Analysis of Crime Pattern using Data Mining Techniques

Chikodili Helen Ugwuishiwu[1], Peter O. Ogbobe[2], Matthew Chukwuemeka Okoronkwo[3]
Computer Science Department, University of Nigeria, Nsukka, Enugu state, Nigeria[1, 3]
National Board for Technology Incubation, Abuja, Nigeria[2]

Abstract—The advancement in Information Technology permits high volume of data to be generated in databases of institutions, organizations, government, including Law Enforcement Agencies (LEAs). Technologies have also been developed to store and manipulate these data to enhance decision making. Crime remains a severe threat to humanity. Criminals currently, exploit highly sophisticated technologies to perform criminal activities. To effectively combat crime, LEAs must be adequately equipped with technological tools such as data mining technology to enable useful discoveries from databases. To achieve this, a Real-time Integrated Crime Information System (RICIS) was developed and mobile phones were used by informants (general public) to capture information about crimes being committed within Southern-East, Nigeria. Each crime information captured is being sent to the LEA responsible for the crime type and the information is stored in the agency database for data analysis. Thus, this study uses data mining algorithms to analyze crime trends and patterns in Southern-Eastern part of Nigeria between 2012 and 2013. The algorithms adopted were Classification and Rule Induction. The data set of 973 were collected from Eleme Police station, PortHarcourt (2012) and Nsukka Police station (2013). The analysis enables identifications of some trends of crimes and criminal activities from various LEAs databases, enhancing crime control and public safety.

Keywords—Information technology; law enforcement agency; data mining; crime; classification and rule induction

I. INTRODUCTION

Technology advancements have made the world a better place including access and manipulation of huge volume of dataset in virtually all fields of life [1]. Criminology is an important area for applying data analysis and it is a practice aimed at discovering crime characteristics [2]. Crime is of immense concern in our world today. Crime ranges from simple violation of civic duties (e.g., illegal parking) to internationally organized crimes (e.g., the 9/11 attacks) [3]. Crime has negatively impact both developed and developing nations because of its emotional, economic and social disruptive tendencies. It threatens the quality of life, human rights and poses severe challenges to any society [4][5]. The motivation of this study is to understand the trend, pattern and the prevalence of each crime type in different parts of the South-East, Nigeria. This properly advises the Law Enforcers (LEs) on the most effective approach on crime management. Criminals today, use sophisticated technological tools [6] not just to commit crimes but also to avoid being detected. Unfortunately, most of the Nigeria LEAs are still using paper-based information systems to capture and store crime data, leading to delay in crime information flow and inefficient crime data analysis. The work aims to analyze crime data generated from the implemented RICIS model to assist the LEAs perform their operations.

Crime control and prevention are fundamental to the welfare, stability and development of any society. Though, Government and communities have been making effort to improve on security standard of the public (through establishing different LEAs and community policing), more effort should be done to equip the LEAs with current technologies (e.g. crime analytic tools) for effective crime management. Timely access to relevant information is also of utmost necessity in day-to-day business of LEA, especially in crime investigation and detection of criminals [7].

Data mining is a tool that enables efficient extraction of useful information and patterns from complex and large datasets [6][8][9][10]. It is primarily aimed at uncovering hidden relationships in data warehouse with the aid of artificial intelligence method [11][12]. Data mining has been applied in the areas of crime analysis and prediction to assist the LEs in crime decision making [13].

Crime analysis is a task that includes exploring and detecting crimes and their relationships with criminals to assist security personnel in planning the deployment of available resources for the prevention and suppression of criminal activities [14].

A crime analysis should be able to identify crime patterns quickly in an efficient manner [7]. Some areas of data mining applications include supermarket, hospitals, banks, insurance companies, airline, governments and many other fields of life [15]. Data mining has been used to solve some challenges such as detection and prevention of fraudulent activities in telecommunication services, criminal's activity, etc.

Some data mining algorithm for knowledge discovery in databases include Classification, Rule induction, Association rules, Clustering, Forecasting, and Visualization, etc. Most times, researchers combine data mining techniques to achieve more precise and accurate extractions. Some software packages (built in mining tools) [16] [17] [18] designed for implementation of these techniques for data analysis include: Waikato Environment for Knowledge Analysis (WEKA), KNIME, ORANGE, etc. PHP and other programming tools were used to implement the algorithms (classification and Rule induction) for data analysis. The research will help LEAs and as well the government in improved decision-making and public safety.

The remaining sections are organized as follows: Section 2 discusses literature review; Section 3 deals with the analysis, design and method; Section 4 is on implementation, analysis and results, and finally Section 5 discusses conclusion.

II. REVIEW OF LITERATURE

This section deals with the data mining basic concepts, the imperatives for crime analysis and review of some existing literatures in the related areas.

A. Data Mining and Imperatives for Crime Analysis

Three steps involved in data mining include: exploration, pattern identification and deployment [1]. Exploration means to clean and transform data into a new form, then important variables and nature of data are determined based on the problem; Pattern Identification means to identify and choose the patterns which make the best prediction. Deployment organizes the patterns for desired outcome [1]. According to [16], [19], [15], crime data analysis provides summary statistics, general and specific crime trends to LEs in timely manner to enable understanding on crime and criminal behaviour. This assists LEAs to be proactive in crime detection and prevention while managing their limited resources effectively.

B. Types of Crime Analysis and Data Mining Tasks

The crime analysis types include:

1) Tactical crime analysis generates information on where, when, and how crimes take place to assist officers and investigators in identifying and understanding specific and immediate crime problems.

2) Strategic analysis examines long term changes in crime, known as "crime trends". Administrative Crime Analysis provides summary statistics, and general trend information.

3) Criminal investigative Analysis involves profiling suspect and victims for investigators based on analysis of available information. One may analyze to find out the type of person committing a particular crime series.

4) Intelligence analysis focuses on organized crime, terrorism, and supporting specific investigations with information analysis and presentation.

5) Operations analysis examines how LEA is using its resources. It focuses on such topics as deployment, use of grant funds, budget issues, etc. [15].

Two stages of data mining Tasks include:

a) Data collection phase: The researcher collects the training dataset from a defined source.

b) Data Pre-processing Phase: Processing the collected data to get it suitable for analysis through data cleaning, integration, transformation, reduction and discretization [16].

C. Types of Data Mining Techniques

Classification deals with discovering and sorting crime data into groups or predefined classes such as type, location, time, etc. based on certain attributes/criteria discovered from databases [20][21]. Classification is the process of learning a function f that maps each attribute of a set $X=\{x1, x2, \ldots, xn\}$ to a predefined class label y *[21]*. The goal is to build a model to predict future outcomes. Data classification helps to predict how new individuals or events will behave based on the classification criteria [17]. Classification has been used to detect email spamming and find authors who send out unsolicited emails [11], [18].

Rule induction is a machine learning technique in which through observations, rules are extracted from a dataset. These rules may possibly denote a scientific model of a dataset or signify a pattern in the dataset. A rule is conveyed with "if-then statements". Rule-based algorithm takes training dataset as input and generates rules by dividing the table with cluster analysis.

Association rules are generating rules from crime dataset based on frequent occurrence of patterns [16]. Association Rule searches for relationship between data that exist together in a given record to uncover crucial information. In [18], association rule detected suspicious e-mails by identifying unusual and deceptive communication in e-mails.

Clustering techniques is used to automatically associate different objects (such as persons, organizations, vehicles) that are similar to one another and dissimilar to objects of another group in crime records. Clustering is based on finding relationships between different Crime and Criminal attributes having some previously unknown common characteristics [15]. In crime analysis, cluster analysis is used to identify areas with high concentration of a particular crime type. By identifying these crime "hot spots" i.e. where a similar crime has happened over a period of time helps to manage law enforcement resources more effectively [3].

Forecasting deals with discovering patterns and data that may lead to reasonable predictions. It estimates the future value based on a record's pattern. It deals with continuously valued outcome. Visualization enables miners to rapidly and efficiently locate vital information that is of interest within the data [17].

D. Review of Related Literature

Data mining technology has been adopted by organization improve on business strategies through performing different knowledge discoveries [1]. Authors in [11] reviewed data mining techniques and presented four case studies of their crime data mining project as follows: entity extraction for police narrative reports, detecting criminal identity deceptions, authorship analysis in cybercrime and criminal network analysis. According to [16], a model was proposed for crime and criminal data analyzes using clustering and association rules algorithm. The work tends to help LEs in discovering crime patterns and trends, making forecasts and as well identifying possible suspects. The intension was to assist the Libyan Government on strategic decisions to reduce the high increase on crime rate.

In [19], k-means clustering algorithm was used to perform data analysis to assist LEs in crime reduction. The goal was to extract useful information from crime dataset to enable LEs to identify and analyze crime patterns for effective crime control and prevention. WEKA and Microsoft Excel were used for analysis. Data mining and machine learning tools such as the

Additive Regression, Linear Regression were used to conduct a comparative study between the violent crime patterns from the Communities and Crime Unnormalized Dataset provided by the University of California-Irvine repository and actual crime statistical data for the state of Mississippi provided by neighborhoodscout.com. The researchers implemented the three selected tools using the same finite set of features on the Communities and Crime Dataset and found that the linear regression algorithm performed best in predicting violent crime patterns [17].

A concise review on Data mining techniques and its applications on over 100 different areas on crime analysis were presented in chronological order [18]. In [22], clustering algorithm was used for crime data analysis to detect crime patterns and speed up the process of solving crime issues. The aim was to formulate crime pattern detection as machine learning task and using data mining to support LEs in crime struggle. The modeling technique was able to identify the crime patterns from the crime data. Research on [3], presented a detailed overview of clustering technique, its types and some of its application areas.

In [2], data clustering method and artificial neural networks were adopted to unsupervisely classify crimes and crime pattern recognition respectively. This research finally proposed a method for crime matching processes. In [23], binary clustering and classification techniques were used to analyze criminal data and was aimed to potentially discover a criminal based on the witness/clue at the crime spot. An auto correlation model was used to validate criminals on the process. Researchers in [24] focused on property crime and explore the best approach to reliably predict the location, time, and/or likelihood of future residential burglary. The work adopted classification model to perform the crime forecasting.

Authors in [13] studied data mining technologies used in criminal investigation, highlight the approach adopted in crime data analytics and summarizes some challenges arising during the analysis process. A machine-learning classification algorithms, such as boosted decision tree and K-Nearest Neighbour (KNN) were adopted to create a model that will perform crime prediction on Vancouver dataset collected for a period of 15 years (2003-2018). It is recorded that prediction accuracy between 39% and 44% was achieved [25].

III. DESIGN AND METHODS

A. Methodology

Data collection and the research output adopted quantitative method. The 973 primary data were collected from Eleme police station PortHacourt (2012) and Nsukka police station (2013) in the South-East Nigeria. Subsequently, data for analysis will be collected from the RICIS developed. To get the desired outcome, the data collected passed through three stages of data mining: exploration, pattern identification and deployment. The data were pre-processed using different pre-processing techniques (cleaning, missing values and removing inconsistency). The pre-processed data were used to find different crimes, criminal trends and behaviour. The crimes and criminals were grouped into clusters according to their important attributes. The data mining algorithm adopted

were classification and Rule Induction. The Psuedocode of the two techniques is presented in section 3.3.

B. Conceptual Framework of the Study

The framework of the proposed research presents the summary of the research in a block diagram as shown in Fig. 1. This gives details of the processes of the mining from data collection to the final report (result). The work aims to analyze crime data generated from the implemented RICIS model shown in Fig. 2, to assist the LEAs perform their day-to-day operations, though, data currently used for analysis in this study was collected from two different Nigerian Police stations as mentioned in Section 3.1.

C. Analysis of the RICIS System

The RICIS is a crime reporting system developed to capture crime information from informant as shown in Fig. 2. The components of the RICIS as shown in Fig. 2 include an informant who reports crime, a mobile phone for crime report, an application for directing crime information to the appropriate agency, a middleware to enable data exchange among the agencies, individual agency's databases for their data autonomy, a centralized database for data analysis and backup, internet to enable communication and finally the bidirectional arrows indicating the direction of information flow.

This system enables LEA to collaborate by sharing resources such as one agency searching for the existence of a criminal's information in any of the participating agency or LEAs redirecting a crime information wrongly sent to any of them to the appropriate agency for instant responses. The system also enables LEAs substitute the traditional paper-based crime reports systems with an electronic system. The system provides instant response to crime and generate huge amount of crime data for data analysis. The reported crime provides data such as location codes, crime type, where, when, and how crimes happens (short crime details), information about suspect (if identified) and the victim. These information enables the law enforcement officers to offer a just-in-time response to crime scene and have quick access to the right data.

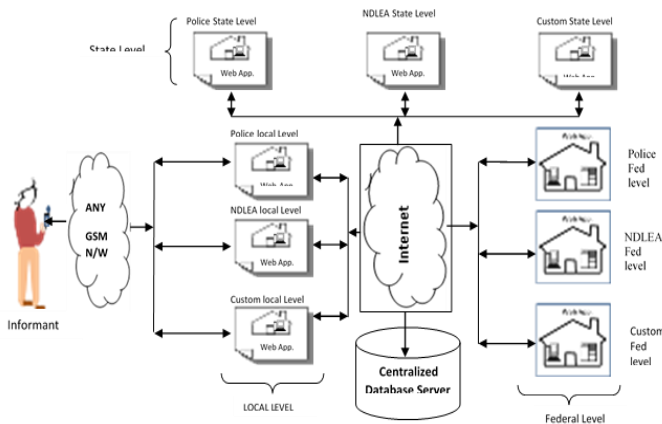Fig. 1. Conceptual Framework of the Data Mining Process.

Fig. 2.   The RICIS Model.

Mining the crime data on RICIC extract some crime pattern and relevant information such as: most prevalent crime in the country, state with highest number of criminals or crimes, age range of criminals with highest crime rate,

location with highest crime rate, the class of people that commit most crimes (literate, illiterate, married, single, jobless etc.), time of the day crime usually occur in a particular place and which week of a month or month of the year or year with highest number of crime, etc. The results from this system can provide statistics of crimes handled in an agency in a specified time (e.g. number of crime incident investigated and handled in a year, etc.). This will to a great extent help LEs to manage their limited resources effectively and improve on the security standard of the nation.

D. Data Mining Algorithm

This section shows the algorithms (Pseudocode) of the data mining techniques (Classification and Association rule) used. The pseudo code shows the sequential steps taken in the implementation of these techniques. The analysis done is a modified algorithm that combines both classification and association rule. Association rule is a learning algorithm based on rule or adaptation [26][27]. The algorithm must be provided with TrainingSet already been classified in order to generate a list of classification rules.

Rule Induction algorithm

1.0   Input data
    1.1   let the ClassificationRuleList be empty
2.0   Perform classification
    2.1   repeat
        2.1.1   let the ConditionalExpressionSet be empty and let the BestConditionExpression be nil
        2.1.2   Repeat
        2.1.2.1 Let the TrialConditionalExpressionSet be the set of conditional expressions, {x and y where x belongs to the ConditionalExpressionSet and y belongs to the SimpleConditionSet}.
        2.1.2.2 Remove all formulae in the TrialConditionalExpressionSet that are either in the ConditionalExpressionSet (i.e., the unspecialized ones) or null (e.g., big = y and big = n)
        2.1.2.3 for every expression, F, in the TrialConditionalExpressionSet
            2.1.2.3.1 If F is statistically significant and F is better than the BestConditionExpression by user-defined criteria when tested on the TrainingSet then replace the current value of the BestConditionExpression by F
        2.1.2.4 While the number of expressions in the TrialConditionalExpressionSet > user-defined maximum
        2.1.2.5 Remove the worst expression from the TrialConditionalExpressionSet
        2.1.2.6 Let the ConditionalExpressionSet be the TrialConditionalExpressionSet
      2.1.3 until the ConditionalExpressionSet is empty
      2.1.4 if the BestConditionExpression is not nil then let the TrainingSubset be the examples covered by the BestConditionExpression
      2.1.5 Remove from the TrainingSet the examples in the TrainingSubset
      2.1.6 Let the MostCommonClass be the most common class of examples in the TrainingSubset
      2.1.7 Append to the ClassificationRuleList the rule 'if ' the BestConditionExpression ' then the class is ' the MostCommonClass
    2.2 Until the TrainingSet is empty or the BestConditionExpression is nil

3.0 return the ClassificationRuleList [28].

The classification algorithm classifies data based on rules. The algorithm goes thus:

1.0 Input data
    1.1 Convert continuous values to categorical values.
    1.2 Set all classified to false (all classified is a variable that checks if all the data values are totally classified or not)
    1.3 Set n = 0
    1.4 Initialize classified dataset to empty
2.0 Repeat

    2.1 pick the value of an attribute
    2.2 generate rule for the combination (inserting value into a classified dataset)
    2.3 classify the value of the attribute
    2.4 set n=n+1
    2.5 if n = total number of data in raw dataset then set all classified to true, go to 3.0
    2.6 else repeat from step 2.1 until all classified = true

3.0 return classified data.

In the combined algorithm, there are key things to note:

1) Attributes – these are the field names.

2) Instances – these are the rows(records) in a dataset.

3) Class – this is the value of attribute (data stored in a field).

4) Subclass – this is the node of a class.

The task here is to loop through all the instances, pick the classes and classify them according to their respective attributes. In each classification, there is also a sub classification. Each classification has inner classifications which are based on the original criteria for mining the data. These are said to be sub attributes. The algorithm employs statistical method in forming a rule. Basically, the system used these techniques to display the crime type that is most committed, location with highest crime, age group with highest crime, state/LGA with highest crime, etc. Sorting rule is used to form this assertion. When an array is sorted in descending order, those elements at the beginning of the array are the ones with highest frequency, therefore to form the rules from the sorted array becomes easier. The data mining algorithm goes thus:

1.0   Input data
*   1.1   Input data mining criteria (select attribute)*
*   1.2   read data from the database and store in unclassified dataset*
*   1.3   initialize sub-attributes*
2.0   Manipulate data
*   2.1   convert continuous values to categorical values*
*   2.2   set classified dataset to empty*
3.0   LOOP through the unclassified dataset
*   3.1   Pick a class from an instance based on selected attribute*
*   3.2   Search through the Classified Dataset for the class*
*   3.2.1   If class is not found then add a new node of the class to the classified dataset*
*   3.2.1.1   Loop through the sub attributes*
*   3.2.1.2   Pick a subclass from the instance*
*   3.2.1.3   Add the subclass as a sub node to the class node*
*   3.2.1.4   Repeat from 3.2.1.2 until all the sub attributes are treated*
*   3.2.2   Else update the node where the class is found in the dataset*
*   3.2.2.1   Loop through the sub attributes*
*   3.2.2.2   Pick a subclass from the instance*
*   3.2.2.3   Search for the subclass in the parent class node in the classified dataset*
*   3.2.2.4   If subclass is not found then add the subclass as a sub node to the class node*
*   3.2.2.5   Else update the subclass sub node in the class node*
*   3.2.2.6   Repeat from 3.2.2.2 to 3.2.2.5 until all the sub attributes are treated*
*   3.3   Repeat from step 3.1 to 3.2.2.5 until all data are classified*
4.0. Generate Rules for the manipulated data
*   4.1   Sort the classified dataset in descending order*
*   4.2   Set n = 0*
*   4.3   Loop through the nodes of the classified dataset*
*   4.3.1   Pick a class at node n*
*   4.3.2   If n = 0 and the number of occurrence of class in node n is greater than the class in node n+1 then the rule follows that the class in node n is the best fit for the test. Go to step 4.4*
*   4.3.3   Else if n > 0 and the number of occurrence of class in node n is greater than the class in node n+1 then the rule follows that all the classes from node 0 to n are the best fit for the test. Go to step 4.4*
*   4.3.4   Else set n = n + 1*
*   4.3.5   Repeat from step 4.3.1*
*   4.4   Return the results*

5.0   Print results
*   5.1   Display the results node by node*
*   5.2   End*

IV. System Implementation

These sections document the implementation and the results of the proposed system. Fig. 3 shows list of all reported cases that are yet to be investigated. To display this, simply login and select state and the LGA where the crime was committed.

Fig. 4 displays the investigated crimes (centralized database) where one can mine data. In mining crime data, the following 13 different parameters were used: Person, Sex, Age, Marital Status, Occupation, Educational Status, State, LGA, Crime Location, Crime Date, Crime Time, Crime Type and Tribe. If a parameter is selected, the program analyzes the crime data based on that parameter. Fig. 5 shows mining by age range. Data mining results on classification based on age showed that out of 973 crime data collected in 2013, criminals within the age range of 31-40 committed the highest number of crimes with 263 cases. In each analysis, there are always sub-analysis which re-analyzes the crime cases using all other parameters except the selected one.

Fig. 3.   List of Crime Cases Yet to be investigated.
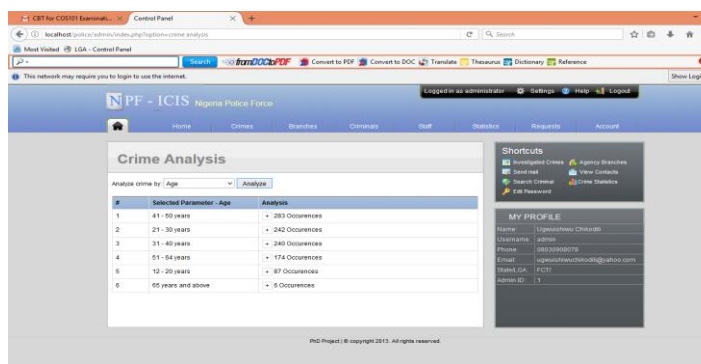

Fig. 4.   List of Investigated Crimes.

Fig. 5.   Crime Analysis by Age Range.

Table I shows a clearer picture of the result in Fig. 5. The sub-analysis of the result on Fig. 5 is shown on Fig. 6. Result

from Fig. 6 shows that Beach Junction Nsukka is a location with the highest number of crimes.

Table II shows a comprehensive crime statistics of all crime types committed in 2013 based on the collected. The result is presented state by state and the total occurrence of each crime type within the year is given. This crime distribution is based on the state of origin of the criminals i.e. citizens of the country that committed crime in Enugu state. Data mining results on classification showed that out of 973 crime data collected in 2013, Anambra state has the highest number of criminal cases with 192 crimes. It may also interest a reader to note that because this research was carried out in the Eastern part of the country, the concentration of these criminals' states of origin was spread mostly in the South eastern part of the country.

TABLE I.   THE CRIME ANALYSIS RESULT BY AGE RANGE AS SHOWN IN FIG. 5

#	Age Range	Analysis
1	41 - 50 years	228 Occurences
2	21 - 30 years	229 Occurences
3	31 - 40 years	263 Occurences
4	51 - 64 years	168 Occurences
5	12 - 20 years	83 Occurences
6	65 years and above	2 Occurences
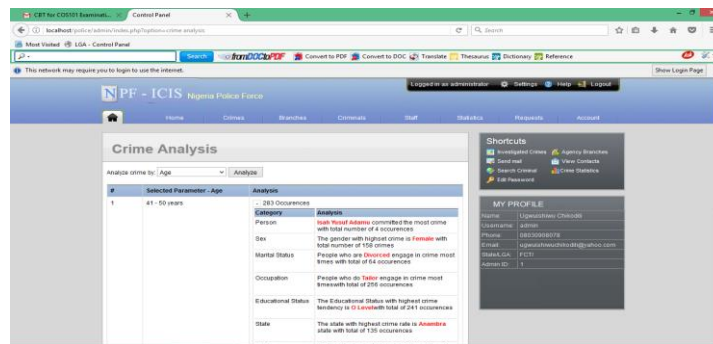
Fig. 6.   The Sub-Analysis of the Result shown in Fig. 6.

TABLE II.   CRIME STATISTICS FOR 2013 BASED ON THE CRIMINAL'S STATE

#	Criminal's State	Abduction	Assault	Assault Occ. Harm	Car Snatching	Car Theft	Certificate Forgery	Conduct	Defilement	false witness	Fire Incident	Forgery of Police ID	G. Harm	House Breaking	Illegal Roadblock	Indecent Assault	Keke NAPEP Snatching	Kidnapping	M. Damage	M/Stealing	Malicious Damage	Missing Person	MT Stealing	Murder	OBT	Phone Snatching	Pocket Picking	Raping	Robbery	Sexually Harassment	Stealing	Store Breaking and S	Threat to life	Threatening Violence	Vandalization	Wounding
1	Abia	2	3	3	5	1	6	4	5	1	1	1	5	1	3	2	3	3	3	2	3	4	1	3	2	0	1	4	4	3	5	4	1	2	5	3
2	Adamawa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Akwa Ibom	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Anambra	5	8	7	7	9	9	5	4	7	2	5	5	4	4	7	5	3	6	5	2	5	8	5	7	7	6	4	4	5	8	3	8	7	8	6

5	Bauchi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	53	0	0	0
6	Bayelsa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	Benue	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	Borno	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	Cross River	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	Delta	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	Ebonyi	1	1	1	1	2	0	1	0	2	0	1	1	3	3	1	1	0	2	1	0	0	3	0	4	1	1	1	1	0	1	0	1	0	0	2
12	Edo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	Ekiti	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	Enugu	1	0	1	0	2	1	2	1	4	2	3	1	0	2	2	3	1	1	2	3	0	1	1	2	1	0	1	3	4	2	4	1	0	2	1
15	Gombe	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	Imo	1	1	1	1	0	2	3	2	4	2	1	2	2	1	2	2	1	2	0	0	2	1	0	2	0	3	2	0	0	2	2	2	2	2	1
17	Jigawa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	Kaduna	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	Kano	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	Katsina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	Kebbi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	Kogi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	Kwara	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	Lagos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	Nassarawa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	Niger	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	Ogun	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	Ondo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	Osun	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	Oyo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	Plateau	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	Rivers	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	Sokoto	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	Taraba	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	Yobe	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	Zamfara	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	FCT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Total	**10**	**13**	**13**	**14**	**14**	**18**	**15**	**12**	**18**	**7**	**11**	**14**	**10**	**13**	**14**	**14**	**8**	**14**	**10**	**8**	**11**	**14**	**9**	**17**	**9**	**11**	**12**	**12**	**12**	**18**	**13**	**13**	**11**	**17**	**13**

V. CONCLUSION

A Real- Time Integrated Crime Information System was developed and data analysis performed. It is intended that when this application is fully deployed, data analysis should be done on crime data available in the RICIS centralized database. This work used classification and Association rule algorithms to identify useful crime trend and patterns. These machine learning algorithms (Psuedocode) were implemented using PHP and other programing language tools. The results of this data mining will support LEAs in effective decision making, management of their limited resources and will considerably reduce crime rate and increase the security standard of the nation.

A. Limitations of the Study

There was a big challenge during data collection because most of these Nigerian LEAs do not have electronic crime data. The data I got was manually picked into a table format one after the other and this made things difficult. Nigeria Custom Services, Nigeria Police and NAFDAC offices were visit by the researcher but data were collected from only two police stations – Eleme Police station and Nsukka police station, the major reason being that there was no electronic database to access.

B. Recommendations

The system is used by virtually every citizen of the country to fight against crime. Specifically, this application is meant for Nigerian LEAs to manage societal crime issues. LEAs of any country, other governmental agencies, private and public organization can equally use this system with a little modifications. Government can use this application to determine the efficiency of LEAs on crime management by taking the statistics of crime handled within a specified time.

WEKA and other mining software tools may be considered in the data analysis instead of developing new mining application, increase in the data size and mining parameters, and as well data mining techniques such as clustering and linear regression may be may be adopted for a more improve analysis.

Comments: The data collected are not sufficient to make for conclusive evidence. How does this provide for privacy of persons under investigation? The supposed 'criminals' are suspects, or have they been convicted by a court of competent jurisdiction?

REFERENCES

[1] Bharati M. Ramageri. "Data Mining Techniques and Applications", Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305. Available at http://www.ijcse.com/docs/IJCSE10-01-04-51.pdf [accessed Jan 26, 2017].

[2] M. Keyvanpoura, M. Javidehb, M. Ebrahimia (2011). Detecting and investigating crime by means of data mining: a general crime matching framework Procedia Computer Science 3 (2011) 872–880.

[3] Neha D. and B.M. Vidyavathi, "A Survey on Applications of Data Mining using Clustering Techniques", International Journal of Computer Applications (0975 – 8887) Volume 126 – No.2, September 2015. Available at http://www.ijcaonline.org/research/volume126/number2/neha-2015-ijca-905986.pdf[accessed Jan 27, 2017].

[4] Xingan Li, "Application of Data Mining Methods in the Study of Crime Based on International Data Sources", 2014. Available at https://tampub.uta.fi/bitstream/handle/10024/95108/978-951-44-9419 2.pdf? sequence=1[accessed Feb 08, 2017].

[5] D. Tyagi, and S. Sharma. (2018). "An Approach to Crime Data Analysis: A Systematic Review." International Journal of Engineering Technologies and Management Research, 5(2:SE), 67-74. DOI: 10.5281/zenodo.11975.

[6] C. Chauhan and S. Sehgal, "A review: Crime analysis using data mining techniques and algorithms," 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, 2017, pp. 21-25, doi: 10.1109/CCAA.2017.8229823.

[7] S.Yamuna and N. S. Bhuvaneswari, "Data mining Techniques to Analyze and Predict Crimes" The International Journal of Engineering and Science (IJES) Vol 1 No. 2, PP 243-247, 2012. Available at http://www.theijes.com/papers/v1-i2/AJ01202430247.pdf [accessed Feb 08, 2017].

[8] H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin and M. Chau, "Crime data mining: a general framework and some examples," in Computer, vol. 37, no. 4, pp. 50-56, April 2004, doi: 10.1109/MC.2004.1297301.

[9] R. Bhargava, P. Singh and R. S. Sangwa (2018). Analysis of Crime Data Using Data Mining Algorithm. International Journal of Engineering Sciences & Research Technology 7(2), 675-681.

[10] Akgöbek, Ömer (2013). A rule induction algorithm for knowledge discovery and classification. Turkish Journal of Electrical Engineering & Computer Sciences, Vol. 21, DO - 10.3906/elk-1202-27.

[11] H. Chen, W. Chung, Yi Qin, M. Chau, J. Jie Xu, G. Wang, R. Zheng and H. Atabakhsh, "Crime Data Mining: An Overview and Case Studies", 2003. Available at https://www.researchgate.net/publication/2870463_Crime_Data_Mining_An_Overview_and_Case_Studies [accessed Jan 26, 2017].

[12] S. P. Deshpande, and V. M. Thakare (2010). Data Mining System and Applications: A Review. International Journal of Distributed and Parallel systems (IJDPS) 1(1), DOI : 10.5121/ijdps.2010.1103 32.

[13] UKEssays (2018). Survey Of Data Mining Techniques On Crime Data Criminology Essay. https://www.ukessays.com/essays/criminology/survey-of-data-mining-techniques-on-crime-data-criminology-essay.php?vref=1.

[14] K. Zakir Hussain, M. Durairaj and G. Rabia Jahani Farzana, "Application of Data Mining Techniques for Analyzing Violent Criminal Behavior by Simulation Model", International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 2, No. 1, 2012. Available at http://ijcsits.org/papers/Vol2no12012/5vol2no1.pdf [Accessed on 26th Jan, 2017].

[15] Z. S. Zubi and A. A. Mahmmud, "Crime Data Analysis Using Data Mining Techniques to Improve Crimes Prevention", International Journal of Computers, Vol 8, 2014, pp.39-45. Available at http://www.naun.org/main/NAUN/computers/2014/a022007-096.pdf. [accessed 26th Jan, 2017].

[16] Z. S. Zubi and A. A. Mahmmud, "Using Data Mining Techniques to Analyze Crime patterns in the Libyan" Available at http://www.wseas.us/e-library/conferences/2013/Budapest/IPASRE/IPASRE-09.pdf [Accesses on 27th Jan,2017].

[17] L. McClendon and N. Meghanathan," Using Machine Learning Algorithms to Analyze Crime Data", Machine Learning and Applications: An International Journal (MLAIJ) Vol.2, No.1, March 2015. Available at http://airccse.org/journal/mlaij/papers/2115mlaij01.pdf [Accessed 26th Jan, 2017].

[18] H. Hassani, Xu Huang, E. S. Silva, and M. Ghodsi. "A Review of Data Mining Applications in Crime" 2016. Available from: https://www.researchgate.net/publication/301579904_A_Review_of_Data_Mining_Applications_in_Crime [accessed Jan 27, 2017].

[19] J. Agarwal, R. Nagpal and N. R. Sehgal, "Crime Analysis using K-Means Clustering", International Journal of Computer Applications (0975 – 8887) Volume 83 – No4, December 2013.Available at https://pdfs.semanticscholar.org/dcb6/bff7931d085e7bf6ff004b0d28fccfca22df.pdf [accessed Feb 08, 2017].

[20] M. Nayak, V. Yadav, Y. Patil (2019), Crimerate Prediction using Datamining. International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS), 8(4), 50-52.

[21] K. B. Al-Janabi and H. K. Abdullah (2010). Crime Data Analysis Using Data Mining Techniques to Improve Crimes Prevention Procedures. Iraqi conference for Information technology (ICIT).

[22] S. Varan Nath, "Crime Pattern Detection Using Data Mining", 2006. Available at http://cs.brown.edu/courses/csci2950-t/crime.pdf [accessed 26th Jan, 2017].

[23] U Mande, Y. Srinivas, J.V.R.Murthy / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 4, July-August 2012, pp.149-153.

[24] S. Rajkumar, P. M. Sakkarai, J. J. Soundarya, P. Varnikasree (2019). Crime Analysis and Prediction Using Data Mining Techniques. Proceedings of the First International Conference on New Scientific Creations in Engineering and Technology (ICNSCET-19), 602-607.

[25] S. Kim, P. Joshi, P. Kalsi, and P. Taheri (2018). Crime Analysis through Machine Learning. 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON).

[26] P. Clark & T. Niblett, "The CN2 Induction Algorithm". Machine Learning Journal vol.3 No 4, Pp2, 1988. http://www.cs.utexas.edu/users/pclark/papers/cn2.pdf.

[27] J. Stefanowski (2010). Lecture note on Induction of Rules, Institute of Computing Sciences, Poznan University of Technology, Poznan, Poland.

[28] Wikipedia,"CN2 algorithm". Available at https://en.wikipedia.org/wiki/CN2_algorithm accessed on 17/04/16.