An Intelligent Crime Prediction System Using Artificial Neural Network And Geographical Information System

P. B. Fajemiseye, C. Ugwu, L. N. Onyejegbu

Abstract

Crime is a social problem that has eaten deep into the society. The rate at which innocent blood is wasted on a daily basis and the display of bottled-up frustration by the citizens remains a cause for concern. Many have lost their loved ones, investment and absence of safety in most parts of the country. The aim of this study is to develop and implement a model that embedded Geographical Information Systems (GIS) and Artificial Neural Networks (ANN) for the analysis of observed data to predict crimes. ANN provides the system with learning ability in such a way that the system can learn from both past and present event. To accomplish this, Self-Organizing Map (SOM) and K-means clustering algorithm are used to cluster dataset to increase the learning rate of the system and Geographic Information Systems (GIS) provide the system with a digital representation that enables the user to map crimes locations analytically and descriptively. The methodology employed in this project is the object oriented methodology. The system was implemented using PHP programming language and MySQL database. The results show that the new system has higher degree of operation compared to the existing system. It was discovered from the result that the new system is faster than the existing system and give better predictive value.

Keywords: Crime, Information System, Spatial dataset of crime series.

INTRODUCTION

In recent times, Crime has become a social canker worm that has eaten profound into the social texture of the Nigerian society with the end goal that its impact is multifaceted. The concept of crime refers to violation of principles and rules that were believed to be unique fitting for the binding any community or a nation. Despite the fact that Adegoke, (2014) opine that crime is inevitable in our social life, it is an integral part of societies and it is functional, but functionality in a society such as ours has to be viewed seriously because of the social and psychological problems it has caused many victims. Truth be told, regardless of the usefulness of Crime in the general public, the demonstration of Crime is condemnable and inadmissible in a solid society, regardless of the defense culprits may show.

The challenges of crime and insecurity have assumed formidable dimensions forcing the entire nation to rue, the loss of their loved ones, investment and absence of safety in most parts of the country. The rate at which innocent and guiltless blood is squandered every day and the presentation of restrained dissatisfaction by the residents remains a reason for concern. Nwaze, (2011) clarifies that the rate of bloodshed during the Nigeria civil war is a piece of cake compared to the terrorist attacks in few months.

During the 1970s 80s, the well-known Crimes that were predominant in Nigeria include: outfitted theft, armed robbery, stealing, assault, burglary, rape etc. However, today terrorism, bomb blasts, kidnapping, drug trafficking, money laundry, child trafficking, assassinations and other criminal exercises have turned into the request of the day (Chinwokwu, 2013). The rate of brutal violations, for example, as terrorism, kidnapping, armed robbery and banditry, suicide bombing, religious killing, ethnic clashes, politically-motivated killings and other forms of criminal activities in the country is becoming increasingly regular occurrence that characterized life in the nation. Nigeria has reliably positioned low in the Worldwide Peace List (GPI, 2012), implying a compounded condition of Crimes and instability in the nation. Victoria (2010) noted that "Nigeria was plagued with myraid of security difficulties, kidnapping, terrorism, civil disturbance, political violence, burglary attacks, fraud, assassination, armed robbery, among others" . Regardless of stringent laws and disciplines to check these violations, they have kept on being on the

expansion with the police apparently powerless and unequipped for savaging the circumstance.
Expanding pattern of Crimes in a general public shows that the security offices need to bear the weight of hoodlums in bigger numbers than the past. An impressive number of lawbreakers has just been sentenced and detained; some of them have been discharged in the wake of finishing their detainment and discharged under supervision, while numerous offenders are submitting their exercises inside the common society without being gotten. This unending pattern has made the way toward examining Crimes extremely perplexing. Sri Lanka Police too has been looking with these challenges for quite a while. As per insights, all things considered, 50000 grave Crimes are accounted for every year in Nigeria. When a Crime is accounted for, it is recorded and examined by the important administrative police headquarters of the specific region.

The basic roles and duty of government of any country is the provision of security and ensuring safety of lives and property. The role is stated in the 1999 constitution of the Federal Republic of Nigeria that "the security and welfare of the general population will be the basic role of the administration" (Area 14(2) (b). Crime has become very sophisticated in Nigeria and fighting crime with the old traditional methods do not seem to yield productive desired results. Because crime impacts on our sense of security, affects our quality of life, and has far reaching economic consequences, crime protection and crime combat have gained significant importance in the general public, police agencies, politics, and science. To counteract the impact of crime throughout society, crime agencies, among other things, have reverted to the application of modern intelligence and geospatial technologies, which have quickly become an emerging scientific research field for tackling such security needs. The development of such technologies is often founded on a solid theoretical basis that includes such well-known theories as routine activities (Felson and Poulsen 2003), rational choice (Clarke and Cornish, 1993), and environmental criminology (Brantingham and Brantingham, 1997). The advent of intelligent crime prediction system (ICPS) will provided law enforcement agencies with a reasonable answer to the above questions and powerful tools that will aid crime control and crime prevention efforts. The system is an hybrid system which is made up of Artificial Neural Network (ANN) and Geographic Information System (GIS).

2. Related Works

Zakaria and Ayman (2014) proposed a model for crime and criminal data analyzes using simple k-means algorithm for data clustering and Aprior algorithm for data Association rules. The paper sought to help specialist in discovering patterns and trends, making forecasts, finding relationships and possible explanations, mapping criminal networks and identifying possible suspects. Clustering is based on finding relationships between different Crime and Criminal attributes having some previously unknown common characteristics. Data for both crimes and criminals were collected from police departments' dataset to create and test the proposed model, and then these data were preprocessed to get clean and accurate data using different preprocessing techniques. The preprocessed data were used to find out different crimes and criminal trends and behaviors, and crimes and criminals were grouped into clusters according to their important attributes.

Madhusmita and Behera (2012) developed a model that uses K-Means Algorithm for information clustering and hereditary calculation is connected for instatement. Adjusted K-Means calculation is tried on an iris informational index and its execution is contrasted and other bunching calculation and is discovered to be more exact and perform superior to standard K-Means with less number of grouping and quantization mistakes and can be connected notwithstanding for high dimensional dataset. It handles a portion of the issues of K-means, for example, obscure number of groups and the affectability to introductory centroid

Majeed (2010) used Regression analysis for building a predictive crime model and combined with Neural Networks with the aim of developing a new hierarchical neural Network approaches to generate a more reliable prediction. It encompasses the GIS, SCS algorithm, statistics and neural networks for developing a hybrid predictive crime model, mapping, visualizing crime data and the corresponding population in the study region, visualizing the location of obtained clusters, and burglary incidence concentration 'hotspots' which was specified by clustering algorithm SCS. The promising results were compared with the non-hierarchical neural Network back-propagation network and multiple regression analysis. The average percentage accuracy achieved by the new methodology at testing stage increased 13% compared with the

non-hierarchical BP performance. Crime detection and prevention techniques were applied to different applications ranging from cross-border security, Internet security to household crimes.
Dawei et. al. (2013) proposed a spatial information mining system to ponder Crime hotspots through their related variables. Creator utilized Geospatial Discriminative Examples (GD Patterns) to catch the critical contrast between two classes (hotspots and typical regions) in a geo-spatial dataset. Using GD Patterns, a novel model was developed Hotspot Enhancement Device (HOT)— to enhance the distinguishing proof of Crime hotspots. At long last, in view of a similarity measure, GD Pattern bunches were gathered and envision the appropriation and attributes of Crime related variables.

The consequences of this methodology was assessed utilizing a certifiable dataset gathered
Javad (2014) gives an audit about mining helpful data by methods for Information mining. Author specified despite the fact that information mining can be connected to different fields, one of the pivotal field is criminology where information digging used for recognizing Crime qualities. Distinguishing and investigating violations and researching their relationship with hoodlums are engaged with the breaking down Crime process. Creator uncovers that criminology is a suitable field for utilizing information mining techniques which shows the high volume and the complexity of relationships between crime datasets. The paper aims to provide useful information by means of Data Mining, so that crime hot spots are identified and crime trends are predicted.
Yamuna et al. (2012) proposed information mining strategies to dissect and anticipate the future crime. The forecast of future Crime patterns includes following Crime rate changes starting with one year then onto the next and utilized deciding to extend those progressions into what's to come. The fundamental strategy includes group the states having the equivalent crime trend and after that utilizing ‖next year‖ bunch data to arrange records. This is joined with the state neediness data to make a classifier that will foresee future Crime patterns. To the bunched outcomes, a characterization calculation was applied to anticipate the future Crime design. The characterization was performed to discover in which classification a bunch would be in the following year. This enables us to manufacture a prescient model on foreseeing one year from now's records utilizing the current year's data. The choice tree calculation was utilized for this reason. The summed up tree was utilized to anticipate the obscure crime trend for the following year. The exploratory outcomes demonstrated that the strategy utilized for expectation is precise and quick.

Malathi (2013) worried that national security has been expanded after Mumbai TajHotel assault on 26.11.2008. This was occurred as criminal and psychological oppressor exercises. In this way, to break down and unravel such kind of issues creator connected information mining with regards to law implementation and insight investigation. Creator used clustering/arrangement based model to envision Crime patterns. The paper likewise convey different information mining approaches and systems which can be connected for Crime design and different examination of city Crime, for example, property crime, violent Crime, Crime against ladies and different violations. The aftereffects of this information mining could conceivably be utilized to lessen and even counteract Crime for the fourth coming years.

Shyam, (2014) proposed an idea to solve crime detection problems using data mining. Crimes are a social nuisance and cost our society dearly in several ways. This author look at the use of clustering algorithm (k-means clustering) to detect the crimes patterns and speed up the process of solving crimes. This clustering technique was applied to real crime data to validate the results. Author also used semi-supervised learning technique for knowledge discovery from the crime records to increase the predictive accuracy.
Anna and Christopher (2010) applied fuzzy association rule mining for community crime pattern discovery. Discovered rules are presented and considered at regional and national levels. To extract rare and novel rules from thousands of discovered rules, support metric was defined. It helps the law enforcement personnel to find out interesting and meaningful crime patterns rather than wasting time in uninteresting and rare (rules) patterns.

Goswami et al. (2010) proposed three diverse continuous example mining approaches that is Record channel, crossing point and Proposed calculation which depends on established Apriori algorithm. Author made similar examination by keeping dataset of 2000 exchanges and therefore Record channel demonstrated better than traditional Apriori calculation, Convergence approach demonstrated superior to anything Record channel approach lastly proposed algorithm demonstrated that it is greatly improved

than other incessant example mining calculation. Revatthy and Satheesh (2012) present a point by point think about on bunching systems and its job on crime applications. Creators presents bunching one of the deciding system to join information objects into groups. The information protests inside the gathering are fundamentally the same as and extremely disparate too when contrasted with articles to other groups. Creators entire thoughts is this investigation will definitely enable the Crime to branch for better expectation and arrangement of crimes. It likewise helps the Crime investigators and law implementers to go before the case in the examination and help solving unsolved violations quicker. Creators likewise exhibits parcel bunching calculation a standout amongst other strategy for finding similarity measures.

P. B. Fajemiseye
Department of Computer Science
University of Port Harcourt
Nigeria
(peterbus2008@gmail.com)

L. N. Onyejegbu
Department of Computer Science
University of Port Harcourt,
Nigeria
(@uniport.edu.ng)

C. Ugwu
Department of Computer Science
University of Port Harcourt
Nigeria
(chidiebere.ugwu@uniport.edu.ng)

3. MATERIALS AND METHODS

In existing systems such as Zakaria and Ayman (2014), Malathi, and Santhosh (2013) and others has proposed a model for data analyzes using simple k-means algorithm for data clustering and Aprior algorithm for data association rules as shown in Figure 1. Clustering is based on finding relationships between different data attributes having some previously unknown common characteristics. The work tends to help specialists in discovering patterns and trends, making forecasts, finding relationships and possible explanations, mapping criminal networks and identifying possible suspects. The K-Mean algorithm used in the system is fast, robust, easier to understand and it gives best results when dataset are distinct or well separated from each other; but this K-Means algorithm fails to give optimum result when it comes to clustering high dimensional dataset because its complexity tends to make things more complicated when more number of dimensions are added.

The K-means algorithm used in the existing work is stated as the cluster analysis procedure and it is analyzed to determine the properties of the dataset and the target variable. It is typically used to determine how to measure similarity distance. Basically, it functions as follows:

1. Input: The number of k and a database containing n objects.
2. Output: A set of k-clusters that minimize the squared-error criterion.

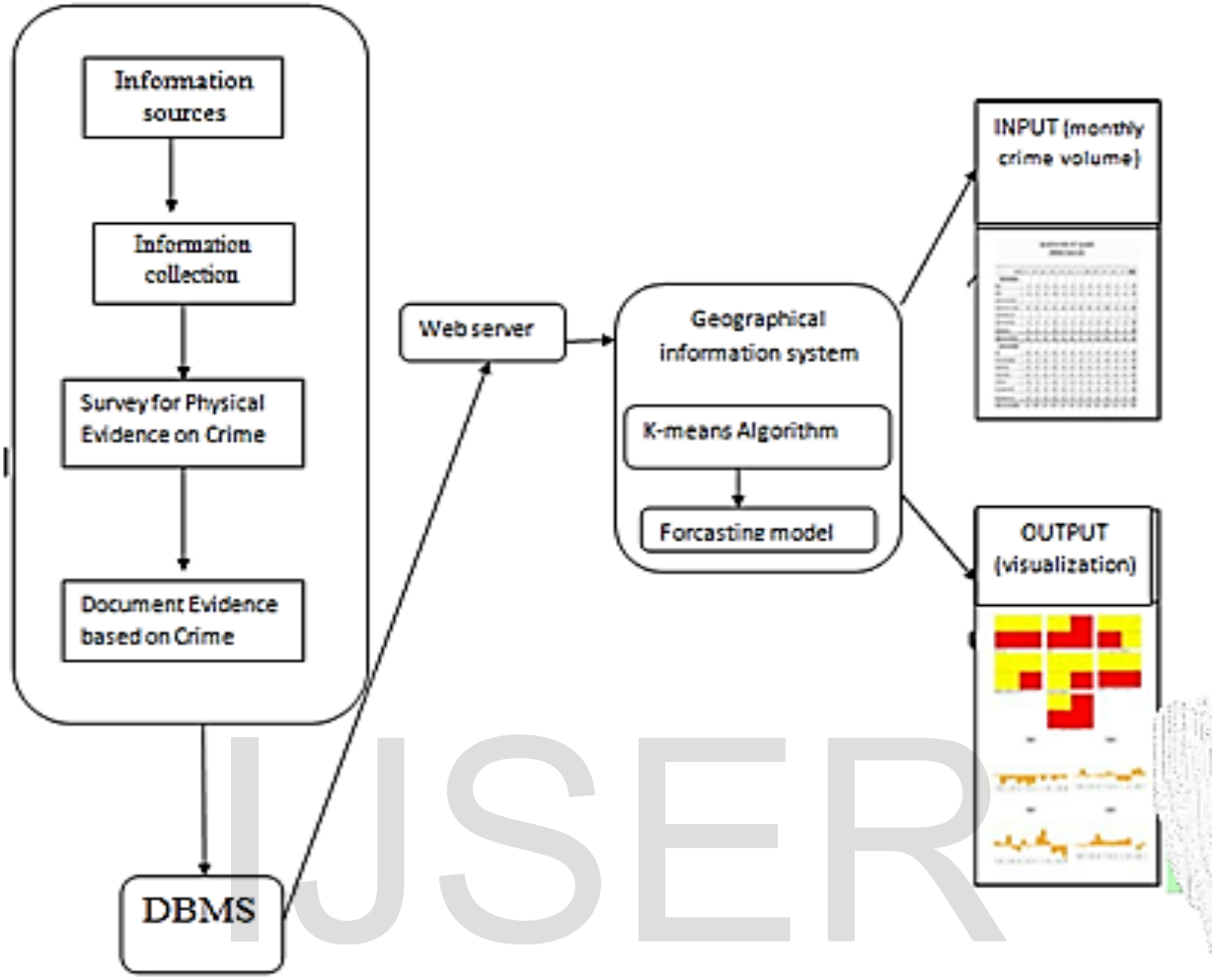

Figure 1: Architecture of the Existing System

4. PROPOSED SYSTEM

The proposed system is a hybridized system which makes use of K-means algorithm and Self-Organized Map (SOM algorithm) to effects the use of crime mapping using Geographical information system (GIS) analysis techniques to help users and law enforcement agencies tracking crime incidents and produce density maps showing overall crime patterns as shown in the Figure 2 below. With improvements in the availability and quality of crime data in digital format and more robust GIS software, researchers have expanded the use of GIS to the area crime prediction. To improve the performance, we proposed SOM with K-Means algorithm. Basically SOM is used in the system for training the dataset and to reduce the dimension of dataset before clustering (dimension reduction), whereas the K-means algorithm is used for clustering and determining the number of clusters.

The Self-Organizing Map (SOM) with K-Means algorithm is tested on large dataset and the performance is compared with traditional K-means clustering algorithm and found to be more accurate, with less number of classification and quantization errors and can be applied even for high dimensional dataset. It was found out that the approach gives better accuracy and better performance in terms of speed for predicting future occurrence of an event.

The model uses the Geographic Information Systems (GIS) to provide the system with a digital representation that enables the user to map crimes locations analytically and descriptively from their different geographical locations.

Data for both crimes and criminals were collected from police departments' dataset to create and test the model, and then these data were preprocessed to get clean and accurate data using different preprocessing techniques such as Conversion, Validation, Sorting, Summarization and Aggregation. The preprocessed data were

used to find out different crime and criminal trends and behaviors, and crimes and criminals were grouped into clusters according to their important attributes.

Finally, the geographical and spatial visualization is used for visualizing the rate at which crime is committed in a geographical location, whereas geographical and spatial visualization refer to a set of tools and techniques supporting the analysis of geospatial data through the use of interactive visualization. Traditional, static maps have a limited exploratory capability; the graphical representations are inextricably linked to the geographical information beneath. GIS and geo-visualization allow for more interactive maps; including the ability to explore different layers of the map, to zoom in or out, and to change the visual appearance of the map, usually on a computer display. Geo-visualization represents a set of cartographic technologies and practices that take advantage of the ability of modern microprocessors to render changes to a map in real time, allowing users to adjust the mapped data on the fly.

5. Algorithm of the Proposed System

This algorithm is an hybrid methods and employs specific properties of statistical distributions to determine centers with boundaries of each cluster drawn by suggesting a possible significance level. This algorithm is significantly less time consuming, efficient and objective for finding clusters within the datasets.

1. Invent the use of K-mean to locate the crime scene
2. Denote G as the node
3. Randomize the node weight vectors in a map
4. Randomly pick an input vector **D(t)**
5. Traverse each node in the map
    1. Use the Euclidean distance formula to find the similarity between the input vector and the map's node's weight vector
    2. Track the node that produces the smallest distance (this node is the best matching unit, BMU)
6. Update the weight vectors of the nodes in the neighborhood of the BMU (including the BMU itself) by pulling them closer to the input vector

$$W_v(s+1) = W_v(s) + \emptyset(u,v,s).\alpha(s).(D(t) - W_v(s))$$

7. Increase 8 and repeat from step 2 while $8 < 1$

A variant algorithm:

1. Initiate the SOM to located clustered environment of crime
2. Randomize the map's nodes' weight vectors
3. Traverse each input vector in the input data set
    1. Traverse each node in the map
    2 Use the Euclidean distance formula to find the similarity between the input vector and the map's node's weight vector
    3 Track the node that produces the smallest distance (this node is the best matching unit, BMU)
4. Update the weight vectors of the nodes in the neighborhood of the BMU (including the BMU itself) by pulling them closer to the input vector
5. $W_v(s+1) = W_v(s) + \theta(u,v,s).\alpha(s).(D(t) - W_v(s))$
6. Increase 8 and repeat from step 2 while $8 < 1$

When fed input data, the Euclidean distance, or the straight-line distance between the nodes, which are given a weight, is computed. The node in the network that is most similar to the input data is called the Best Matching Unit (BMU). As the neural network moves through the problem set, the weights start to look more like the actual data. The neural network has thus trained itself to see patterns in the data much the way a human sees. The approach differs from other AI techniques such as supervised learning or error-correction learning, but without using error or reward signals

to train an algorithm. Thus, a self-organizing map is a kind of unsupervised learning.

The algorithm was tested using artificial datasets with very promising results. Experimental results demonstrate that it is especially suitable for large dataset and even for small sample size. The attribution of the Hybridized algorithm are: easy to implement; no previous knowledge of the data set requirements; less number of performed steps leads to a reduction in clustering time; and the results provide the detailed information about the distribution of cases within the dataset. It performs reasonably well in terms of memory requirements; running time; cluster quality. This process was used to delimit the concentrated location.

6. METHODOLOGY

The methodology employed in this research is the object-oriented methodology. This system development methodology uses a recursive object-oriented development method for which the entire system is broken down into subsystems and modules. Due to the vast need to eradicate crime, information are needed in high demand in order to forecast crime using GIS and Artificial Neural Network. This reason serves as one of the limitations for researches in this area. The dataset used in this system was obtained from the open source crime data and GIS repository (Crime Index). The dataset consists of a thousand instances and each instance has twenty attributes that gave detailed information on how crime could be predicted, but these attributes were coded in numerical and symbolic form for confidential reasons. The two ways of interpreting SOM were used. Because in the training phase weights of the whole neighborhood are moved in the same direction, similar items tend to excite adjacent neurons. Therefore, SOM forms a semantic map where similar samples are mapped close together and dissimilar ones apart. This may be visualized by a U-Matrix (Euclidean distance between weight vectors of neighboring cells) of the SOM.

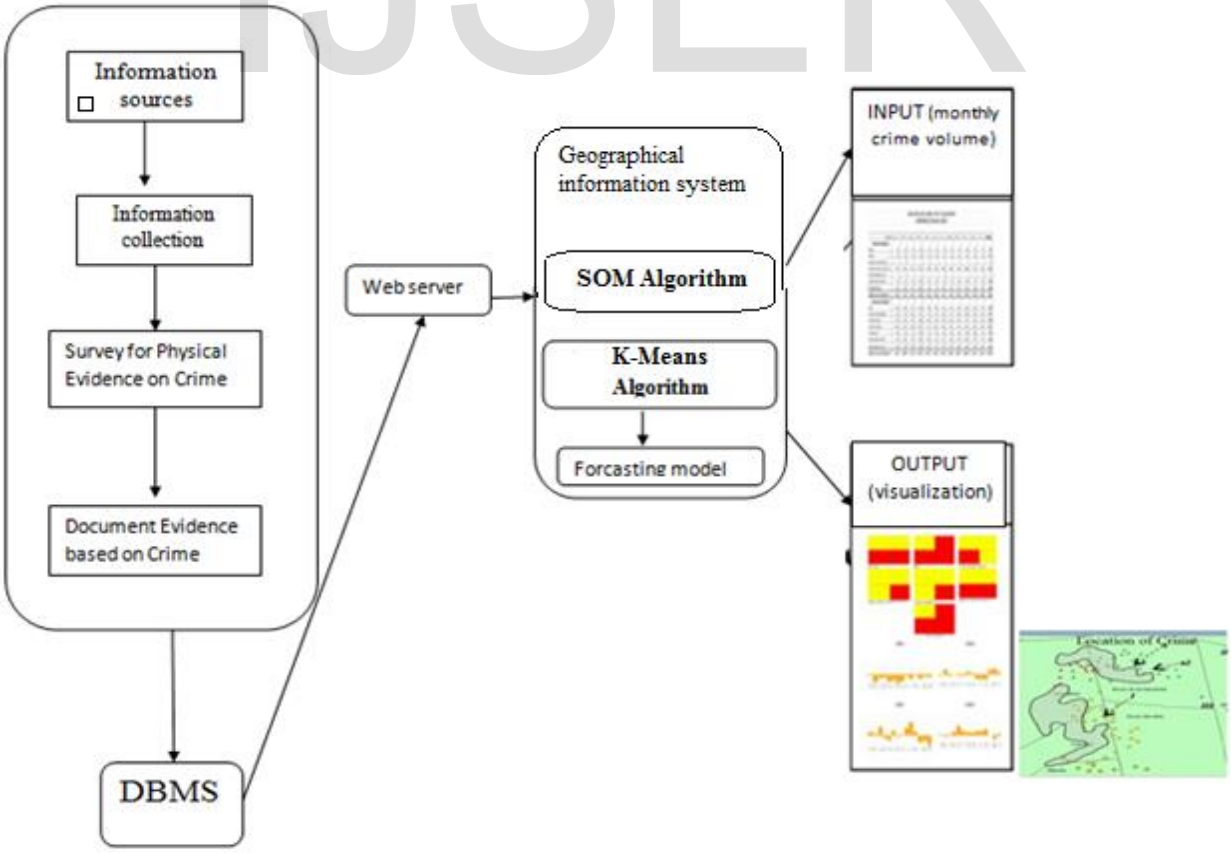

Figure 2: Proposed System Architecture

7. RESULTS

The result or the output of this research will be examined here. From the program screenshots. The result for selected tasks in comparing the K-means algorithm and our Hybrid algorithm are based on software quality assurance measurement called METRICS. The performance of the system is to actualized from the number of occurrence of the particular crime in that location is as follows

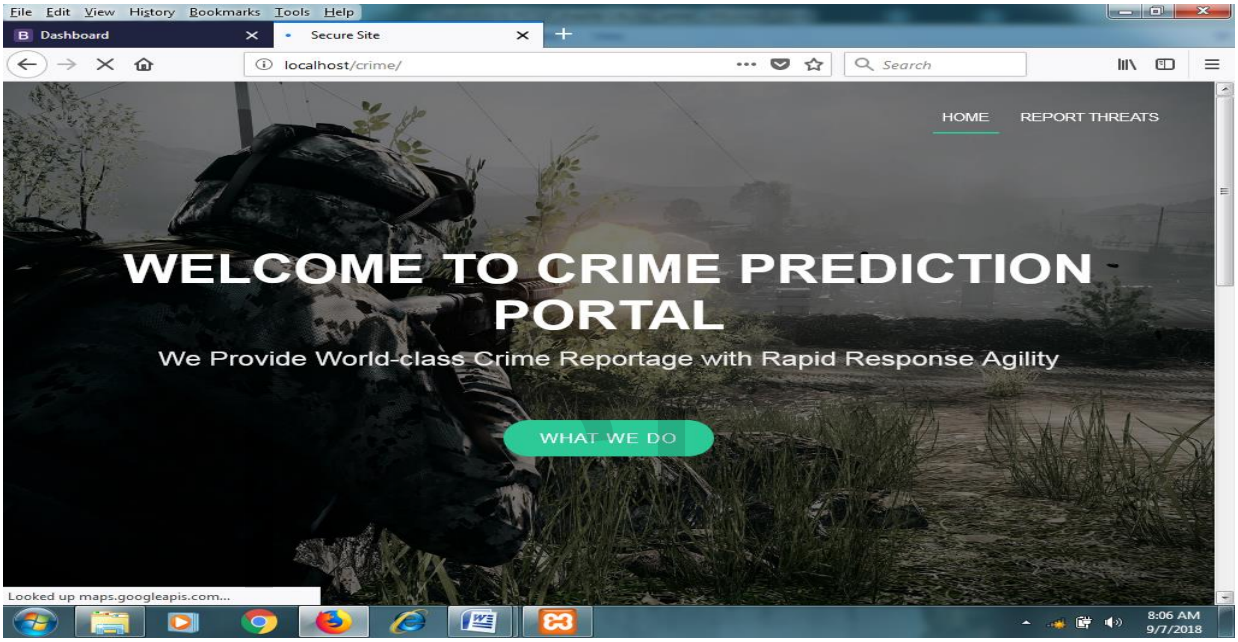

Figure 3: System Home page

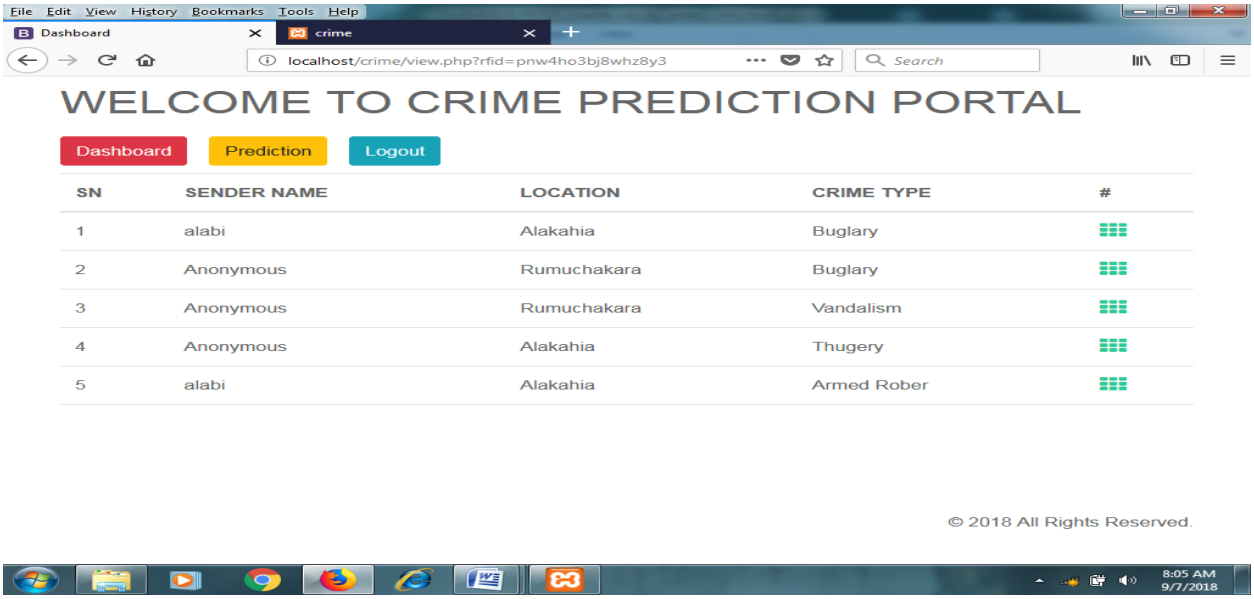

Figure 4: Data Page

Table 1: Aluu crime list

Crime Name	Occurrence	Percentage
Rape	10	8
Vardanlism	15	12
Cultism	30	24
Thugery	10	8
Burglary	40	32
Armed Robbery	20	16

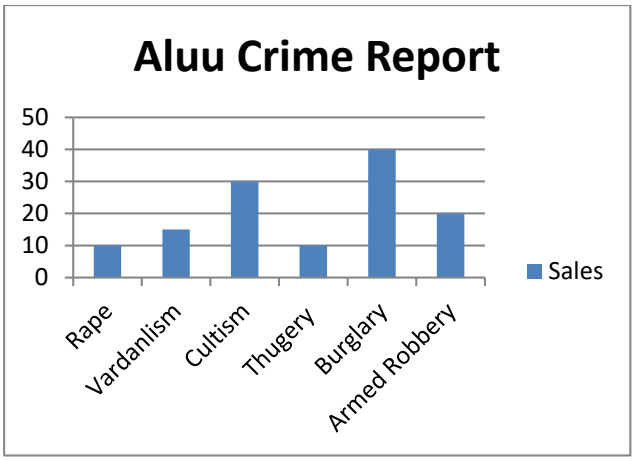

Figure 5: Aluu Crime Report

Table 2: Alakahia crime list

Crime Name	Occurrence	Percentage
Rape	22	12.57
Vardanlism	5	2.85
Cultism	30	17.14
Thugery	18	10.28
Burglary	80	45.71
Armed Robbery	20	11.42

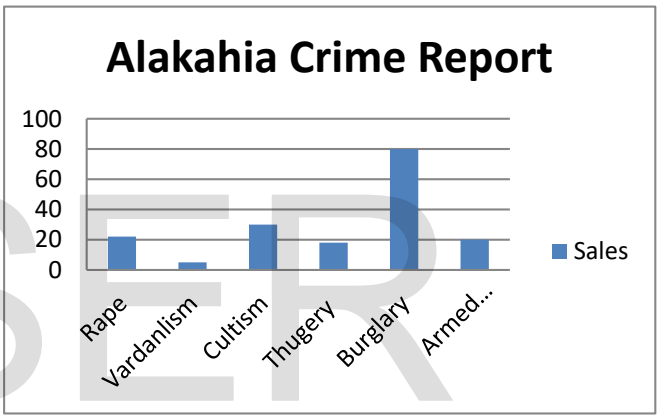

Figure 6: Alakahia Crime Report

Table 3: Rumuchakara Crime list

Crime Name	Occurrence	Percentage
Rape	22	12.02
Vardanlism	8	4.37
Cultism	35	19.12
Thugery	28	15.30
Burglary	70	38.25
Armed Robbery	20	10.92

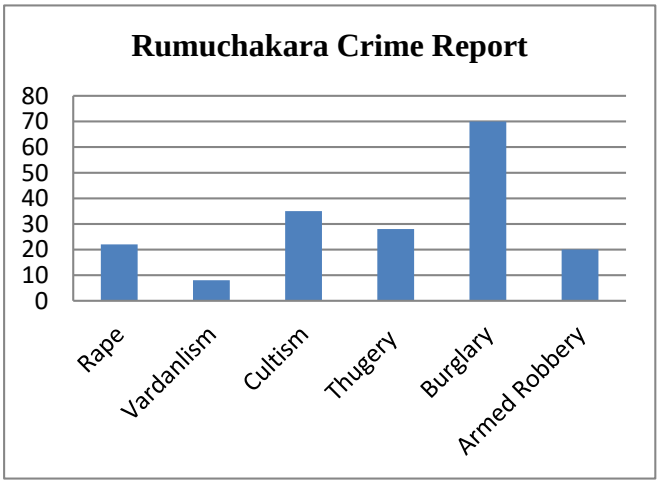

Figure 7: Rumuchakara Crime Report

Figure 4 illustrated that the proposed system data which was gotten from the location where the system is tested and validated as a useful model to predict crime occurrences.

8. Result Discussion

Figure 4 shows all the reported crime incident data have been captured by the system and saved.

The prediction page of the framework demonstrates the crime correspondents data, guide of that specific area where the crime occurred which makes it less demanding to find the area for legitimate examination. Figure 5 to 7 demonstrates the examination and spatial appropriation of crime episode information focuses on the guide of the considered areas with the assistance of ArcGIS.

From the result in the table 1 to table 3 and the display figure 5 to figure 7 (Graph), it was indicated the most occurrences crime committed in the three location unstudied which shows that Burglary crime is the most committed crime which needed to be addressed while and how it could be reduced to the minimum and also the way to prevent the occurrence of such crime.

9. Conclusion

This research reached to a good result when using the hybrid algorithm to predict the crime that occurs in a particular area in a single dataset. ANN provides the system with learning ability in such a way that the system can learn from both past and present event using SOM and K-means to give optimal result on large dataset. GIS provide the system with a digital representation that enables the user to map crimes locations analytically and descriptively from their different geographical location and visualize data in many ways that reveal relationships. GIS technology also has the ability to separate this information in layers,

This thesis has presented a methodology with the aim of developing a hybrid model approach. The methodologies are applied to real data on burglary incidence distribution in the study region. Relevant principles of statistics, GIS, clustering algorithm and ANN are utilized for the analysis of observed data. The statistical technique, simulation analysis described in this thesis was used to identify potentially significant predictive variables among characteristics of burgled households and the level of their contribution in the performance of the model and predicting of future crime in the study region. The model involves historical data (crime), population and census data such as Resident Population, Occupation, Qualifications, Socio-economic status, Household composition and Household spaces. The datasets were prepared using statistical and GIS technologies.

10. Contribution to Knowledge

1 The system combines K-means and SOM algorithm to effectively cluster high dimensional dataset to lower dimensional spaces in order to generate accurate result and better predictive value with less number of classification and quantization errors.

2 Users can easily get feedback responses to a report on instant crime occurrence within the location and can actually locate the current location of the committed crime.

11. Recommendations

The research is recommended to organizations both private and public (e.g. surveillance, Universities, E-government system etc.) that are looking for a better and faster way to predict crime and way to minimize the occurrences of crime.

References

Adegoke, N. (2014). Kidnapping, Security Challenges and Socio-economic Implications to the Niger Delta Region of Nigeria. Central Point Journal: A Journal of Intellectual Scientific and Cultural Interest. 16 (2) 205-216.

Anna L. Buczak , Christopher M. Gifford. T. (2010). Fuzzy Association Rule Mining for Community Crime Pattern Discovery

Chinwokwu, E.C (2013). The challenges of Conflict Management in a Democratic Society: An Overview of Insecurity in Nigeria. American International Journal of Social Science, 2(3). 93 – 106.

Chinwokwu J.J., (2013).Predicting the geo-temporal variations of crime and disorder. International Journal of Forecasting, 623-634.

Dawei Wang et. al. 2013. ―Understanding the spatial distribution of crime based on its related variables using geospatial discriminative

Durkheim E. B. (1958). The Regional Crime Analysis Program (RECAP): A Framework for to Catch Criminals.2848-2853.

Ekhomu R. (2010). Forecasting The Future of Predictive Crime Mapping. Crime Prevention Studies, 13, 29-57.

Goswami D.N., Chaturvedi Anshu., Raghuvanshi C.S. 2010. ―An Algorithm for Frequent Pattern Mining Based On Apriori‖, International Journal on Computer Science and Engineering. 2(4) 942-947.

Javad Hosseinkhani(2014). Detecting Suspicion Information

on the Web Using Crime Data Mining Techniques‖, International Journal of Advanced Computer Science and Information Technology. 3(1)32-41.

Madhusmita Mishra and H.S. Behera (2012). Kohonen Self Organizing Map with Modified K-means clustering for High Dimensional Data Set. International Journal of Applied Information Systems (IJAIS) 34-39.

Majeed Salar Mustafa (2010). Cluster Detection and Analysis with Geo-Spatial Datasets Using a Hybrid Statistical and Neural Networks Hierarchical Approach. Thesis of PhD, University of Glamorgan.

Malathi, A. Santhosh B., S. (2013). An Intelligent Analysis of Crime Data for Law Enforcement Using Data Mining, International Journal of Data Engineering.1 (5), 77-83.

Nwaze,C. (2011). Corruption in Nigeria; Terrorism in Nigeria. University of Ibadan Press, Ibadan

Revatthy K., J. and Satheesh K. (2012).Survey Of Data Mining Techniques On Crime Data Analysis‖, International Journal of Data Mining Techniques and Applications, 1(2), 117-120.

Shyam V. N. (2014). Crime Pattern Detection Using Data Mining and Geographical Information System, 1-4.

Yamuna, N. SudhaBhuvaneswari.(2012). — Datamining Techniques to Analyze and Predict Crimes‖, The International Journal of Engineering And Science, 1(2) 243-247.

Zakaria S. Z., and Ayman A. M. (2014). Crime Data Analysis Using Data Mining Techniques to Improve Crimes Prevention.International Journal of Computers. 8. 39-45.