Additive Regression Tree (BART), Support Vector Machines (SVM), Random Forest (RF), and Neural Network (NN) using precision, recall and F-measures. He concluded that LR is a more preferred option among users due to low false positive mechanism. Also, it had the highest precision and relatively high recall in comparison with other classifiers under contemplation. Karie, Kenande and Venter [11] presented diverging deep learning computing techniques into cybercrime. They stated that in cyber forensics, what makes the process even tough for investigators is the fact that Big Data often comes from multiple sources and has different file formats. Forensic investigators often have less time and budget to handle the increased demands when it comes to the analysis of these large amounts of complex data for forensic purposes. They proposed a generic framework for diverging Deep Learning (DL) cognitive computing techniques into Cyber Forensics (CF) hereafter referred to as the Deep Learning and Cyber Forensics (DLCF) Framework. Their result was promising and it can be used to mitigate cybercrime.

3. RESEARCH METHODOLOGY

This work adapted object-oriented analysis and design methodology for the design of the proposed system. It entails the design of various mechanisms to explain the connectivity and functionalities of the system components. The modules involve: to develop effective web scraping application that will extract the required variables or elements used to describe the harmful behavior and content that correlate with Boko Haram Insurgency attacks in Nigeria. Veracity of the extracted variables (elements) will be achieved using data mining model such as selection, cleaning, integration and storage. It will proceed to use hybrid of Machine learning and Meta heuristic algorithms, subset of artificial intelligence algorithms to analyze the extracted big dataset of Boko Haram insurgency attacks from social media websites.

3.1 Flow Chart Diagram

The system flow chart is shown in Fig. 1.

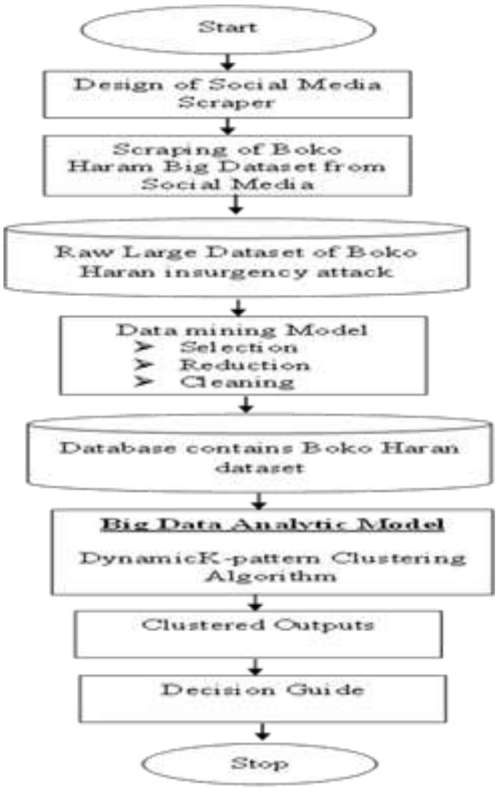

Start
Design of Social Media Scraper
Scraping of Boko Haram Big Dataset from Social Media
Raw Large Dataset of Boko Haran insurgency attack
Data mining Model
➤ Selection
➤ Reduction
➤ Cleaning
Database contains Boko Haran dataset
Big Data Analytic Model
DynamicK-pattern Clustering Algorithm
Clustered Outputs
Decision Guide
Stop

Fig-1: Flow Chart Diagram

The flow chart diagram describes the connectivity and functionalities of the components of the proposed system. The components of the flow chart diagram are as follows:

- Web Scraping software: ScrapeStorm software and Twitter/Facebook Developer's Application Programming Interface (API) Graph were used to develop web scraper (crawler) to extract Boko Haram big datasets from social media sites (Twitter, Facebook, etc.). This provided efficient mechanism for scraping very large data from social media sites and other web sites at real time and none-real time. The developed Scraping Software was used to scrape 3,000,050 tweeters and Facebook user's reports on Boko Haram insurgency attacks in Nigeria.

- Data mining model: Data mining models were used to preprocess Boko haram attacks extracted dataset from social media. This involved using preprocessing mining tools (such as: data selection, cleaning, reduction and etc.). Map Reduction Frameworks were used to reduce the complexity and voluminous nature of the extracted big datasets. The preprocessed dataset was presented in a matrix computing form for the next phase assignment.

- Big Data Analytic Model: Big data analytics focus on data discovery using data sciences, data mining techniques such as supervised learning algorithms (e.g. prediction and classification algorithms), unsupervised learning algorithms(e.gs clustering and association algorithms), optimization algorithms (e.g. swarm intelligences, etc.), evolutional algorithms, advanced statistical functions, visualization tools, and etc. These techniques can accelerate analysis, leading to insights from both traditional and nontraditional data sources. This work adapted DynamicK-reference clustering algorithm [5] an improved data analytic model for the analysis of the heterogeneous data types of the extracted Boko Haram insurgency attacks mixed large dataset.

- DynamicK-reference clustering algorithm: DynamicK-reference clustering algorithm is a subset of Artificial Intelligence model. It is a hybrid of swarm intelligence and unsupervised machine learning algorithms which possesses robust and efficient analytic strength to overcome the challenges and difficulties big data analysis posed on traditional data mining models. It is the integration of Dynamic Multi-Swarm Optimization (Swarm Intelligence algorithm) and K-reference clustering (Partitional Clustering algorithm) algorithms. K-reference clustering algorithm is a Partitional Clustering algorithm that is a hybrid of RDW_K-means and K-representative clustering algorithms. "It has the potentials to cluster large mixed dataset [5]". Reference Distance Weighted (RDW) is a distance measure parameter used to improve the performance of the traditional K-means clustering algorithm [5].

From Fig. 1, the Boko Haram insurgency raw dataset attacks were scraped from social media and stored in the database for data mining model preprocessing. The veracity, variability and validity of the extracted datasets were achieved using data mining model which includes: selection, cleaning, integration and storage. The data mining preprocessing removed the "abnormalities, outlier and redundancy" (variability) in the big dataset and makes it informative (veracity) and clean (Validity) for analytical processing in the database. As the veracity, variability and validity of the dataset were achieved using data mining model, next is the mechanism to solve variety of big data analysis. Variety implies that big data are heterogeneous; such that data come from different sources with different types such as unstructured, structured and semi-structured. As the volume of data is high, it requires intelligence to identify efficient techniques to archive accurate data analysis.

The perfect technique to handle this is the application of DynamicK-reference clustering algorithm. Its mechanisms are in three phases. The first phase is to deploy intelligent sub-warms (search agents) to exploit the high dimensional and massive datasets of the n-dimensional space. The intelligent search agents provide information on the positions of the instances or particles (i.e. the best K value) of the unstructured mixed dataset. The second phase is to benchmark the different best k values obtained by each sub-warm (intelligent search agents) and select the global best k values. In the third phase the global k value obtained by the intelligent search agents in the second phase guides the K-reference clustering algorithm to accurately cluster the mixed datasets into similar and dissimilar clusters. The clustered outputs are visualized and used for decision control.

3.2 RWDK-means Clustering Algorithm

The RWD-K-means Clustering Algorithms are as follow:

"**Step one**: obtain the global best value of clusters k from the result of Dynamic Multi-swarm algorithm.

Step two: Randomly selecting the centroids (v_1, v_2...v_k) in the data set.

Step three: Calculate the Reference Distance Weight (RDW) [(1)] of the corresponding centroids (v_1, v_2...v_k) of the data points in the dataset.

Step four: Calculate RDW(s, t, α)=$\frac{\sum_{i=1}^{n}\alpha_i\left|\frac{s_i-t_i}{s_i}\right|}{n}=\frac{\sum\left|\frac{s_i-t_i}{s_i}\right|}{n}$

i = 1, 2,..,n. [12] (1)

Where: RDW is the corresponding reference weight vector to the $s_{i,}$

s = {s_1, s_2... s_n} is the features-vector of reference, associated to a class, vector from whom the distance of the data points is measured [12];

t = {t_1, t_2... t_n} is the feature-vector associated to the problem that must be solved or object that must be classified in the dataset.

α = {α_1, α_2.., α_n}, is a vector called relevance vector, whose components α_i are parameters specific for each feature in part and assigned to each feature. Relevance factor, is proportional to the importance/weight of the respective feature under the conditions of problem to be solved or the objects (data points) of the datasets to be clusters [12].

Step five: Find the distance between the centroids using the integration of Reference distance weighted and Euclidean Distance equation.

$$d_{ij} = \left\|RDW_i(s_i - t_k)\right\|^2 \qquad (2)$$

Step six: Update the centroids Stop the process when the new centroids are nearer to old one. Otherwise go to step-four, [5]".

3.3 Dissimilarity Measure of Categorical Domain

The categorical domain of mixed dataset analysis from the review of literatures is analyzed using the dissimilarity measures between categorical objects and the representative of a cluster, is defined based on simple matching approach of chi-square expression. The process is as follows:

Given C = {S_1, ..., S_P} is a cluster of categorical objects, with

S_i = (s_1, ..., $s_{i,m}$), 1 ≤ i ≤ p, and S = (s_1, ... , s_m) a categorical object.

In some cases S may or may not belong to C. Assume that Q = (q_1, ..., qm), with t_j = {(c_j, rf_{cj})|c ϵ D_j}, is a representative of cluster C.

Now, the dissimilarity between object S and representative Q is defined by:

$$d(S,Q)=\sum_{j=1}^{m}\sum_{c_j \varepsilon D_J} rf_{cj}.\delta(s_j,c_j) \qquad (3)$$

Where c = cluster and rf = relative frequency between the clusters [15].

Under such a definition, the dissimilarity d (S, Q) is mainly dependent on the relative frequencies of categorical values within the cluster and simple matching between categorical values. It is also of interest to note that the simple matching dissimilarity measure between categorical objects can be considered as a categorical counterpart of the squared Euclidean distance measure. It is easily seen that

$$d(S,Q)=\sum_{j=1}^{m}\sum_{c_j \varepsilon D_J} rf_{cj}.\delta(s_j,c_j) \qquad (4)$$

$$= \sum_{j=1}^{m} \sum_{c_j \varepsilon D_j, c_j \neq x_j} rf_c \qquad (5)$$

$$\sum_{j=1}^{m}(1 - rf_{x_i}) \qquad (6)$$

where rf_{xj} is the relative frequency of category x_j within C.

The enhanced k-means algorithm RDW_K-means clustering algorithm was integrated with k-representative clustering algorithm to produce K-reference clustering algorithm for clustering mixed dataset. The K-reference clustering algorithm is a hybrid of RDW_K-means algorithm and k-representative algorithm. "The dissimilarity between two mixed-type objects S and T, which are described by attributes Ar_1, $A^r{}_2$,.., $A^r{}_p$, $A^c{}_{p+1,\dots}$ $A^c{}_m$, can be measured by [5]":

$$d_2(S,T) = RDW_i \sum_{j=1}^{p}(s_j - t_j)^2 + \gamma \sum_{j=p+1}^{m} rf\delta(s_j, t_j) \qquad (7)$$

Where RDW = Reference Distance Weighted, and the numerical attribute is the first term in question (7) while the second term is the simple matching dissimilarity measure on the categorical attributes. To avoid the computing process favoring either the categorical or numerical attributes the parameter weight γ is added to the process. The influence of γ in the clustering process is discussed in [13].

The input of the proposed K-reference clustering algorithm is the global best value of k obtained from the swarm intelligent mechanism. It is used by the clustering algorithm to produce efficient clustering outputs (see Fig. 1).

The distance d ($\vec{x}$; $\vec{s}$) between two points and in the n-dimension space is defined as the Euclidean distance as shown in equation (8)

$$d(\vec{x}, \vec{s}) = \sqrt{\sum_{i=1}^{n}(x_i - s_i)^2} \qquad (8)$$

According to our experimental experience, the more peaks in the landscape, the more child swarms are relatively needed. r is relative to the range of the landscape and the width of peaks. In general, we set r according to the following equation:

$$\mathrm{r} = \sqrt{\sum_{i=1}^{n}\left(x_i^u - x_i^l\right)} / \left(W_{\min} + c\left(W_{\max} - W_{\min}\right)\right) \qquad (9)$$

where Wmin and Wmax are the minimum and maximum, they are the lower and upper bound on the i-th dimension of the variable vector of n dimensions [14].

3.4 Clustering Accuracy

A cluster is called a pure cluster if all the objects belong to a single class.

The clustering accuracy 'r' is defined as

$$r = \frac{1}{n}\sum_{i=1}^{k} a_i \qquad (10)$$

Where a_i is the number of data objects that occur in both cluster C_l and its corresponding labeled class. a_i has the maximal value and n is the number of objects in the data set. The clustering error e is defined as:

e = 1 – r. (11)

3.5 DynamicK-reference Clustering Algorithm

The DynamicK-reference Clustering Algorithm includes:

Step 1: "Initialize the swarm by randomly identifying each particle's (data points or objects) position and velocity in search space. Such that all attractors are set to randomized particle position; set swarm attractors (the best of the particle attractor) to particles attractor 1(best position obtain by that particle so far) and store all function values to function floor.

Step 2: REPEAT FOR EACH swarm f (f is the number of swarms to be deployed)// Test for change. Compute function at swarm attractor for swarm n (n is the sub-swarm deployed at that region). Note: the swarm attractor enables information sharing between particles while the particle attractor severs are individual particle memories.

Step 3: IF value (best position) of next result is distinct from previous iteration THEN

3.1: Re-compute function value at each particle attractor.

3.2: Bring up-to-date swarm attractor and stock function values.

Step 4: WHILE particle k of swarm f, // Bring up-to-date Attractor.

4.1: Compute function at updated position and store value.

4.2: IF value of next result is better than particle attractor value THEN

4.3: Particle attractor k: = position and value of particle k. (where k is the current position of the particle)

4.4 IF value of next result is better than swarm attractor value THEN

4.5: Swarm attractor: = position and value of particle k.

FOR EACH swarm y ≠f (y is the sub-swarms) //Exclusion

Step 5: IF swarm attractor p_f is within r_{excl} of p_y THEN

5.1: Randomize the swarm with the worse swarm attractor value.

5.2: Re-instruct particle attractors, compute f at each new position, stock these values, and place attractor of swarm to the position of its best particle). UNTIL number of functions evaluates performed > max

Step 6: Obtain the global variable from best variable K of each sub-swarm.

Step 7: Calculate the Reference Distance Weighted of the corresponding object representative. (see equation (2)).

Step 8: Calculate similarity and dissimilarity of the particles (see equation (7)).

Step 9: Compute dissimilarity measure for the mixed dataset (categorical and numerical)

Step 10: Compute the Relative Frequency (rf) of the dissimilarity matrix of the mixed dataset

Step 11: Proceed to compute the clusters using the rf of the dissimilarity matrix as a parameter.

11.1: For each X_i, calculate the dissimilarities d (X_i, Y_i), i = 1, . . ., k. Then, reassign X_i to cluster C_l (from cluster Cu, say) such that the dissimilarity between Xi and Y_l is least. Update both Y_l and Q.

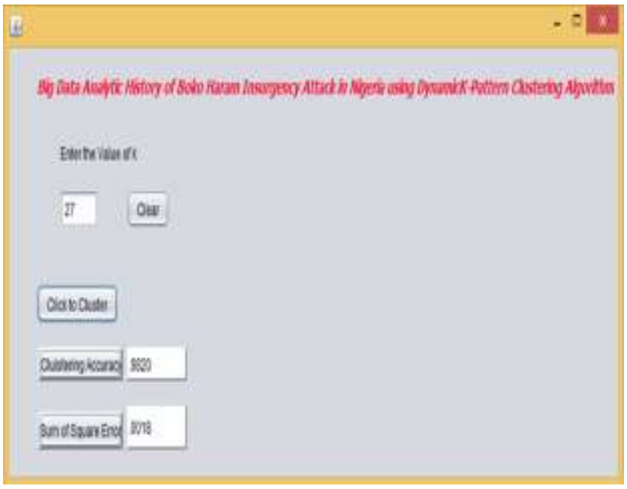
Step 12: If no object has changed clusters after a full cycle test of the whole data set, go to Step7

Step 14: else output results.

Step 13: Stop, [5]".

4 RESULTS

The system interface is shown in Fig. 2.



Big Data Analytic History of Boko Haram Insurgency Attack in Nigeria using DynamicK-Pattern Clustering Algorithm

Enter the Value of k

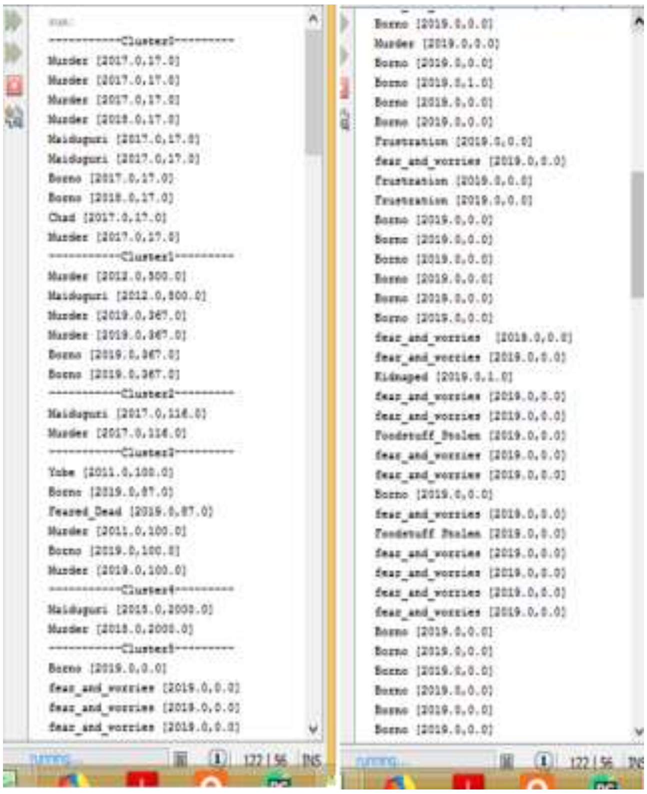
27

Clustering Accuracy 9920

Sum of Square Error 8018

Fig-2: Clustered output of Boko Haram Attacks in Northern Nigeria

The user enters the number of k clusters and then clicks on "Click to cluster" to analyze the results. As the user clicks on "Click to cluster" the system uploads the dataset to produce clustered outputs. The clustered analysis of Boko Haram insurgency attacks in Northern Nigeria is shown in Fig. 3.



```
------------Cluster0----------
Murder [2017.0,17.0]
Murder [2017.0,17.0]
Murder [2017.0,17.0]
Murder [2018.0,17.0]
Maiduguri [2017.0,17.0]
Maiduguri [2017.0,17.0]
Borno [2017.0,17.0]
Borno [2018.0,17.0]
Chad [2017.0,17.0]
Murder [2017.0,17.0]
------------Cluster1----------
Murder [2012.0,500.0]
Maiduguri [2012.0,500.0]
Murder [2019.0,367.0]
Murder [2019.0,367.0]
Borno [2019.0,367.0]
Borno [2019.0,367.0]
------------Cluster2----------
Maiduguri [2017.0,116.0]
Murder [2017.0,116.0]
------------Cluster3----------
Yobe [2011.0,100.0]
Borno [2019.0,87.0]
Feared_Dead [2019.0,87.0]
Murder [2011.0,100.0]
Borno [2019.0,100.0]
Murder [2019.0,100.0]
------------Cluster4----------
Maiduguri [2015.0,2000.0]
Murder [2015.0,2000.0]
------------Cluster5----------
Borno [2019.0,0.0]
fear_and_worries [2019.0,0.0]
fear_and_worries [2019.0,0.0]
fear_and_worries [2019.0,0.0]
```

```
Borno [2019.0,0.0]
Murder [2019.0,0.0]
Borno [2019.0,0.0]
Borno [2019.0,1.0]
Borno [2019.0,0.0]
Borno [2019.0,0.0]
Frustration [2019.0,0.0]
fear_and_worries [2019.0,0.0]
Frustration [2019.0,0.0]
Frustration [2019.0,0.0]
Borno [2019.0,0.0]
Borno [2019.0,0.0]
Borno [2019.0,0.0]
Borno [2019.0,0.0]
Borno [2019.0,0.0]
Borno [2019.0,0.0]
fear_and_worries  [2019.0,0.0]
fear_and_worries [2019.0,0.0]
Kidnaped [2019.0,1.0]
fear_and_worries [2019.0,0.0]
fear_and_worries [2019.0,0.0]
Foodstuff_Stolen [2019.0,0.0]
fear_and_worries [2019.0,0.0]
fear_and_worries [2019.0,0.0]
Borno [2019.0,0.0]
fear_and_worries [2019.0,0.0]
Foodstuff Stolen [2019.0,0.0]
fear_and_worries [2019.0,0.0]
fear_and_worries [2019.0,0.0]
fear_and_worries [2019.0,0.0]
fear_and_worries [2019.0,0.0]
Borno [2019.0,0.0]
Borno [2019.0,0.0]
Borno [2019.0,0.0]
Borno [2019.0,0.0]
Borno [2019.0,0.0]
Borno [2019.0,0.0]
```

Fig-3: Clustered Output of Boko Haram Insurgency from cluster 0 to clusters 5.

The results show the output clusters of "number of deaths resulting from Boko Haram attacks"; "different areas the attacks took place"; and "date, month and year of the attacks" from cluster 0 to cluster 27. The output results of the proposed system gave different clusters assigned to different group of similarities and dissimilarities clusters. The percentage distribution of clustered result of Boko Haram insurgency's area of attack is shown in Fig. 4.

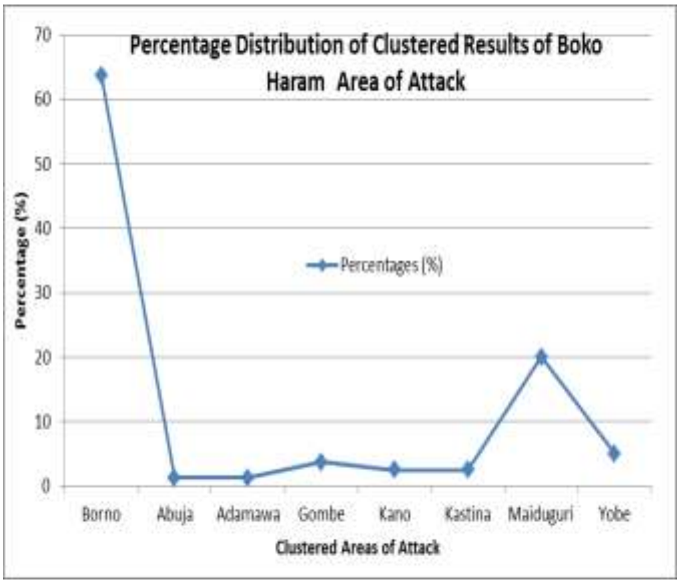


Fig-4: Percentage Distribution of clustered Boko Haram Attacks in Different Areas in Nigeria.

The results show that Borno State has 63.75% attacks in its other territories and Maiduguri its capital has 20% attacks which gave a total of 83.75% attacks in Bonor state. Abuja has 1.25% attacks, Adamawa State has 1.25 attacks, Gombe State has 3.75% attacks, Kano State has 2.5% attacks, Kastina State has 2.5% attacks, and Yobe State has 5% attacks. Borno State has the highest Boko Haram attacks followed by Yobe state and the others. The clustered distribution of number of deaths for Boko Haram attacks in Nigeria is shown in Fig. 5.

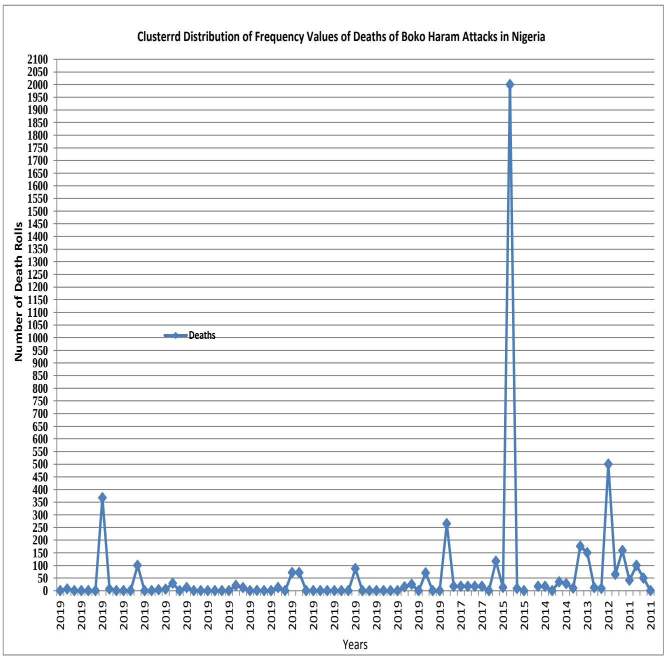


Fig-5: Graphical representation of the clustered distribution of Boko Haram attack of number deaths in Nigeria.

The clustered results were plotted to observe and compare the extent of the attacks at different presidential regimes in Nigeria, and the efforts of Nigeria army to control and eliminate Boko Haram insurgency attacks in Northern Nigeria. From the clustered results more than 2,000 people were killed from 2009 to 2015 during Boko Haram Insurgency attacks. The number of attacks and killings reduced in 2016 and early 2017, this was the period when Mr. President Muhammadu Buhari claimed that Boko Haram insurgency has been defeated. From 2018 till date, there have been constant insurgency's attacks in Northern Nigeria. The Boko Haram attacks in this present administration are worse more than when it started during Mr. President Goodluck Jonathan's regime in 2008.

A clear description of monthly attacks of Boko Haram insurgency attacks in the Northern Nigeria from the year 2008 to 2019 is shown in Fig. 6.

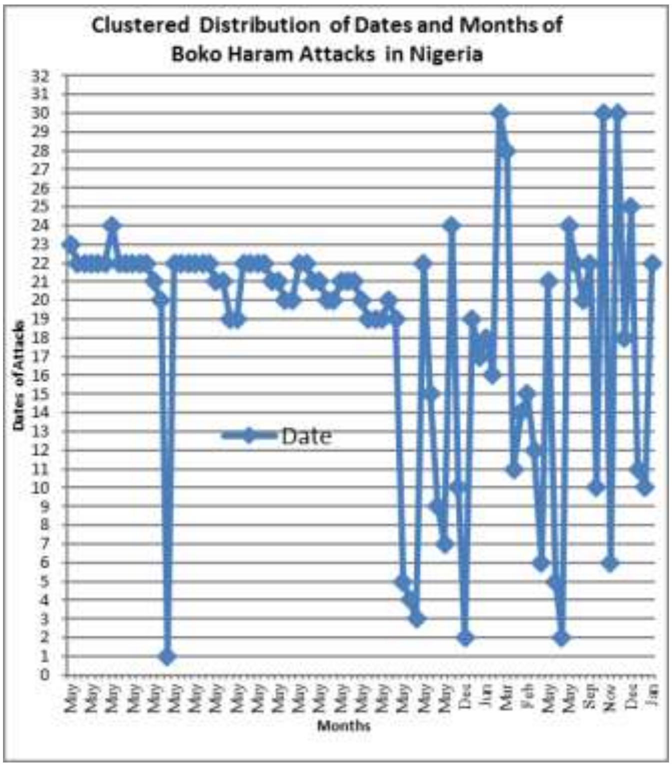


Fig-6: Graphical Representation of the clustered distribution of dates of attacks and months of Boko Haram attacks in Northern Nigeria.

From the results, it was observed that in the month of May in 2019 there have been constant attacks of Boko Haram insurgency in Northern Nigeria. There was an attack on 23th of May 2019, and more than eight attacks on 22nd of May at different time intervals of that date. Furthermore, there have been various attacks recorded on 21st, 20th and 19th of May 2019. On 19th May, there was five (5) attacks at different time intervals. Also, on 20th May, there was seven (7) attacks and on 21st of May, there was ten (10) attacks at different time intervals of that date. This implies that Northern Nigeria borders are extremely porous for insurgency attacks; there were little or no resistance of the militaries defense before the attacks took place. The clustered distribution of dates and years of Boko Haram attacks in Nigeria is shown in Fig. 7.

The clustered results were plotted to observe and compare the extent of the attacks at different presidential regimes, and the effects of Nigeria army to control and eliminate Boko Haram in Nigeria. The clustered output shows that the attacks of Boko Haram insurgency reduced at the first year of President Muhammadu Buhari administrations in 2015.

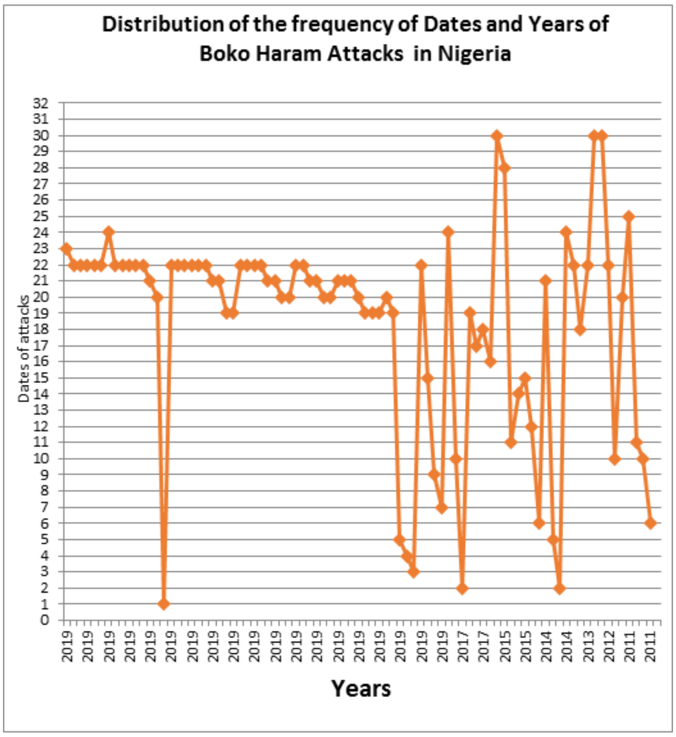


Fig. 7: Graphical Representation of the clustered distribution of dates of attacks and years of Boko Haram attacks in Nigeria.

At that period Boko Haram attacks was reduced in Nigeria, Mr. President claimed that Boko Haram has been defected. From 2017 the insurgency attacks begin again and gets worsen in 2019 in Borno most especially in its capital Maiduguri and other state that is close to the Northern Nigeria's borders. In the regime of President Muhammadu Buhari mostly in 2019 Boko Haram attacks in the northern Nigeria have increased more than when comparing to the time it started on 2008 during the regime of President Goodluck Jonathan. This ugly situation of killings of Nigerians and destroying of properties prompted the former President, Olusegun Obasanjo to comment on the implications and hopeless state of the displaced thousands of Nigerians in the northern states of Nigeria.

5. CONCLUSIONS

In this work, Boko Haram Insurgency Attacks datasets was scraped from social media sites. The dataset was analyzed using the mechanisms of big data analytics and DynamicK-reference clustering algorithms. The mechanisms provided an achievable analytical platform to analyze the historical big dataset of terrorist attacks in Northern Nigeria for military decision support system. Considering the analytical challenges and unique characteristics of big data analytical processes, the DynamicK-reference clustering algorithm was adopted to unveil the hidden patterns of the mixed lager datasets of insurgency attacks in Nigeria. The DynamicK-reference clustering algorithm was resourceful enough to provide clustering accuracy of 0.9820 and clustering sum of square error of 0.0018 from the analysis of Boko Haram

Insurgency attacks dataset. The clustered results identified various strategies of Boko Haram attacks in Northern Nigeria. There were constant attacks of Boko haram Insurgency in the northern Nigeria. The continued attacks led to millions of people currently displaced and killed in the Northern Nigeria most especially in Borno, State.

The repeated Boko Haram insurgency attacks in Northern Nigeria at different time intervals on the same day indicate that the military defense frontline is weak. The security lapses were due to the insufficient commitment of soldiers at the frontline. The setback in operations of soldiers in the ongoing counter-insurgency and counter-terrorism operations in the north-east was also confirmed by the Chief of Army Staff, Lt.-Gen. Tukur Buratai's reports, during military transformational workshop in Abuja on June 18th, 2019. We therefore, recommend intelligent automated systems (such as: crime-mapping platforms, gunshot-detection system, video analytics, smart lights and etc.) or robotic detection systems to be implemented at the country's border frontlines for efficient monitoring and communication of attacks. This will eliminate the setbacks due to insufficient willingness of soldiers to perform assigned tasks.

ACKNOWLEDGEMENT

We would like to thank and appreciate the Departments of Computer Science, Federal University of Technology, Owerri, Imo State, Nigeria for providing the enabling environment that facilitated this work.

REFERENCES

[1] K. D. Strang and S. Alamieyeseigha, What and where are the risks of international terrorist attacks: Adescriptive study of the evidence, International Journal of Risk and Contingency Management. 4, 1-18, 2015.

[2] Strang, D. K. and Sun Z. Analysis relationships in terrorism big data using Hadoop and statistics, Journal of Computer Information Systems, DOI:10.1080/08874417.2016.1181497, 57(1): 67-75. 2017.

[3] J. Rivinius, Majority of 2013 terrorist attacks occurred in just a few countries, in press Release. National Consorttium for the study of Terrorism and Responses to Terrorism (START), University of Maryland, Baltimore, 1-2, 2014.

[4] Abiodum, O. i., Janta, A., Omolara, A. E. Singh, M. M.m Abukakar, L. Z., and Umar, A. M. Big Data: An Approach for Detecting Terrorist Activities with People's Profiling. Proceedings of the International MultiConfrence of Engineering and Computer Science, VoL I IMECS, Hong Kong, 2018

[5] C. P. Oleji, E. O. Nwachukwu, G, Chukwudebe and E. C. Nwokorie. Improved Model for Big Data Analytics Using Dynamic Multi-Swarm Optimization and Unsupervised Learning Algorithms. International Research Journal of Engineering and Technology (IRJET).Unpublished.

[6] Rebollo, C. S.,Puente, C., Palacious, R., Piriz, C., Fuentes, J. P. and Jarauta, J., Detection of Jihadism in Social Network Using Big Data Techniques Supported by Graphs and Fuzzy Clustering. Hindawi https://doi.org/10.1155/2019/1238780. 1-14.

[7] Kolajo, T. and Daramola, O. Leveraging big data to combat terrorism in developing countries. Conference on Information Communication Technology and Society (ICTAS), DOI:.110910/ICTAS.2017 .7920662, (2017).

[8] Tour, I. and Gangopadhyay, A. (2016) Real time big data analytics for predicting terrorist incidents. IEEE Symposium on Technologies for Homeland Security (HST), Waltham, MA, USA, DOI: 10.1109/THS.2016.7568906.

[9] Einstein, A. (2018). Addressing the cybersecurity skills gap through cooperation, education and emerging technologies. https://cybertechaccord.org/uploads/prod/2019/01/TechAccordWP-AddressingCyberSkillsGap_Jan2019.pdf. 1-7.

[10] M. K. Maya, Application of Machine Learning and Deep Learning in Cybercrime Prevention- A Study. National Conference on Research Trend in Big Data and Intelligent Computing organized by Departement of Computer Science. International Journal of Trend in Research and Development (IJTRD), issn: 2394-9333. www.ijtrd.com. 1-4, 2019

[11] Karie, N. M., Kenande, V. R. and Venter, H.S. (2019). proposed diverging deep learning computing techniques into cyber forensics. Forensic Science International: synergy, 1, 61-67.

[12] L.O. Mafteiu-Scai, A new Dissimilarity Measure between Feature-Vectors. International Journal of Computer Applications, Vol.17. pp. 0975-8887, 2013.

[13] Z. Huang: Extensions to the k-means algorithm for clustering large cybernetics 28C, (1998), pp. 219-230, 1998.

[14] T. Backwell, and J. Branke: Multi-swarm Optimization in Dynamic Environments, Department of computing, Goldsmiths College, University of London New Cross, London SE14 6NW, U.K., pp. 1-12., 2004.

[15] P. C. Oleji, N. E. Nwokorie and D. O. Onuodu, Clustering Mixed Dataset with Multi-Swarm Optimization and K-prototype Clustering Algorithm. Nigeria Computer Society (NCS) 26th National Conference & Exhibition; Information Technology for National Safety & Security, Abuja Nigeria, 27, 205-221., 2016.