COMBATING CRIME AND TERRORISM USING DATA MINING TECHNIQUES

Raphael Obi Okonkwo[1] and Francis O. Enem[2]
[1]Department of Computer Science, Nnamdi Azikiwe University, Awka
[2]Management Information System Unit, University of Nigeria, Nsukka,

ABSTRACT

In the wake of 9/11 terrorist attack on America; many countries became more particular about their national security. In the light of this, various means have be adopted to help the law enforcement agencies to identify terrorist and to counter-terrorism. One of such measure is the use of computer technology and computer analysis for effective analysis of criminal activities. Data mining can be used by law enforcement agencies to analyze information by applying the various data mining techniques. In this paper, we analyzed how data mining techniques can be adopted by law enforcement agencies in tracking the activities of terrorist and their criminal activities; also the paper examines the limitation of data mining in fighting crime in Nigeria.

Keywords: Dataset, terrorism, data mining, crime, knowledge discovery database, data warehouse

0 INTRODUCTION

After the terrorists attack on world trade centre popularly known as 9/11, the American government has devised various means by which the activities of terrorists and other crime perpetrators would be monitored before unleashing their evil plots on the general public. Other nations through out the world having seen the devastating effect of 9/11 in an attempt not to have a repeat of such in their soil also adopted and welcomed means which geared towards preventing terrorism in their countries and indeed world over.

In Nigeria today, many terrorist networks have sprouted in many parts of the country, MEND, Boko Haram and MASSOB to mention just but a few, have been unleashing terror to the Nigerian public. The government is extremely concern in curtailing the activities of these extremist as well as other crime perpetrators ranging from mobile phone theft, cult activities, drug trafficking, gang related offences, fraud, kidnapping for ransom, organized crime and others.

Today world over, with the sophistication in technology so is crime, organized crime and terrorism utilize these sophistications in technology to carry out their nefarious activities world wide. Terrorist seldom operate in a vacuum but interact with one another to carry out their nefarious activities. Worrisome though is the fact that in most terror networks, organized crime etc, there seems to be a degree in associative relationship among members of the network. From message gathering, information leaking, down to the execution and in most cases one network can be link to another.

Detecting criminals and solving crime is not an easy task at all and have been a prerogative of the law enforcement

agencies, the initiative of crime fighting is solely the responsibility of law enforcement agencies concern, however with the increasing sophistication in technology, computer system are now being used in tracking criminals and their activities and computer data analysts have started helping the law enforcement officers and detectives to speed up the process of solving crimes.

2.0 RELATED WORK

Data mining is relatively new technology in the country though it has been in use in most of the advance country, According to Cate (2008) Data mining is a promising tool in the fight against terrorism. It already plays a number of important roles in counter terrorism including locating known suspects, identifying and tracking suspicious financial and other transactions, and facilitating background checks. Seifert (2007) positions data mining as a key feature in the fight against terrorism and crime Data mining is emerging as one of the key features of many homeland security initiatives. Often used as a means for detecting fraud, assessing risk, and product retailing, data mining involves the use of data analysis tools to discover previously unknown, valid patterns and relationships in large datasets.

Thuraisingham (2008) said data mining can be used to detect unusual patterns, terrorist activities and fraudulent behavior. DeRosa (2004) Data-mining and automated data-analysis techniques are powerful tools for intelligence and law enforcement officials fighting terrorism. Data mining has been increasing supporting the law enforcement agencies in the fight against terrorism. This lead the United States of American's government in the wake of 9/11 to establish the Total information Awareness later renamed the terrorist Information Awareness according to Seifert (2007), though both program has since been discontinued noted Seifert.

DeRosa (2004) pointed out that automated data-analysis techniques can be useful tools for counterterrorism in a number of ways. One initial benefit of the data-analysis process is to assist in the important task of accurate identification. Technologies that use large collections of identity information can help resolve whether two records represent the same or different people. Accurate identification not only is critical for determining whether a person is of interest for a terrorism-related investigation.

Aiding data mining in countering terrorism is however due to the increase in speed of computers processing power, the declining cost of technology as well. Wiess et al. (2010) said that data mining developed as a new discipline for several reasons. First, the amount of data available for mining grew at a tremendous pace as computing technology became widely deployed. Specifically, high speed networks allowed enormous amount of data to be transferred and rapidly decreasing disk costs permitted this data to be stored cost-effective.

2.1 Data Mining Process

The process of data mining is simply the collection of data into a single repository where data mining algorithms are applied for knowledge discovery and pattern recognition. Chuck Ballard et al. (1999) highlights that a data warehouse provides the base for the powerful data analysis techniques that are available today such as data mining and multidimensional analysis, as well as the more traditional query and reporting. Fayyad et al. (1996) pointed out that data warehousing helps set the stage for KDD. They also noted that the most data-mining algorithms such as statistics, pattern recognition, and machine learning assume data are in the main memory. However, Kurt (2010) opines that most of the data mining can exist with a data warehouse. Figure 1 shows the simple process of data mining, data from various sources are gathered together in a repository commonly known as data warehouse before data mining techniques are applied for pattern evaluation, recognition and analysis.

The purpose of this paper is not to give details of data warehousing, but

mention have to be made about how it does set the stage for data mining.

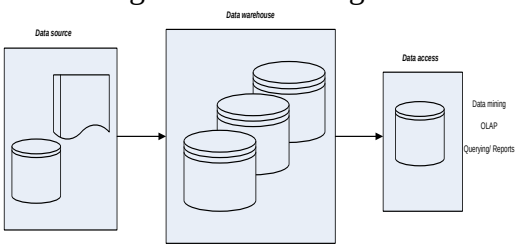

Figure1. The concept of data mining and data warehousing

In view of the above, we have to x-ray the definition of data mining and Knowledge discovery database according to Fayyad et al. (1996), Fayaad et al. defined data mining as a process in the knowledge discovery database (KDD) which is a nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data. Their views are diagrammatized in Figure 2 and show data mining as a continuous process; from a large dataset, valid data are selected, processed and transformed into a more useful dataset before data mining techniques are applied for valid patterns. Dissecting further the definition and concept of data mining according to Fayyad et al. (1996), the following are evident:

Datasets: Data are set of facts (database) and pattern describes a subset of the dataset.

Model: Designates extracting and fitting a model to the data

Process: The fact that KDD and data mining comprise many processes

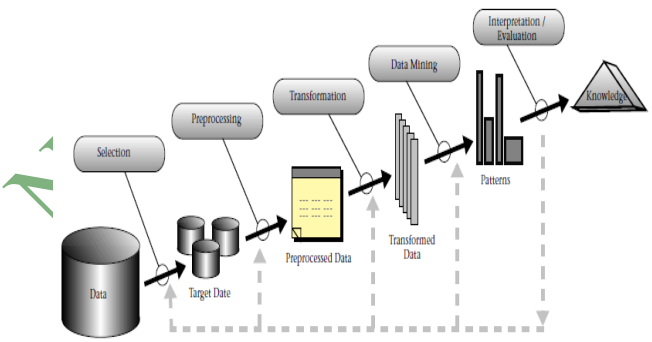

Figure 2. The process of data mining in the Knowledge Discovery Data-Base

Source: Usman Fayyad et al. (1996)

3.0 COMBATING CRIME USING DATA MINING TECHNIQUES

Using data mining, various techniques and algorithms are available to analyze and scrutinize data. However, depending on the situation, the technique to be used solely depends upon the circumstance. Also one or more data mining techniques could be used if one is inadequate. Data mining applications also uses a variety of parameters to examine the data.

In using data mining techniques in crime detection, Thuraisingham (2010) said that the methods used are top down reasoning where we start with a hypothesis and then determine whether the hypothesis is true or bottom up reasoning where we start with examples and then come up with a hypothesis. In the case of determining terrorist, first method adopted is usually the top down approach by asking or stating the hypothesis; who carried out the attack/crime or it is certain person that carried out the attack.

The next step is to start investigation as to the likely causes of the attack and the individuals who might have responsible attack**.** We have stated that crime investi-gation remains the prerogative of the law enforcement agencies concern, but computer and computer analysis can be useful in solving detecting.

3.1 Classification

According to Wies *et al*. (2010), classification task is the most commonly encountered data mining task. Classification in a broad sense is a data mining technique that produces the characteristics to which a population is divided based on the characteristic. The idea is to define the criteria use for the segmentation of a population, once this is done, individuals and events can then fall into one or more groups naturally. Thuraisingham (2008), said that classification divides the population (dataset) based on some predefined condition.

When classification is used, existing dataset can easily be understood and it will in no doubt help to predict how new individual or events will behave based on the classification criteria. Two crown (2010), states that data mining creates classification models by examining already classified data (cases) and inductively finding a predictive pattern. These existing cases may come from an historical database, such as people who have already undergone a particular medical treatment or moved to a new long-distance service. They may come from an experiment in which a sample of the entire database is tested in the real world and the results used to create a classifier

4.0 DISCUSSIONS

Classification assumes that we have some certain idea of the individuals (suspects in case of crime) based on the predefined criteria. Applying classification algorithm is such that it divides a given population (dataset) based on the criteria formed. For example, assuming that a kidnap has been reported to law enforcement agency, they may try to form the idea of the kidnapper(s), say 3 males, between 28 and 32 of age, Christian, speaks Ibo fluently, between 5inch and 7inch tall. The classification has been made and it becomes imperative to place all males meeting the above criteria under proper observation.

The algorithm to be adopted will be such that the population will be divided into two distinct parts, male and female.

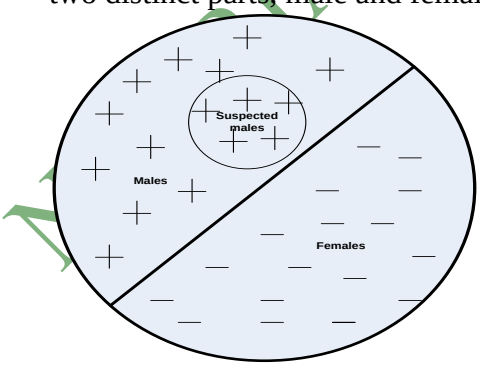

Figure 3. Classification algorithms depict the segmentation of a dataset based on some data criteria

The algorithm will also segment the males according to their suspected criteria. Figure 3 shows a typical segmentation scenario where males are segmented from females. Within the male dataset, it could be further segmented based on our criteria. The next step is to scan the historical dataset (data warehouse) for related matches. If the search does not yield positive result, we could narrow the search by reclassifying the classified dataset. Assuming that we were not able to find a class, the new criteria will then be used to update out historical database (data warehouse), if another related crime occurs, it becomes easier to form a match.

Figure 4 depicts a classification algorithm, in it, the data warehouse classified based on the criteria for suspected individuals. In our example above, the dataset is classified into males and females, obviously, this classification is not adequate, therefore the process continues until we have a dataset that can be matched against the criteria. Our classified dataset is now matched against the criteria, if we have a related match, it is then place under surveillance otherwise the process continues until there are no more matches.

If after analyzing a crime with classification technique, it is possible that a number of suspected individuals will come up. In order to narrow down to the suspects; their activities, links, associations, and relations could further be analyzed using the Link Analysis data mining technique.

4.1 Link Analysis

Link Analysis (LA) is another data mining technique that is useful in detecting valid and useful patterns. The theoretical framework of Link Analysis (LA) is based on the fact that events are linked to one another and are hence mutually exclusive. Link Analysis framework is that if A is link to B and B in linked to C and C to D, then A could be linked to D. When any link scenario is visualize, it collapses into a form of graph, link analysis uses various graph

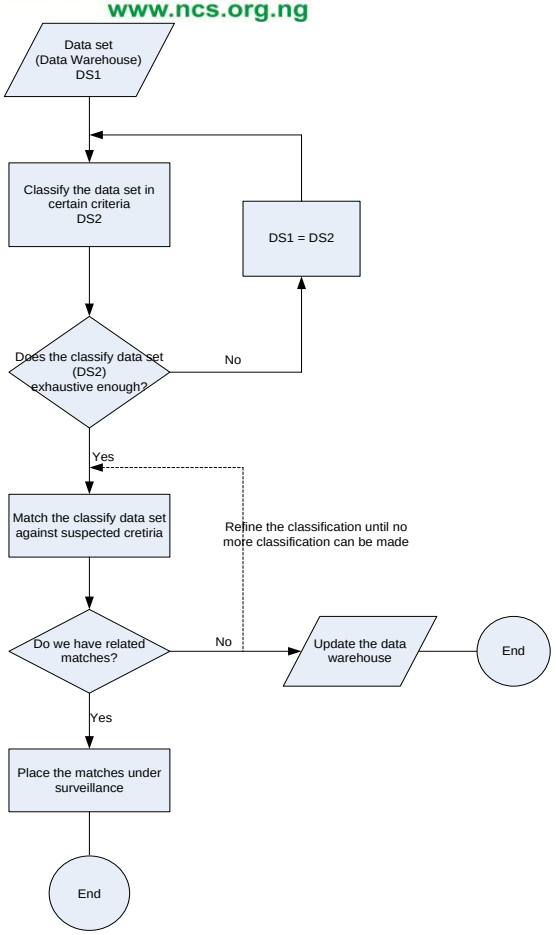

Figure 4. Classification flow for detecting crime and combating crime and terrorism

theoretic techniques. It is essentially about analyzing graphs

According to Memon and Qureshi (2005), the data mining technologies like link analysis (LA) can be employed by law enforcement investigators and intelligence analysts to help them to examine graphically the anomalies and inconsistencies; and connect networks of relationships, and contacts hidden in the data. LA is the first level by which networks of people, places, organizations, vehicles, bank accounts, telephone calls, email contacts, and other tangible entities can be discovered, linked, assembled, examined, detected, and analyzed.

With link analysis, law enforcement agency will have to mine data pertaining to any suspected individual. How do we relate a telephone call to another? Ho do we relate a bank transfer to another? How do we relate a travel document to another? One of the most basic problems is using computer and computer analysis to solve crime is Nigeria is that data are hardly kept. For instance buying pattern of individual Nigeria, when there are no historical data, data mining becomes pretty very much difficult.

4.4.1 Discussions

Using link analysis (LA) is purely based on graph. The ability to analyze graph gives the technique its strength. Once we will be able to analyze graphs properly, then we will be able to use this technique. Consider the example in figure 5, when the target dataset has been classified, we will have quite a number of individuals to put under surveillance. How then do we arrive at some suspected individuals? During the introduction, we have noted that terrorists and crime perpetrators seldom operate in a vacuum; that is there must be links to other individuals.

Link analysis can then be used to analyze the activities of individuals by forming a link of their activities. These links might be in form of telephone conversation, places visited, bank transactions etc as depicted in figure 6. The link analysis algorithm can then build links between suspected individuals. Consider the simple graph as depicted in figure 6 that represents the telephone conversation of some suspected individuals, B and C can be linked to Q, while B can be linked to P but the same can not be said of C to P

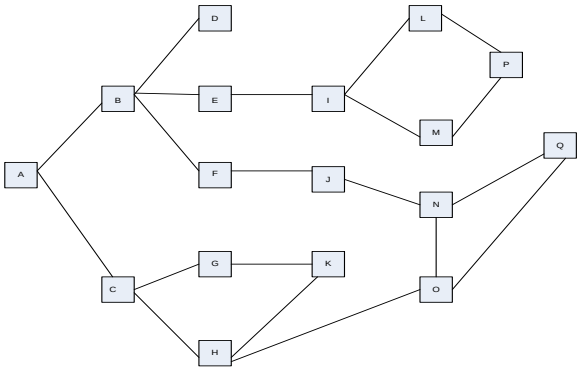

Figure 5. Link Analysis Graph

Link analysis can be so ambiguous and over crowded. The idea will be to

reduce the graph in manageable chuck so that vital information can be deduced from such graphs. Thuraisingham (2008) said that with link analysis one needs to reduce the graphs so that the analysis is manageable and not combinatorial explosive.

In Nigeria, the ownership of mobile telephone is being address with the registration of SIM cards both by the mobile operators and the regulatory body the Nigerian Communications Commission (NCC). Activities of suspected individuals could be monitored by law enforcement agencies via their activities. Telecom-munication providers on request provides the law enforcement agencies with telephone call manifest, with details such as the International Mobile Equipment Identity (IMEI), the location of the emanating and terminating calls via the Global Positioning System (GPS) and other pertinent details regarding calls, activities of the suspected crime perpetrators can be duly analyzed leading to possibilities of the attackers in case of crime.

Another problem with graph is that it might not necessary represents the entire activities of the suspected individuals and thus we would have partial information. Many other agencies could have activities (graph) about the individual and when you merge them, a clearer picture might be formed. The bank activities of the suspected individuals might provide a complementary link. We might want to know the deposit and withdrawal patterns of some suspected individuals. Figure 6 demonstrates a simple data warehousing; banking activities could be warehoused at a central position where transactions could be further analyzed.

Banks do furnish the Economic and Finance Crime Commission (EFCC) with relevant information on transaction beyond certain amount of money; however, an individual can circumvent this by dividing the amount of money into several parts and

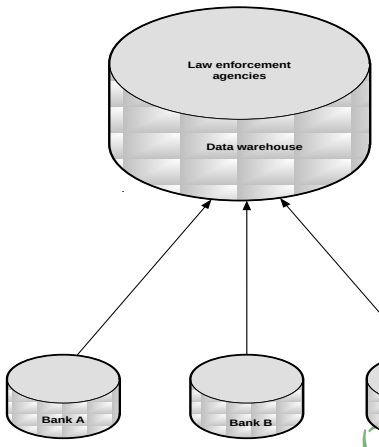

Figure 6. Shows the transactions of different banks to a central repository

deposits same into several banks as noted earlier. If we have a data warehouse where bank transactions are warehoused as shown in figure 6, transactions could be mapped and various patterns could be matched in terms of some biometric features, deposit and withdrawals.

4.2 Clustering

The goal of clustering is to group similar object into one cluster and dissimilar object into another cluster based on characteristics of data, also knows as segregation. Clustering can easily be used to automatically segregate individual into a defined group. Based on the characteristics that we have used in the segregation, certain occurrences of data can be further placed under detailed surveillance.

Clustering has been with us over time. People naturally cluster together based on some certain qualities, attributes and characteristics. People from the same country, religion, tribe, race etc cluster together. According to Wiess et al. (2010), the main reason for data clustering is that it allows us to build simpler more understandable models of the world, which can be acted upon more easily. It is imperative to understand that clustering differs from classification in the sense that classification is based upon a defined variable. Two crowns (2005), clustering is a way to segregate data into groups that are

not previously defined, whereas classification is a way to segment data by assigning it to groups that are already defined.

4.2.1 Discussions

Clustering is another data mining technique that can be used to detect crime and terrorism. Clustering and classification are almost the same however, in classification; there is basic parameter but clustering does not require any parameter. Clustering techniques and algorithm are based on real-life model that individual with certain qualities must cluster together.

Most crime and crime related activities can be clustered in Nigeria, the activities of the kidnappers can be clustered to south east though sometimes scattered all over the country, MEND activities can be found in Niger delta until recently, the activities of Boko-Haram can also be clustered to the Northern part of the country. Majority of the bank robbery and other incidents of robbery are more likely in the Lagos area. Figure 7 depicts a sample of cluster cases of kidnapping is some selected areas.

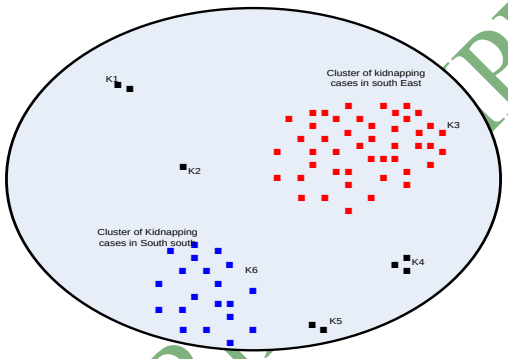

Figure 7. Cluster cases and anomalies

Clustering also assumes that in a crime dominated area, individuals with some certain crime specialties will cluster together. For example, individuals who specializes in kidnapping tends to cluster together while those that does car snatching will cluster. In terms of crime, often times it involves a syndicate with high degree of interconnection amongst them. Jonas J and Harper J (2006) said that the terrorists that attacked the world trade center at one point clustered together. Clustering algorithms will then be to identify given clusters and their areas of operation, anytime a crime is reported, law enforcement agencies can look for related clusters, and then examined them for clues.

Figure 8 shows a typical clustering algorithm, the dataset is clustered into various clusters, the crime suspects in then place under each cluster for matches. If there are matches, they are then place under surveillance otherwise, if there are no more clusters to be formed, a new cluster is then formed and the dataset updated.

Often times in clustering anomalies do occur in what is referred to anomaly detection. This is a case of a prevalent event that happens within a particular place and an event or activity happening elsewhere. For example Boko-Harem is noted to operate in the north mainly, what if all of a sudden; they attack in the Western or Eastern part of the country. Kidnapping for ransom is mainly a feature in the South East, if we have a case of kidnapping for ransom in Abuja metropolis, then an anomaly has happen as depicted in Figure 7. When an anomaly occurs, it also shows the starting of another cluster. K3 and K6 are the normal clusters of kidnapping cases; K1, K2, K4 and K5 are all anomalies of kidnapping.

4.3 Nearest Neighbour (K-NN)

One of the oldest techniques of data mining is the nearest neighbour. Alex B et al (2010), The nearest neighbour prediction technique of data mining is quite very simple in that is assumes that in order to predict what the predication value in one record, look for records with similar predictor values in the historical database and use the predication value for the record that is nearest. The principle of NN is just about solving a problem by looking at past problems and finding a similar one that matches the current problem. Two crowns (2010), when trying to solve new problems,

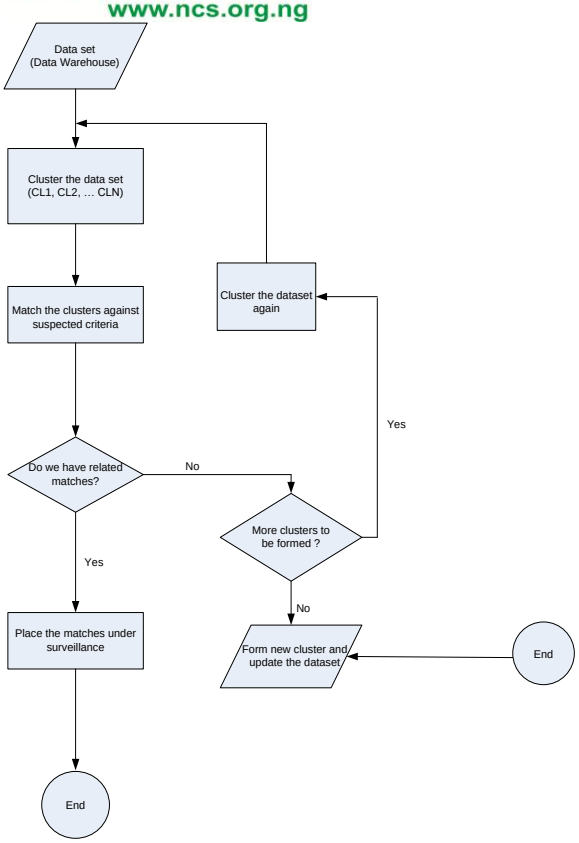

Figure 8. The use of cluster algorithm to identify cases of crime

people often look at solutions to similar problems that they have previously solved.

Two crowns (2010), K-nearest neighbour (k-NN) is a classification technique that uses a version of this same method. It decides in which class to place a new case by examining some number - the "k" in k-nearest neighbor - of the most similar cases or neighbors (Figure 9). It counts the number of cases for each class, and assigns the new case to the same class to which most of its neighbors belong.

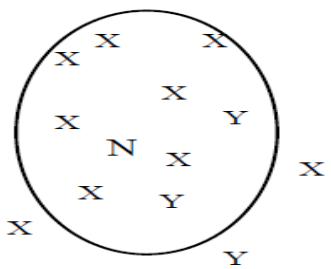

Figure 9. The k-Nearest Neighbour (K-NN), assigns the new case N to X as X outnumbers Y

4.3.1 Discussions

The Nearest Neighbour data mining technique and algorithm has been used for ages. By intuition, when one sees a good person that becomes close to suspected criminals, he or she will cringe. This technique works well when we have identified a group, the chances are that any other person among the group will likely be associated with them. Take for instance a group of notorious armed bandits, when any person is seen near each member of the gang, our perception is that he or she is one. Another example is that in a place notorious for its gang activities, any person sees loitering around there is automatically assumed to be one of them.

Consider figure 9, the k-nearest neighbour counts the occurrences of cases "k" and assigns the new case to the highest number in the group. Another example, in a place noted for its cult activities, some of the people residing in the place might not likely be cultist but because the cultist out numbered the non-cultist, any time a new person is seen in the place, it is assumed that the person is also a cultist.

Also the k-nearest neighbour assume that when a crime is committed, in order to find the perpetrator, look for a crime of similar pattern and try to find a pattern. Continue with this until no more similar cases could be matched with the new case as depicted in Figure 10.

4.4 Limitations of data mining

Data mining is an emerging tech-nology and has advanced a great deal. No doubt data mining have been accepted and are being applied in several human endeavours. But the question that has been raised all over remains; how far can data mining go? Can it be really applied in detecting/or preventing terrorist activities. Data mining can be useful in detecting and solving crime related issue, the main limitation of data mining is personnel.

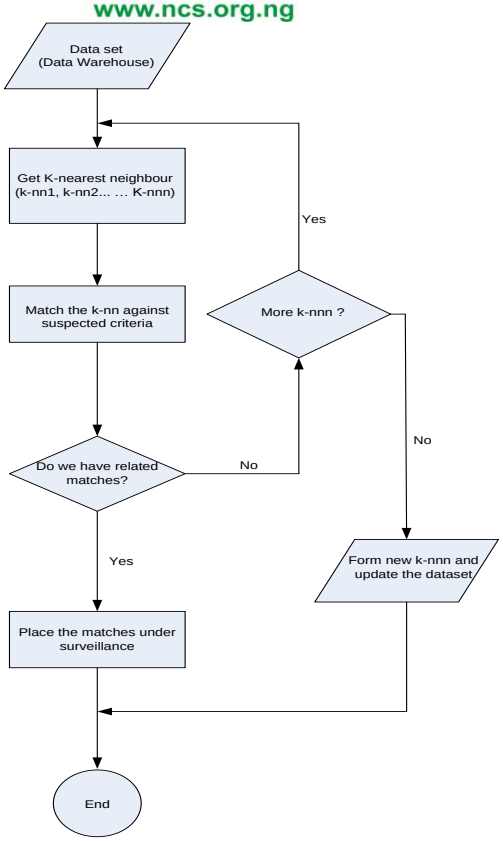

Figure 10. The use of K-NN algorithm to identify cases of crime

Two Crown (1999) While data mining products can be very powerful tools, they are not self sufficient applications. To be successful, data mining requires skilled technical and analytical specialists who can structure the analysis and interpret the output that is created. Consequently, the limitations of data mining are primarily data or personnel related, rather than technology-related

Another limitation of data mining is that we do not have already at hand data on individuals. The individual data and corporate data that are required to be used to track the potential terrorists and crime perpetrators are certainly not in existence with the Nigeria Law enforcement agencies. Though Patches of data might exist here and there but we do not have a comprehensive historical database where we can draw analysis and inference from.

Data mining can be very sensitive in the quality of data input. Often times, data might be incorrect, error prone and above all a large quantity of dataset need to be dealt with in order to ferret useful data that can assist the detectives and law enforcement agencies.

Data mining awareness simply does not exist in the country and government has no clear blue print on the use of technology to resolve crime. Cate F(2008), pointed out that in United States of American, that government data mining in general is widespread and expanding. According to him, a 2004 report by the Government Accountability Office found 42 federal departments to have started data mining and planning to do so. Not so can be said about Nigeria as where most of the federal ministries do not have adequate computer infrastructure.

Finally, we might not remain adamant of the one problem that data mining is going to take away from us. This is our privacy! Data mining having been proven to be useful will in no doubt erode our privacy.

4.5 Recommendations

Government agencies should setup data mining agencies within the law enforcement agencies where various data should be consolidated and mined. Data on individuals such as voter's registration, national identity, population census information, etc should be linked together to profile the identity of an individual. Phone numbers, bank accounts and other related activities could easily be traced to any individual.

Companies and organization to cooperate with the law enforcement agencies more by reporting cases of irregular transaction pattern to the government agencies on real-time basis. They do not have to wait until the law information agencies request them to provide information on some certain transaction.

To complement the above, a centralized data warehouse should be established by the police where data on each individual could be established. This will enable the mapping of real data to the data mining attributes. Data mining techniques

will not replace detectives or tell at once whom a terrorist is but a careful study and analysis can help identify ways by which crime could be reduce.

Law enforcement to be better equipped in the area of computer technology and computer analysis in order to crack some of the grueling data mining techniques and be able to link cases to suspected individuals. Adequate training and retraining should be provided for the staff of the law enforcement agencies.

Government should also be encourage to invest more into communication technology as well as information technology by creating adequate enabling and necessary field for investment and development. Though much has been done recently on this regard, however other supporting factors such as adequate electricity supply, good road network to mention but a few.

Government and corporate bodies to set up data warehousing and mining institute where academics, professionals, law enforcement agencies will interact to develop better data mining algorithms capable of detecting terrorist and crime activities easily.

5.0 CONCLUSIONS

Combating terrorism, terror attacks and other criminal activities requires the adequate attention of the government. Law enforcement agencies should better be equipped in this information age on how to use computer and computer analysis to track the nefarious activities of the hoodlums. Corporate bodies especially banks should also play vital role in the fight against terrorisms.

Crime pattern analysis and detection can only help the law enforcement agencies and are not intended to replace them. Also, data mining techniques are not going to crop up and say that the bad guy is this or that rather it will help the detectives and law enforcement agencies in crime fighting.

Data mining technique used in crime detection depends solely on the situation at hand. Most cases require the combinations of two or more techniques used alongside. For instance classification and link analysis techniques can be used to complement each other.

Data mining techniques and algorithms can be applied in most cases to solve the issue of crime by identifying the activities of these criminals and tracking them down. A well implemented data mining algorithm will definitively go a long way in this quest to minimize terrorism.

Data mining in not all to counter-terrorism as there are various drawbacks which included the issue of skilled manpower, inadequate investment in telecommunication and IT infrastructure, inadequate data mining policies and above all legal issues that characterize unwanted tracking of innocent citizens.

6.0 REFERENCES

Alex Berson, Stephen Smith and Kurt Thearling (2010). "An Overview of Data Mining Techniques" retrieved from the web 14-12-2010

Carlile of Berriew Q.C. (2007). "Data mining: The new weapon in the war on terrorism" retrived from the Internet on 28-02-2011
http://fcw.com/articles/2006/05/29/data-mining-the-new-weapon-in-the-war-on-terrorism.aspx

Cate H. Fred (2008). "Legal Standards for Data Mining" retrieved from the internet on 12-03-2011
http://www.hunton.com/files/tbl_s47Details/FileUpload265/1250/Cate_Fourth_Amendment.pdf

Chuck Ballard, Dirk Herreman, Don Schau, Rhonda Bell, Eunsaeng Kim and Ann Valencic (1998). "Data Modeling Techniques for Data Warehousing" IBM Corporation.

Clifton Christopher (2011). "Encyclopedia Britannica: data mining", Retrieved from the web on 20-01-2011
http://www.britannica.com/EBchecked/topic/1056150/data-mining

DeRosa Mary (2004). "Data mining and Data Analysis for counter terrorism", CSIS report-2004

Wiess M. Gary and Davison D. Brain (2010). "Data Mining", Retrieved from the Internet on 18-03-2011
http://storm.cis.fordham.edu/~gweiss/papers/data-mining-chapter-2010.pdf

J.R. Global Security Resources (2002). "Terrorism", retrieved from the Internet on 18-03-2011 http://www.angelfire.com/ca7/Security/TERRORISM.html

Jonas, Jeff and Harper, Jim (2006). "Effective Counterterrorism and the Limited Role of Predictive Data Mining" retrieved from the web 12-02-2011
http://www.thebreakingnews.com/files/articles/datamining-cato-report.pdf

Kurt Thearling (2010). "An Introduction to Data Mining; Discovering hidden value in your data warehouse", retrieved from the web 14-12-2010
http://www.thearling.com/text/dmwhite/dmwhite.htm

Memon Nasulla and Abdul Qureshi (2005). "Investigative Data Mining and its Application in Counterterrorism", proceeding of the 5th international conference on Applied Information and Communication, PP (397-303)

Osmar R. Zaïane (1999). "An Introduction to Data Mining: Principles of Knowledge Discovery in Databases"

Seifert Jeffrey W. (2004)."Data mining report, CRS report for congress", order code RL31789.

Thuraisingham Bhavani (2010). "Data Mining for Counter-Terrorism" The Mitre Cooperation, Retrived from the Internet on 18-03-2011

Two Crows Corporation (1999). "Introduction to Data Mining and Knowledge Discovery", 3rd Edition, Potomac

Usman Fayyad, Gregory Piatetsky-Shapiro and Padhraic Smyth (1996). "From data mining to knowledge discovery in databases", American Association for Artificial Intelligence/The MIT press.