International Journal of Advances in Scientific Research and Engineering (ijasre)

E-ISSN : 2454-8006

DOI: 10.31695/IJASRE.2023.9.5.1

Volume 9, Issue 5
May - 2023

Natural Language Processing-Enabled Framework for Regional Crime Analysis Using Unstructured Text: A Case of Abuja, Nigeria

Jalila Ibrahim Coomassie[1], and Dr. Ahmed Babalaji Ndanusa[2]

Research Scholar[1], Supervisor[2]

Department of Computer Science

University of Abuja

Abuja, FCT,

Nigeria.

ABSTRACT

Background: *Globally, criminal activities have significantly increased over the years, which makes the need for a reliable and insightful analytical approach increasingly vital to take proactive measures. Specifically, Abuja is one of the fastest-growing cities in sub-Saharan Africa. The city is however constrained by the current approaches in crime analysis toward proactive decision-making. Thus, presenting an ardent need for this research to provide safety and security to the civilian population, especially those within the crime-prone area(s).*

Objective: *Objective of this research is to develop a framework for crime analysis by identifying the regional frequency of the crimes committed in some selected parts of Abuja, FCT.*

Method: *This research with the aid of the Natural Language Processing approach, crime reports from 2015 to 2021 of Wuse, Garki, Gwagwalada, Nyanya, Maitama and Utako comprising murder, kidnapping, theft, juvenile delinquency and homicidal acts established by the Nigerian Police Command was used to extract textual data therefrom. Using Orange data mining, which is a python-based environment; the reports were transformed to and pre-processed. Thereafter, crimes were mapped regionally using a visually appealing visualization technique known as Word cloud.*

Results: *This essentially, clustered crimes based on their frequencies and regionally. This approach revealed critical details required proffering insights towards decision-making required for crime mitigation. In addition, to ascertain the relationship between crime hotspots the police divisional stations and slum settlement in the study area.*

Conclusion: *The higher education attainment will be the cure for criminal activities in Nigeria. Government should also create more jobs because high unemployment rates will compel people to commit crimes and this will increase crime rate in Nigeria. Lastly, there should be high budgetary provision towards poverty alleviation programme because higher poverty may lead to higher crimes rate due to depression or mental illness associated with being poor and this will decrease the rate of return of legal activities and more likely to increase return of illegal activities.*

Keywords: *Crime hotspot*, Text Analytics, Text *visualization, Text pre-processing, Word Cloud,*

---

1. INTRODUCTION

1.1 Background of Study

Inherently, crimes are defined as an unlawful or illegal infliction of harm to individuals or animals and/or dispossession of properties from an individual. Globally, crime is one of the human security issues confronting humanity. With numerous countries facing increasing cases of homicide, armed robbery, kidnap, drug trafficking, sex trafficking, illegal gun running etc. United Nations Office on Drugs and Crime in 2011 estimated global homicides at 468,000 and more than a third (36%) was estimated to have occurred in Africa, 31% in the Americas, 27% in Asia, 5% in Europe and 1% in the tropical Pacific region. Africa has remained a leading continent in global crime statistics. Prevalence of violent and non-violent crimes has remained high in South Africa and Nigeria.

In Nigeria, crimes manifest in a convulsive increase of both violent and non-violent crimes. Cases of ransom-driven kidnapping, armed robbery, assassination and are now ravaging the polity like a tsunami and spreading a climate of fears and anxieties about

public safety [1]. These increased cases of crimes in Nigeria has placed Nigeria on the global crime map leading to intermittent travel advisory warnings of foreign countries to expatriates in Nigeria or intending visitors to Nigeria. Especially, since these crimes portrays the inability of government to provide a secure and safe environment for lives, properties and the conduct of economic activities considering the alarming increase in criminal activities in Nigeria.

Abuja, the capital city of Nigeria, is one of the fastest growing capital cities in sub-Saharan Africa. Recently, the city is experiencing throes of crime of armed robbery, homicide, as well as kidnapping in an unprecedented rate. The scope of crime prevention has grown considerably in the last few years. What was previously the sole concern of the police and the private security industry has spread into areas from real estate developer, car manufacturers, residents' groups, building public facilities like society offices and shopping centres, all these calls for continuously using improved new ways to prevent crime. Criminal victimization has serious consequences for the citizens and society because high standard of living is undermined by high level of criminal victimization [2]. The issue of crime in the city of Abuja has increased since the relocation of the federal capital to Abuja. While, there have been some measures to mitigate threats emanating from these heinous activities, the current approaches employed using crime reports established by the Nigerian Police or historical data on crimes are still constrained, as discussed in the ensuing subsection.

The underpinning aim of this study is to develop a framework for crime analysis. This essentially was attained by reaching the following research objectives

i. To eliminate the less relevant textual data from the crime reports.
ii. To identify the regional frequency of the crimes committed in Abuja.
iii. To form a reliable means of crime visualization.
iv. To assess the accuracy of the developed framework.

2. LITERATURE REVIEW

Essentially, the primary step in developing a body of knowledge commences by considering previous studies in order to gain insights on the extent to which the extant studies have attained, and identify the procedures employed in addressing the focus of these studies [3]. Therefore, a critical review of relevant literature was ensured to proffer insights on the major challenges associated with crime mitigating strategies, the approaches that have been employed to pre-process unstructured data and the approaches used in crime analysis. Therefore, this chapter provides a comprehensive discussion on crimes in section 2.1. Section 2.2 discusses extant approaches employed in crime analysis. While sections 2.3 and 2.4 present global crime trend and the study area with its associated crimes.

2.1 Background on Crimes

Crimes are defined by criminal law. If a person engages in acts of behaviours that are considered to be harmful to society, they could be found guilty of committing a crime. Crimes are generally prosecuted in a criminal court. Someone convicted of a crime may be forced to pay fines and may also lose their personal freedoms and privileges by being sentenced to jail or prison time [4]. Personal crimes are most commonly generalized as a violent crime that causes physical, emotional, or psychological harm to the victim. These crimes are offenses against the person, and can include but are not limited to crimes presented and defined in the ensuing subsection.

i. Assault and Battery: Assault refers to the intentional creation of a reasonable apprehension of harm. In other words, assault is a situation in which one person causes another person to fear being harmed. Assault and battery are most commonly considered two distinct personal crimes, although many states do merge the two crimes into the one crime known as “assault and battery. Battery refers to the unauthorized application of force against another person’s body. This results in offensive touching, or actual physical injury. As some jurisdictions define assault as attempted but failed battery, battery charges are commonly grouped with assault to form the single charge of assault and battery [5].
ii. False Imprisonment: False imprisonment refers to one person forcibly restraining another person, against their will, with a risk of being seriously injured or killed. Any person who intentionally restricts another person’s freedom of movement, without their consent, may be liable for false imprisonment [6].
iii. Kidnapping: Kidnapping is defined as the carrying away or confinement of a person by force or deception, without that person’s consent. In other words, kidnapping is the seizure and detention of a person without their consent, while intending to carry away the person at a later time, hold the person for ransom, etc.
iv. Homicide: Homicide includes crimes such as first and second-degree murder, involuntary manslaughter, and vehicular homicide; and

DOI: 10.31695/IJASRE.2023.9.5.1

v. Rape: Rape also includes statutory rape and sexual assault.
vi. Property crimes**:** or offenses against property, do not necessarily involve the harm of another person. Rather, these crimes involve interference with another person's right to use or enjoy their own property. Some examples of property crimes include but are not limited to:
vii. Theft: Larceny refers to a type of theft in which a person takes another person's property and carries it away, with the intent to permanently deprive the legal owner of their property. Robbery is known as theft by force, and may also be considered a personal crime as it often results in physical and mental harm. Burglary occurs when a person breaks and enters into a home or building, intending to commit a crime. This crime is generally theft, although assault or arson may also constitute burglary;
viii. Arson: Arson is the wilful and malicious burning or charring of another person's property or structure;
ix. White Collar Crimes: Embezzlement refers to a type of white-collar crime in which a person entrusted with the finances of another person or business illegally takes that money for their own personal use. Forgery is another example of a white-collar property crime, because it is the creation, alteration, forging, or imitation of any document with the intent to defraud another person of their property;

Inchoate, or incomplete, refers to crimes that were initiated but not brought to completion. A person would need to take a substantial step towards completing a crime, as opposed to simply intending to commit a crime. A few examples of inchoate crimes include:

i. "Attempted" crimes, such as attempted robbery, attempted murder, etc.;
ii. Solicitation: Crimes involving requesting, asking, hiring, commanding, or encouraging someone else to commit a crime; and
iii. Conspiracy: Crimes involving multiple actors coming together to engage in criminal activity.
iv. Statutory crimes are violations of specific state or federal statutes. They may involve either property offenses or personal offenses. An example of this would be alcohol related crimes, such as DUI or selling alcohol to a minor [7].

After a succinct discussion on crimes and related terms, the ensuing sub-section presents various approaches currently employed in analysing criminal activities in different parts of the world.

2.2 Global Trends of Crime Rate

Globally, the rate of crime is increasing significantly. Since crime is neither systematic nor random, it cannot be expected. The trendy technologies and advanced strategies facilitate criminals in achieving their black act. In step with the Crime Records Bureau, crimes like felony, setting fireplace to a property (arson), etc. are relatively remittent however on the opposite hand, crimes like murder, regulatory offense, etc. are enlarged, places wherever the rate lies high are thought about as crime hotspots [8]. Thus, it is necessary to predict the prevalence of future crime hotspots and crime suspects by exploitation crime trends. Crime trend prediction is useful to form a call-in crime interference activity [9].

Crime analysis is a crucial approach which will be optimized using various techniques and processes. Several studies have been completed by various researchers within this domain. The current study is limited to using the databases to identify criminal hotspots. None of them, however, took into account the type of crime or the time of the incident as a factor. The static maps provided by [10]. Yu et al. (2017) lack any interactive features. To overcome these drawbacks, the system suggests a framework that provides visualization tools that take into account the type of crime to identify crime hotspots and aids in visualizing these regions with the use of interactive Google maps.

Similarly, some papers focused on usage of decision trees for crime analysis. [11]. Ahishakiye et al., (2017) used the attributes population of country, Median Household income, percentage of individuals who are unemployed with age greater than 16, sort of crime, etc. which only predicts whether in a neighbourhood there'll be high, medium or low percentage of violent crimes which will happen in future. The methods proposed by them didn't predict the sort of crime which will happen. [12]. Chen et al., (2018) explored time series of data, and forecast crime trends in the following years. They predicted crime category given time and location, to overcome the problem of imbalance, they merged multiple classes into larger classes K-NN, Naïve Bayes approach over sampling strategy and Random Forest, using feature selection to improve accuracy. Also, they applied several state-of-the-art data mining and time Series of Data and Forecast techniques that are specifically used for crime prediction. The experimental results will show that the Tree classification models performed better on the classification task over k-NN and Naive Bayesian approaches. Holt-Winters with multiple seasonality which gives best results when predicting crime trend, to minimize the occurrence of Crime.

DOI: 10.31695/IJASRE.2023.9.5.1

If we can know where, who and when with sufficient time lag, we can reach a goal of "zero crimes". Hence the need for a crime prediction system. By proposing a work on a 4-dimensional crime prediction model based on the activity of individuals that we could model it in mathematical forms [13].

To provide new perspectives for understanding crime, and how to control it using POI (point of interest) features to assist the demographic features. Propose the use of POI features to assist the demographic features, and to use taxi flow as hyperlinks to supplement the geographical neighbours. The intuition behind the hyperlinks is that the taxi flow models the social interaction among non-adjacent regions, which potentially propagate crime or resources and information used in crime control. Adopt the negative binomial regression modal over the linear regression model, mainly because the count-based regression models guarantee positive prediction, while the linear regression may give negative crime rate as prediction [14].

With the increasing concern of the crime rate, there is need for law enforcement Officers to discover patterns and trends for making forecasts, Data mining being sensitive to quality of input data, may be inaccurate having missing information (noise, redundant data). Data mining attributes could be challenging in its own ways. Big data can extensively be used to transform large unstructured or structured raw data into crucial and meaningful information which will help in forming a healthy decision support system for the judiciary and legislature to enforce law and order towards keeping crimes in check and making strategic decisions for safety and well- being of the society [15].

Crime prediction and finding relevant data from great amount of crime knowledge aims to stop the crime rates. A survey was conducted so Crime prognostication will be improved by the employment of economical knowledge assortment and different ways. Combined techniques are needed to make a higher crime prediction by desegregation multiple models to resolve single downside for better performance [16].

Also, there is need to highlight the most dangerous areas in research and compare it with crime rate to other neighbourhoods. Elaborate contributions to topic crime prevention. Identify a research gap which includes cyber security, and the measurement of the crime rate as well as a stabilizing process on the crime rate which started in 2015 in the area of study.

Despite considerable research efforts, yet there is a need to have a better analytical approach, which direct police patrols toward criminal activities. Previous studies are lacking to achieve a reliable crime analysis based on historical data with the aid of a framework, which is capable of guiding towards an insightful analysis therefrom. Overall, these results provide early identification of crime, hot spots with higher crime rate, and future trends with improved predictive accuracy than with other methods and are useful for directing police practice and strategies. Ability to control Crime rate by designing a computational prediction system that can improve crime analysis to further enhance the safety and security of cities and help to prevent crimes. Ability to control Crime rate by designing a computational prediction system that can improve crime analysis to further enhance the safety and security of cities and help to prevent crimes [17].

2.3 Review of Related Studies

Many researches have tried explaining why crimes occur in a certain area or is there any pattern that can be concluded from the past events. One of such theory that answers these questions is the crime prediction theory. According to the prediction theory by ([18]. McCaghy et al., 2016), crime does not happen in a random fashion, it is either opportunistic or planned. It states that any criminal activity occurs when there is intersection of work space of a target and the offender. The people's work space is comprised of places he/she visits in a day-to-day routine, like workplace, educational institutes, shopping malls, recreational areas etc. All these specific locations of the offender or victim are also called nodes.

Personal paths, the routes people take every day connect with various nodes creating a circumference of personal space. This personal area is also person's awareness space. Thus, crime pattern theory states that crime involving two people can only occur when the personal spaces of both intersects at one point or another. Thus, it can be considered that crimes are not completely random, they can be studied and analysed using various approaches. For instance, data mining in the study and analysis of criminology can be categorized into main areas, crime control and crime suppression. ([19]. Parsons et al., 2011) introduced a framework for crime trends using a new distance measure for comparing all individuals based on their profiles and then clustering them accordingly, [20]. Gupta (2013), highlighted the existing systems used by Indian police as e-governance initiatives and also proposes an interactive query-based interface as crime analysis tool to assist police in their activities.

Similarly, an interface which is used to extract useful information from the vast crime database was proposed by National Crime Record Bureau (NCRB) to identify crime hot spots using crime data mining techniques such as clustering etc. The effectiveness of the proposed interface has been illustrated on Indian crime records. [21]. Fadzil et al., (2022) discussed the development of Visual

DOI: 10.31695/IJASRE.2023.9.5.1

Interactive Malaysia Crime News Retrieval System (i-JEN) and describe the approach, user studies and planned, the system architecture and future plan.

Evidently, their main objectives were to construct crime-based event; investigate the use of crime based event in improving the classification and clustering; develop an interactive crime news retrieval system; visualize crime news in an effective and interactive way; integrate them into a usable and robust system and evaluate the usability and system performance and the study will contribute to the better understanding of the crime data consumption in the Malaysian context as well as the developed system with the visualization features to address crime data and the eventual goal of combating the crimes [22]. Thiprungsri (2010a), examines the application of cluster analysis in the accounting domain, particularly discrepancy detection in audit. The purpose of his study is to examine the use of clustering technology to automate fraud filtering during an audit.

Equally, cluster analysis was employed to help auditors focus their efforts when evaluating group life insurance claims. [23]. Nazlena et al., (2010a), look at the use of missing value and clustering algorithm for a data mining approach to help predict the crimes patterns and fast up the process of solving crime. [24]. Malathi. Nazlena et al., (2010b), used a clustering/classify based model to anticipate crime trends. The data mining techniques are used to analyse the city crime data from Police Department. The results of this data mining could potentially be used to lessen and even prevent crime for the fourth coming years [25]. A research work focused on developing a crime analysis tool for Indian scenario using different data mining techniques that can help law enforcement department to efficiently handle crime investigation. The proposed tool enables agencies to easily and economically clean, characterize and analyse crime data to identify actionable patterns and trends. ([26]. Al-Janabi, 2011), presents a proposed framework for the crime and criminal data analysis and detection using Decision Tree Algorithms for data classification and Simple K Means algorithm for data clustering.

The paper tends to help specialists in discovering patterns and trends, making forecasts, finding relationships and possible explanations, mapping criminal networks and identifying possible suspects. [4]. Agarwal, (2013), applied myriad of tools on crime data sets to mine for information that is hidden from human perception. With the help of state-of-the-art visualization techniques, we present the patterns discovered throu.gh our algorithms in a neat and intuitive way that enables law enforcement departments to channelize their resources accordingly. ([27]. Thiprungsri, 2010b), examine the possibility of using clustering technology for auditing. Automating fraud filtering can be of great value to continuous audits.

The objective of their study was to examine the use of cluster analysis as an alternative and innovative anomaly detection technique in the wire transfer system. ([3]. Rabia & Farzana, 2012) tried try to capture years of human experience into computer models via data mining and by designing a simulation model.

2.4 Research Gap and Our Contribution

From the preceding discussion, it has been revealed that, despite the remarkable successes recorded in the highlighted studies in the domain of criminology, Visual and intuitive criminal and intelligence investigation techniques were recommended by the study conducted in ([3]. Rabia & Farzana, 2012). Concisely, the approaches employed in various phases of this research suggest substantial contributions to the existing body of knowledge; in the areas of Natural Language Processing, Criminology and Unstructured textual data pre-processing. The study demonstrated the novelty and the applicability of such datasets in regional crime identification. These contributions are further explained in the following subsections:

While theoretically, this research explained the role of NLP technique in crime analysis. In addition, it has also demonstrated how handwritten reports can be transformed and employed in analytical inferences towards regional crime identification in the developed framework. This framework has proven to be an auspicious step for understanding of regions and their corresponding levels of crime vulnerability. Thus, this research has contributed to the growing body of knowledge.

Broadly, a framework should have the ability of guiding researchers in practical and design considerations. To this effect, the phases embedded in the developed framework can fill this gap in addressing similar issues when implemented. Additionally, the study conducted by [28]. Migueles et al., (2019), equally identified the need for an open-source environment to improve the adoption of large datasets in deriving analytical insights therefrom.

As inherently established, the underpinning aim of this research is to aid the Security Agencies in mitigating and averting crimes within Abuja. It is therefore important to reveal that, the results provided in the phase of regional crime visualization present a very auspicious means of meeting this need. Essentially, the regional crime visualization will advertently aid in decision-making, since it has clustered regions based on their frequencies of crimes.

DOI: 10.31695/IJASRE.2023.9.5.1

3. METHODOLOGY

Within the scope of this research, this chapter presents the tools and methods employed in bridging the identified gaps in the preceding chapter. Concisely, the research design is explained in section 3.1. While the conceptual framework illustrating the phases of the research is explained in section 3.2.

3.1 Research Design

A research design plays a crucial role in any meaningful research. It represents the general guiding precepts utilized by a researcher to combine several components of a research in a rational and structured manner, in order to successfully attain research objectives. Similarly, a research design ensures the adequacy of the procedures employed in obtaining valid and accurate answers to research questions [29]. It serves as a focal point for well-conducted research and provides comprehensive processes for the collection, analysis and interpretation of data. Consequently, Figure 3.1 illustrates the components of the adopted design in the course conducting the research.

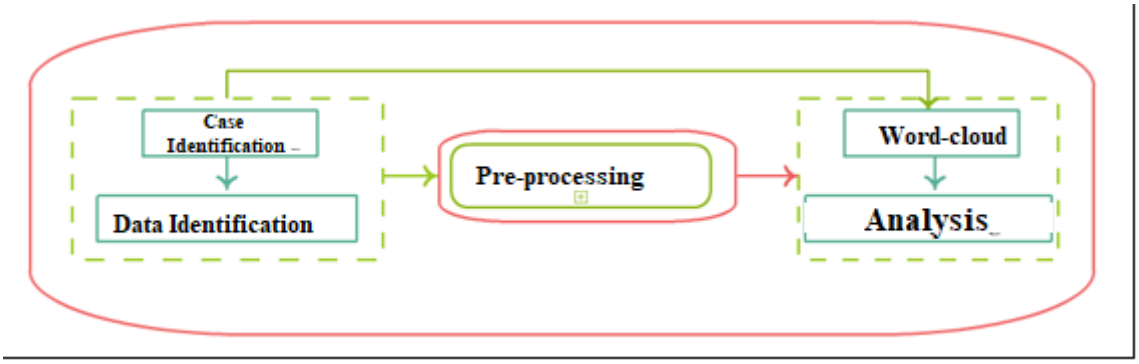


Figure 3.1: Research Design: Adapted from (Pandey & Pandey, 2013)

Within the scope of this present research, it can be concluded that the utilization of unstructured textual data classifies regional crime vulnerability, reflect effectively on the Qualitative mode of research. Therefore, by implementing this research design as shown in Figure 3.1, various tasks were accomplished in phases in order to meet the objectives of the research. The phases with the corresponding tasks are detailed in Figure 3.2.

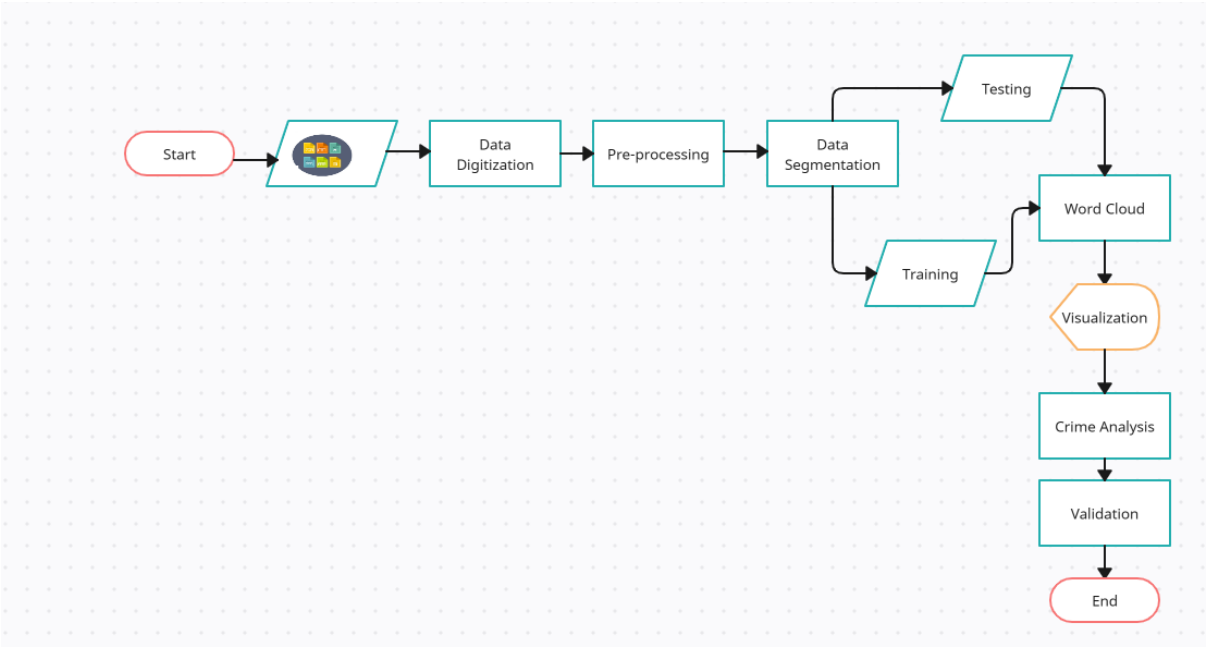


Figure 3.2: Conceptual Framework

3.2 Conceptual Framework

As illustrated using the preceding figure, the ensuing subsections describe various segments of the conceptual framework.

DOI: 10.31695/IJASRE.2023.9.5.1

3.2.1 Data input

Illustratively, the required data for the developed framework was obtained from the Police command depicting the reports of crime cases within Abuja. And it encompasses culpable homicide, theft, rape, juvenile delinquency from 2015 to 2021.After the collection of these reports, which were initially in the manuscript forms. The documents were transformed into digital formats as explained in the ensuing sub-section.

3.2.2 Data Digitization

The most obvious benefit of digitization is to preserve and provide the aggregation of various resources in digital form, using digital technology and resources are preserved for future generation and simultaneously made accessible for current use ([30]. Stolterman et al,2004). In this research, the police reports were digitized in order to provides a means of preserving the content of the materials by creating an accessible facsimile of the reports needed for research, which is illustratively represented using Figure 3.2

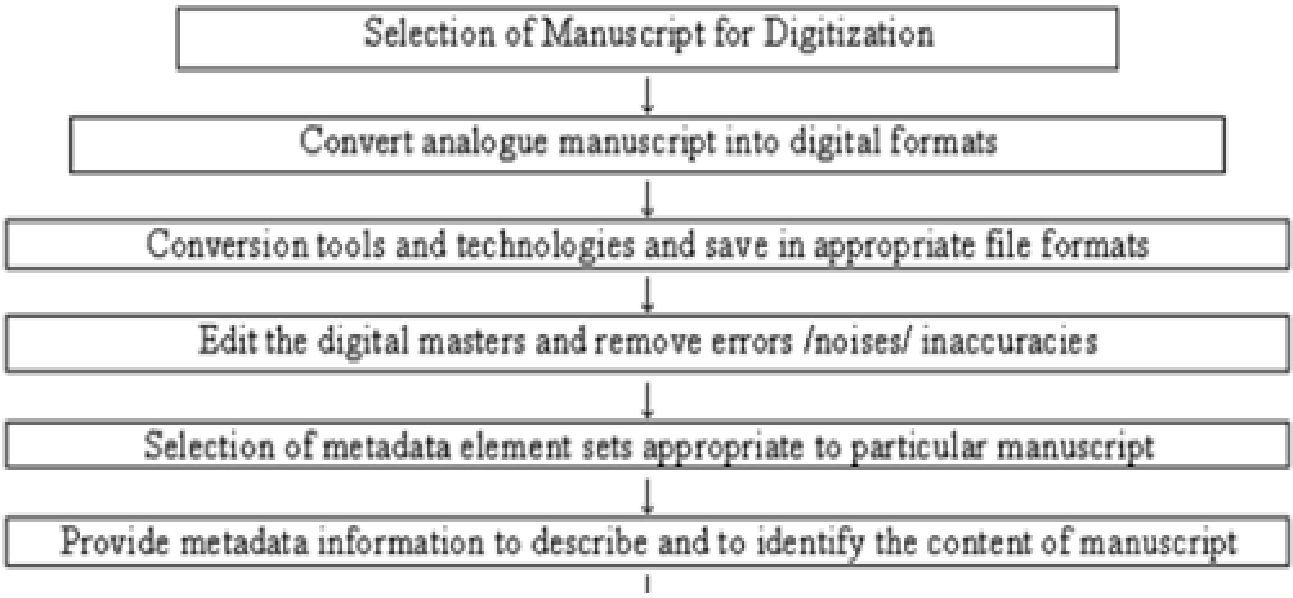


Figure 3.3: Data Digitization Phase

3.2.3 Pre-processing

Here, the digitized textual data were pre-processed to obtain the required records, which were further processed for regional crime identification. Generally, pre-processing involves data cleaning as well as various tasks involved in transforming a set of data to an enhanced format prior to the analytical tasks ([31]. Fatima et al., 2017). In this research, the collection of data sets was done from a data-intensive sources and as such, the uncertainty of these data is intrinsic. Also, as a result of the large volume of digitized data collected for this research, and due to the uncertainty of involved therein, pre-processing this sets of data remains a very complex practice. Lower Casing basically means converting all the words in a text in lower case. This reduces dimensionality because, for example, a system given a text with both “Maitama” and “maitama” will consider them different and so it would be a waste of time and resources to have to process each repeated word.

Also, in Natural Language Processing, before a machine can understand text as it is presented to it, it will have to pre-process it and translate it into a more readable format also referred to as normalization. There are quite a few techniques that are used for pre-processing text, some are general and are used automatically whenever building an NLP program; others are more task specific and depend on the type of text and final use of the system.

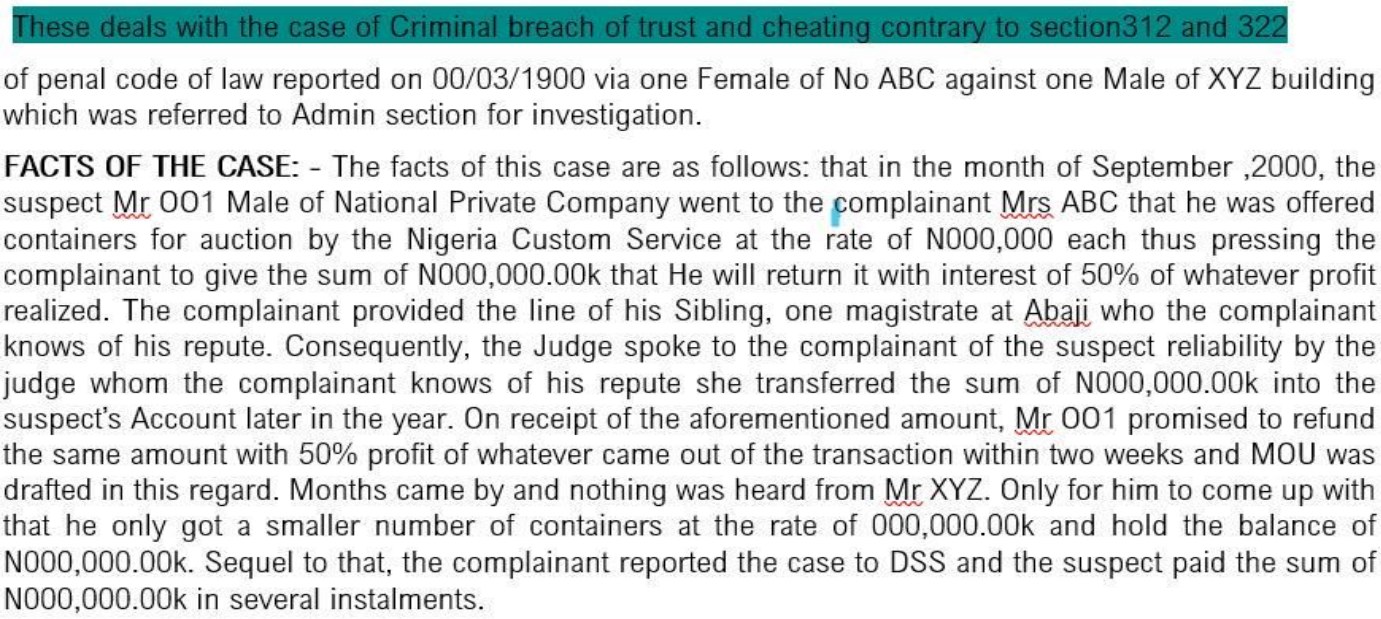
3.3 Removing Noise with Regular Expression (REGEX)

Noise removal is used to get rid of all sorts of characters that could be misleading for the machine, such as punctuation and symbols that are not part of natural language, like HTML tags, or the numbers in enumerated lists. This also reduces dimensionality because words like “me!”, “me”, and “<a>me” will be treated differently if not for the noise removal. Regular Expressions (RE) are a powerful tool that is used for text analysis because it also allows to search for patterns in a text and provides operators for accurate pattern matching when looking to clean the sentences of repeating patterns.

DOI: 10.31695/IJASRE.2023.9.5.1

```
# removing white space
text = re.sub('\t ', '', str(text))
text = re.sub('\t', ' ', str(text))
# removing new line characters
text = re.sub('\n ','',str(text))
text = re.sub('\n',' ',str(text))
# removing apostrophes
text = re.sub("'s",'',str(text))
# removing hyphens
text = re.sub("-",' ',str(text))
text = re.sub("— ",'',str(text))
text = re.sub('_', ' ', str(text))
```

Hence, this section addresses the inherent complexity in pre-processing this vast volume of data that are crucial in this research which in turn, fills the gap of research objective one, by defining a pre-processing technique to eliminate the less relevant textual data from the digitized crime reports. For instance, the established police report on crimes contains words describing the events:



These deals with the case of Criminal breach of trust and cheating contrary to section312 and 322 of penal code of law reported on 00/03/1900 via one Female of No ABC against one Male of XYZ building which was referred to Admin section for investigation.

FACTS OF THE CASE: - The facts of this case are as follows: that in the month of September ,2000, the suspect Mr OO1 Male of National Private Company went to the complainant Mrs ABC that he was offered containers for auction by the Nigeria Custom Service at the rate of N000,000 each thus pressing the complainant to give the sum of N000,000.00k that He will return it with interest of 50% of whatever profit realized. The complainant provided the line of his Sibling, one magistrate at Abaji who the complainant knows of his repute. Consequently, the Judge spoke to the complainant of the suspect reliability by the judge whom the complainant knows of his repute she transferred the sum of N000,000.00k into the suspect's Account later in the year. On receipt of the aforementioned amount, Mr OO1 promised to refund the same amount with 50% profit of whatever came out of the transaction within two weeks and MOU was drafted in this regard. Months came by and nothing was heard from Mr XYZ. Only for him to come up with that he only got a smaller number of containers at the rate of 000,000.00k and hold the balance of N000,000.00k. Sequel to that, the complainant reported the case to DSS and the suspect paid the sum of N000,000.00k in several instalments.

Figure 3.4: Sample of a Digitized Report

From the digitized textual data, in order to identify the frequency of the regional crimes, the relevant features are the locations, as well as the crimes. Therefore, all other details were removed from the digitized report to further ensure a relatively clean data. After this approach, the datasets were further segmented into training and testing, as detailed in the following sub-section.

3.4 Training and Testing

In this phase, the entire datasets were segmented into two pieces; a training set and a testing set. This consists of random sampling without replacement about 75 percent of the contents representing the training set. While the remaining 25 percent is put into your test set. Essentially, this was done using the following codes:

DOI: 10.31695/IJASRE.2023.9.5.1

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import tree
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor
```

In the code ensuing, train_test_split splits the data and returns a list which contains four NumPy arrays, while train_size = .75 puts 75 percent of the data into a training set and the remaining 25 percent into a testing set.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0, train_size = .75)
```

Concisely, Train-Test split is a model validation process that allows this research to simulate how the developed framework is likely going to perform with new data.

3.5 Word Cloud

Generally, visualizing the documents with different time stamps using conventional word clouds is not an easy task for most text analysts, because the word clouds are typically designed for static documents. Words between two consecutive word clouds usually vary with font sizes and positions. Some words may even appear or disappear frequently over time. Too many of those word variations often distract users. Thus, text analysts may find the words hard to follow and track over time. In this research, word cloud layouts specifically for documents with different time stamps alongside bag of words were employed in order to cluster words based on their regional frequencies. This approach was able to organize the layouts according to different semantic coherence criteria, including a similarity criterion, an importance criterion, as further illustrated in Chapter Four.

3.6 Crime Analysis

As earlier mentioned, potential source of insight exists in unstructured textual data commonly collected by police for the purposes of investigation or administration. Yet, police agencies do not typically have the skills or resources to analyze these data at scale. In this research, the outputs of the word cloud were transformed into regional crime maps depicting the locations, as well as the visual interpretive frequencies showing regions and their related levels of vulnerability to crimes.

3.7 Framework Assessment and Validation

The validation phase was conducted by comparing levels of crime in the intervention and control areas pre- and post-response. This was carried out using simple count data, as well as expert Review provided by the Command. The assessment provided by the command is entirely in harmony with the analysis provided by this research. And as a result, the developed framework is considered valid, as further explained in the next chapter.

4. RESULTS AND DISCUSSION

In the preceding section of this thesis, the research methodology was demonstrated. The various phases employed provide an illustrative idea of the expected outputs from various phases of the proposed framework. In the development of this framework, this chapter is presented structurally to provide details of every segment of the framework, as explained using the ensuing Components in Figure 4.1.

DOI: 10.31695/IJASRE.2023.9.5.1

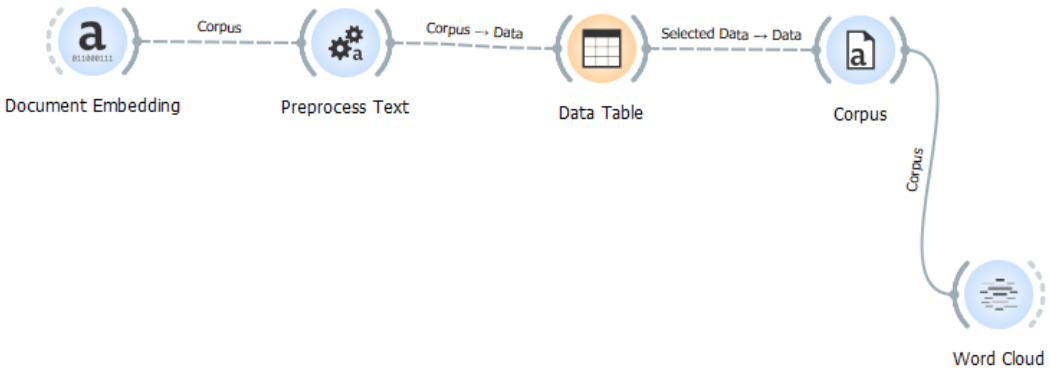


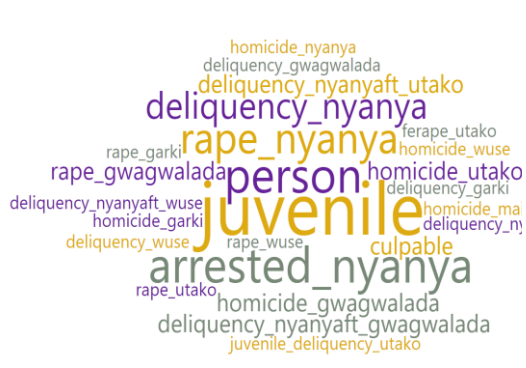
Figure 4.1: Components of the Framework

Illustratively, the digitized unstructured data was embedded in to the framework, which later was pre-processed to provide the required results as shown in Figure 4.2.

Trained Raw Text **Tested Raw Text**

Figure 4.2: Trained and Tested Output

From figures 4.1 and 4.2, the results were obtained after training and testing the raw data. i.e. prior to pre-processing, which apparently provides no insightful details for decision-making. Thus, the need to pre-process the data, which provides the ensuing, results.



Trained **Test**

DOI: 10.31695/IJASRE.2023.9.5.1

Figure 4.3: Output from Pre-processed Data

From the emanating results, the pre-processed data provides an insightful visualization from the structured text. This essentially, reveals regions with their associated crimes, which can be employed for decision-making.

4.1 Crime Frequent Hotspots

The first goal of our study was finding spatial and temporal criminal hotspots. We have successfully achieved this goal using Apriori algorithm on the datasets. We have extracted all the interesting patterns based on our predefined thresholds. We found that the first location has 62 interesting frequent patterns while the second has 59 patterns. Table 4.1 report our Apriori results for crime frequent patterns. The frequent item sets ordered by the location and year of occurrence. With these different frequent item sets, we are able to conclude the most likely crime locations along with their frequent occurrence.

4.2 Crime Frequency

The second target for our study was to predict the crime type that might occur in a specific location within a particular year. The KNN classifier enabled us to reach this target with a reasonable accuracy. To predict an expected crime type, you need to provide at least two related features of the crime. The required features are the occurrence day, month or year of the occurrence and crime location. All features can be submitted in their nominal values.

The location has to be one of its 78 neighbourhoods. Every given result indicates the predicted crime type for a given set of crime features. Table 4.1 gives the corresponding crime type for each location and year.

Table 4.1: Apriori results for the crime frequency pattern

Location & Year / **Crime**	**Culpable Homicide**	**Theft**	**Rape**	**Juvenile Delinquency**
Garki, 2015	6	34	19	16
Garki, 2016	8	13	6	1
Garki, 2017	7	8	4	3
Garki, 2018	3	6	4	1
Garki, 2019	7	4	6	5
Garki, 2020	2	7	11	NIL
Garki, 2021	2	3	2	NIL
Wuse, 2015	3	25	24	8
Wuse, 2016	11	14	15	4.
Wuse, 2017	4	10	6	2
Wuse, 2018	5	4	2	1
Wuse, 2019	11	3	3	NILL
Wuse, 2020	3	2	4	1
Wuse, 2021	1	5	1	NILL

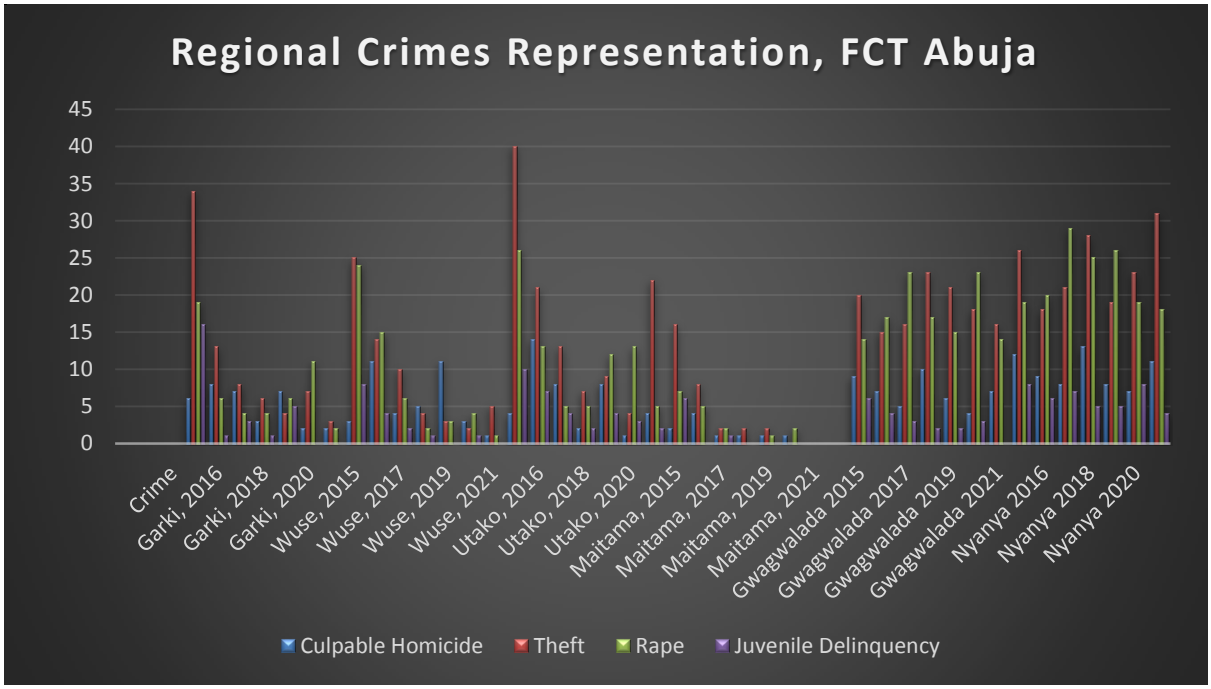
DOI: 10.31695/IJASRE.2023.9.5.1

Utako, 2015	4	40	26	10
Utako, 2016	14	21	13	7
Utako, 2017	8	13	5	4
Utako, 2018	2	7	5	2
Utako, 2019	8	9	12	4
Utako, 2020	1	4	13	3
Utako, 2021	4	22	5	2
Maitama, 2015	2	16	7	6
Maitama, 2016	4	8	5	NILL
Maitama, 2017	1	2	2	1
Maitama, 2018	1	2	NIL	NILL
Maitama, 2019	1	2	1	NILL
Maitama, 2020	1	NIL	2	NILL
Maitama, 2021	NIL	NILL	NILL	NILL
Gwagwalada 2015	9	20	14	6
Gwagwalada 2016	7	15	17	4
Gwagwalada 2017	5	16	23	3
Gwagwalada 2018	10	23	17	2
Gwagwalada 2019	6	21	15	2
Gwagwalada 2020	4	18	23	3
Gwagwalada 2021	7	16	14	NILL
Nyanya 2015	12	26	19	8
Nyanya 2016	9	18	20	6
Nyanya 2017	8	21	29	7
Nyanya 2018	13	28	25	5
Nyanya 2019	8	19	26	5
Nyanya 2020	7	23	19	8
Nyanya 2021	11	31	18	4

DOI: 10.31695/IJASRE.2023.9.5.1

Grand Total of Crime Occurrence 2015 2016 2017 2018 2019 2020 2021	 360 240 190 168 173 149 148			
Grand Total of People Arrested 2015 2016 2017 2018 2019 2020 2021	 432 285 232 200 231 176 217			

Furthermore, figure 4.6 presents the graph of the regional crimes within the study area.



DOI: 10.31695/IJASRE.2023.9.5.1

Figure 4.6: Regional Crime

4.3 Crime Hotspots Demographics Analysis

By studying the demographics data for each neighborhood, we found that dangerous neighborhoods are associated with large population and large number of housing units. Additionally, there is an interesting correlation between the big numbers of vacant houses and the dangerous locations. Moreover, we found that people's age and gender distribution vary between dangerous and safe locations. Specifically, dangerous neighborhoods have more male while the safe neighborhoods have more female. Furthermore, dangerous neighborhoods tend to have bigger number of people with ages from 20 to 29 years whereas the safest neighborhoods seem to include larger number of people with ages from 50 to 59 years. However, we have not found any relationship between crime hotspots and people's race distribution. Ultimately, in other to ascertain the attainment of the underpinning aim and objectives of this study, various tasks conducted in order to attain the objectives of the study are highlighted as thus:

4.4 Research Objectives

To eliminate the less relevant textual data from the crime reports;

To this end, it is pertinent to mention that, eliminate the less relevant textual data from the crime reports in form of data pre-processing was required towards proving a cleaned data, which will in turn enhance the analytical inference. Therefore, this was achieved by removing
their frequencies. This facet essentially provided relevant insight on the regional crimes in Abuja, as elaborated in the preceding chapter.

To form a reliable means of crime visualization

Decision-making requires a reliable means of identifying a trend through visual representation. Therefore, the provision of an appealing visualization results represents a very critical segment in this research. To this effect, an approach based on Word Cloud was utilized in order to visually reveal the frequencies of regional crimes in Abuja to aid decision-making by the authority.

4.5 Finding in Relation to Relevant Studies

Within the scope of some of related studies, various approaches were employed. With relatively reduced level of analytical inference and insight derivable therefrom. And without revealing much details of highly vulnerable regions depicting the frequency, as well as the type of the crime. While this project focused on crime analysis by implementing NPL using unstructured Textual data by considering Culpable homicide, theft, rape and juvenile delinquency, which were not initially considered by any studies. Within the scope of this present study, the consideration of these crimes has revealed the vital roles of employing several regional datasets for crime analysis. This is essentially needed in order to provide a means of identifying crimes and its level of vulnerability regionally.

4.6 Summary

In this chapter, the formulation of the various components for the proposed framework was explained. The theoretical background that requiring the need to pre-process unstructured textual data was equally discussed with illustrative outputs. Furthermore, the relationship between regions and crimes were mapped to present an appealing visualization. In the course of performing various tasks in developing the framework, various findings were made, which are discussed in the ensuing chapter.

ACKNOWLEDGMENT

In the name of Allah, the most gracious the most merciful. First and foremost, I am thankful to the Almighty for giving me the strength, knowledge and the opportunity to undertake this study.
My Late parents Alhaji Dr. Ibrahim A. Coomassie GCON, NPM, MNI and Hajia Baraka Ibrahim Coomassie (May their souls continue to rest in Jannatul Firdaus) without you I would not be who I am.
I would like to thank my respected supervisor Dr. Ahmed Ndanusa for his guidance and professional attitude throughout this research program.
My appreciation also goes to Professor F N Ogwueleka, Head of Department of Computer Science, University of Abuja who worked tirelessly in ensuring maximum success in the completion of these program. I thankfully acknowledge the support of my lecturers especially Dr. Hashim Bisallah for all the encouragement and words of advice during the research period.
The Nigeria Police FCT Command Force CID you gave me the maximum cooperation all through.
I cannot forget my friends, colleagues at work, course mates who supported me and offered deep insight into the study.

DOI: 10.31695/IJASRE.2023.9.5.1

Finally, my husband, children and my siblings for the moral support and encouragement. Thank you all and God bless.

REFERENCES

[1] Ghani, Z. A. (2017). A comparative study of urban crime between Malaysia and Nigeria. *Journal of Urban Management*, *6*(1), 19-29.

[2] Alemika, E., & C. I. C. (2005). Criminal victimization and fear of crime in Lagos Metropolis, Nigeria. *Cleen Foundation Monograph*, *3*(5), 23–31.

[3] Rabia, G. & Farzana, J. (2012). Application of Data Mining Techniques for Analyzing Violent Criminal Behavior by Simulation Model A STUDY ON CONSUMER ATTITUDES TOWARDS SOLUS PER ACQUE IN NATURALS SALON AND SPA View project. In *IRACST-International Journal of Computer Science and Information Technology & Security (IJCSITS)* (Vol. 2, Issue 1).

[4] Agarwal, J. (2013). Crime Analysis using K-Means Clustering. In *International Journal of Computer Applications* (Vol. 83).

[5] Ramachandran, G. (2017). Assault and battery on property. *Loy. LAL Rev*, *12*(3), 12–18.

[6] Mahbub, N. T. (2016). Sustainable Development and its Evolution in the Realm of International Environmental law. *Nnamdi Azikiwe University Journal of International Law and Jurisprudence*, *7*, 1-16.

[7] Peeler Y., & Rivera P., (2020), A descriptive framework for the field of data mining and knowledge discovery, International Journal of Information Technology & Decision Making, 7 (4) (2008) 639–682.

[8] Zhang X., Liu L., Xiao L., & Ji J., (2016) ``Comparison of machine learning algorithms for predicting crime hotspots," IEEE Access, vol. 8, pp. 181302181310, 2020.

[9] Retnowardhani E., & Triana G., (2016), Neighborhood disinvestment, abandonment, and crime dynamics, Journal of Urban Affairs (2014), http://dx.doi.org/10.1111/juaf.12102.

[10] Yu et al. (2017). Developing a Web service-based application for demographic information modeling and analyzing**. DOI:** 10.1109/Agro-Geoinformatics.2017.8047069.

[11] Ahishakiye et al., (2017), Crime Prediction Using Decision Tree (J48) Classification Algorithm.

[12] Chen, H., Chung, W., Qin, Y., Chau, M., Xu, J. J., Wang, G., Zheng, R. & Atabakhsh, H. (2018). *Crime Data Mining: An Overview and Case Studies*. http://ai.bpa.arizona.edu/

[13] Ounacer, S., Ait, H., Bour, E., Oubrahim, Y., Gho umari, M. Y. & Azzouazi, M. (2018). *Using Isolation Forest in anomaly detection: the case of credit card transactions*. *6*(2), 394–400.

[14] Wang, H., Kifer, D., Graif, C. & Li, Z. (2016). Crime rate inference with big data. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 635–644.

[15] Aarathi et al., (2018). Legal Files Management System Using Big Data. DOI: 10.1109/CESYS.2018.8723904.

[16] Kumar, R., & Nagpal, B. (2019). Analysis and prediction of crime patterns using big data. *International Journal of Information Technology*, *11*, 799-805. DOI:10.1007/S41870-018-0260-7

[17] Safat, W., Asghar, S. & Gillani, S. A. (2021). Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques. *IEEE Access*, *9*, 70080–70094.

[18] McCaghy, C. H., Capron, T. A., Jamieson, J. D., & Carey, S. H. H. (2016). *Deviant behaviour: Crime, conflict, and interest groups*. Routledge.

[19] Parsons, M. A., Godøy, Ø., Ledrew, E., de Bruin, T. F., Danis, B., Tomlinson, S. & Carlson, D. (2011). A conceptual framework for managing very diverse data for complex, interdisciplinary science. *Journal of Information Science*, *37*(6), 555–569.

[20] Gupta, M., G. J., A. C. C., & H. J. (2013). Outlier detection for temporal data: A survey. IEEE Transactions on Knowledge and data Engineering, 26(9), 2250-2267. *IEEE Transactions on Knowledge and Data Engineering*, *26*(9), 23–29.

[21] Fadzil, N. H. M., Shahar, S., Rajikan, R., Singh, D. K. A., Ludin, A. F. M., Subramaniam, P., Ibrahim, N., Vanoh, D. & Ali, N. M. (2022). A Scoping Review for Usage of Telerehabilitation among Older Adults with Mild Cognitive Impairment or Cognitive Frailty. In *International Journal of Environmental Research and Public Health* (Vol. 19, Issue 7). MDPI.

[22] Thiprungsri, S. (2010a). *Cluster Analysis for Anomaly Detection in Accounting Data* (Issue 1).

DOI: 10.31695/IJASRE.2023.9.5.1

[23] Nazlena Mohamad Ali, Masnizah Mohd2, Hyowon Lee3, Alan F. Smeaton3, Fabio Crestani4 and Shahrul Azman Mohd Noah2 ,2010.

[24] Nazlena Mohamad Ali, Masnizah Mohd2, Hyowon Lee3, Alan F. Smeaton3, Fabio Crestani4 and Shahrul Azman Mohd Noah2 ,2010 Visual Interactive Malaysia Crime News Retrieval System

[25] Malathi, A. & Santhosh Baboo, S. (2011). Evolving Data Mining Algorithms on the Prevailing Crime Trend-An Intelligent Crime Prediction Model. *International Journal of Scientific & Engineering Research*, *2*(6). http://www.ijser.org

[26] Al-Janabi, K. B. S. (2011). A Proposed Framework for Analyzing Crime Data Set Using Decision Tree and Simple K-Means Mining Algorithms. In *Journal of Kufa for Mathematics and Computer* (Vol. 1, Issue 3).

[27] Thiprungsri, S. (2010b). *Cluster Analysis for Anomaly Detection in Accounting Data* (Issue 1). DOI-10.4192/1577-8517-v11_4

[28] Migueles, J. H., Rowlands, A. V., Huber, F., Sabia, S., & van Hees, V. T. (2019). GGIR: a research community–driven open source R package for generating physical activity and sleep outcomes from multi-day raw accelerometer data. Journal for the Measurement of Physical Behaviour, 2(3), 188-196.

[29] Pandey, Prabhat. & Pandey, M. Mishra. (2013). *Research methodology: tools & techniques*.

[30] Stolterman E. and A. C. Fors (2004). "Information Technology and the Good Life," in Information Systems Research: Relevant Theory and Informed Practice, B. Kaplan et al.(eds), London, UK: Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-8095-6_45

[31] Fatima, A., Nazir, N. & Khan, M. G. (2017). Data Cleaning in Data Warehouse: A Survey of Data Pre-processing Techniques and Tools. *International Journal of Information Technology and Computer Science*, *9*(3), 50–61.

DOI: 10.31695/IJASRE.2023.9.5.1