Tatari Ali Journal,
Volume 2, Issue 1, 2024. Pp. 1 – 15
ISSN - Online: 2006-6427
https://tatarialijournal.com

Received: date: 20th April, 2024
Revised: date: 29th April, 2024
Accepted: date: 8th May, 2024
Published: date: 30th May, 2024

Machine Learning Techniques for Prediction of Metropolitan Cities Crime Rate

Abdulmumini Mohammed[1], Dr. Yusuf Musa Malgwi[2] and Tirmizi Mohammed[3]
[1]Department of Mathematical Sciences, Bauchi State University, Gadau
[2]Department of Computer Science, Modibbo Adama University, Yola
[3]Department of Computer Science Taraba State University, Jalingo
Correspondence email: abduljarmai@gmail.com (08032343624)

ABSTRACT: We all believed that, crime become one of the most predominant aspects in our societies today and is increasing in an alarming rate which effects affects day to day activities of the entire societies. Traditional crime solving methods are in effective in the current situation, therefore, developing model for crime rate prediction by applying machine learning technique K Nearest Neighbor Algorithm would lessen the strain on law enforcement agencies and aid in crime prevention by forecasting areas that might have high crime rate for necessary action to be taken. The model was validated by using new instance to calculate various distances of the training dataset which was ranked based on the shortest distance or nearest neighbor while the system performance was evaluated using confusion matrix which comprises of precision, recall and f-1 score. While the results of this model highlighted that, Doubeli and Yelwa was predicted and having high crime rate in the years 2021 and 2022 respectively. While Zango and Demsawo was predicted and having low crime rate in the years 2020 and 2023 respectively. Therefore, for future prediction, wuro Jabbe was predicted to have high crime rate in the year 2024. Conclusively, this may help concern authority to take necessary actions which will safeguard the life and properties of the people as well as promote peace within the society.
Key words: *Model, Machine Learning, Crime, Prediction and K-NN*

1. Introduction

As crime become most predominant and alarming aspects in our society include house trespass/burglary, Mischief, defilement, rape, homicide, extortion, shila boys activities etc and its cost and effects affects almost everyone to some degree: such as medical cost, property losses, loss of life as well as socio economic development of the entire society. Umamaheswaran, *et al.* (2020) also said Crime is one of the biggest and dominating problems in our society and its prevention is an important task. Daily there are huge numbers of crimes committed frequently. These require keeping track of all the crimes and maintaining a database for same which may be used for future reference. The current problems faced are maintaining of proper dataset of crime and analyzing this data to help law enforcement agencies to detect, predict and solve crimes at much faster rate and thus reduces the crime rate in the society.

Therefore, coming up with a model of Machine Learning techniques such as K Nearest Neighbors (KNN) Algorithm for prediction of metropolitan cities crime rate become necessary to assist the law enforcement agencies within area of study to take necessary actions against

future crime occurrence by predicting area/location which was having higher crime rate as well as areas that have probability for having high crime occurrence.

Machine Learning is a subfield of computer science, but is often also referred to as predictive analytics, or predictive modeling which goal and usage is to build new and /or leverage exiting algorithms to learn from data, in order to build generalizable models that give accurate predictions or to find patterns, particularly with new and unseen similar data. In Machine Learning, artificial knowledge is generated on the basis of experience. In order to enable the software to independently generate solutions, the prior action of people is necessary. For example, the required algorithms and data must be fed into the systems in advance and the respective analysis rules for the recognition of patterns in the data stock must be defined. Once these two steps have been completed, the system can perform the remaining tasks (Umamaheswaran *et al.*, 2020).

Therefore, Machine learning offers methods of regression and classification used to predict rates of crimes. Arthur Samuel, a pioneer in machine learning and artificial intelligence defined machine learning as a field of study that gives computers the ability to learn without being explicitly programmed. In essence, machine learning is a computer system's method of learning by way of examples. Machine learning, or automated learning, is a subcategory of Artificial Intelligence, where the model tries to make a machine to learn from past experiences. In other words, this means turning experience into expertise and knowledge (Deprez *et al.*, 2017).

The machine continues learning as the amount of data increases and adjusts itself iteratively (Alpaydin, 2014). The resurgence of different technologies, such as remote and real-time data collecting, has made it possible to collect vast amounts of data about almost anything. This has enabled the use of machine learning in scenarios that seem completely random and would be thought difficult to predict. Alpaydin (2014) consider consumer behaviour, where people do not go to a grocery store and buy products at random, but they do buy certain items at the same time. These are the kinds of patterns that a machine learning model is trying to find, and which companies then use to extract valuable information.

The aim of this study is to develop a crime rate prediction model using machine learning technique (KNN) Algorithm to predict the crime rate of the areas/locations within Jimeta metropolis, Yola North Local Government Area of Adamawa State Nigeria. Through the following objectives:

Create a dataset to be used by the model
Train the model for prediction using training data set
Validate and evaluate the system performance using the test dataset and confusion matrix respectively

It is expected that, the finding and suggestions arising from this study will provides an insight into criminal activities in the society as well as recommendations for investigation. The model will also help to identify places that most crimes occurred, raise alarm to the authority to take necessary action to prevent future crime occurrence by predicting areas that might have higher probability of future crime occurrence.

There are two types of data that will be utilized in the model building and measuring processes: training data and testing data. The difference between the training data and the test data is that training data is used to build up a model, to fit the data and to find the patterns in the data. On the other hand, the testing data is used to validate the model's accuracy and performance. The major goal is to fit the model so that the features relate to the future observations (prediction) as accurately as possible, or that a better understanding is obtained about the relationship of these two (inference) (James, Witten, Hastie & Tibshirani, 2017).Two of the most common methods of supervised learning are classification methods and regression methods. The difference between these two methods is that regression is trying to predict continuous, infinite numerical outputs, whereas classification methods try to classify the predictions to previously determined output categories (Bhuvaneswari & Sarma, 2013).

Unsupervised learning algorithms, however, aim to find hidden structures in unlabeled class data. In essence, the algorithms learn more about the dataset as it is given more examples to be implemented on. However, humans are capable of learning in multiple ways, which also applies for machines. One of the learning type is called unsupervised learning, where no target variable is given in the learning data. Bekkerman *et al.*, (2012), Alpaydin (2014), Awad and Khanne (2015) describe unsupervised learning as a technique that "groups instances without a pre-specified dependent attribute". Awad & Khanne (2015) one method of unsupervised learning is finding similar groups in the input data, which is defined as a clustering model. For example, customer segmentation is a form of clustering, where the clusters are segments in the customer base and the outliers may imply the existence of a niche market (Alpaydin, 2014). Balaji & Srivatsa (2012) used unsupervised learning method in their paper "Customer Segmentation for Decision Support using Clustering and Association Rule based approaches". Their objective was to find the right segmentation methods to help decision makers to provide the right products to the right customers in the health insurance field. They utilized Waikato Environment for Knowledge analysis workbench, which is a free software for data mining and machine learning (Bhuvaneswari & Sarma, 2013).

K-Means Clustering algorithm is use for unsupervised learning to determine the crime rate. The model was then analyzed, pre-processed and implemented to taste the set of information and trained the algorithm. K-Means Clustering algorithm provided more than 75% (Munasinghe, Perera, Udeshini, & Weerasinghe, 2015) Classification is one kind unique methods of information mining used to classify each object in an information set into one of the predefined

classes or groups, The idea is to define the use of the Criteria for the segmentation of the entire database, once this is done, individual datasets can then fall naturally into one or more groups. By means of classification, existing datasets can be easily understood and it also helps to predict how new individual datasets will behave based on the classification. Data mining generates models of classification by observing classified data and discovering a predictive pattern between those data. Naive Bayes is a classification algorithm used to predict that it works on the principle of Bayesian. Munasinghe, *et al.* (2015) stated that, multi-linear regression is use to find the link between dependent and independent variables. K-Nearest neighbors is used for classification to single and multi-class variable. The Neural Network is used for the prediction's precision.

It is the grouping of a set of data in such a way that data in the same group (cluster) are very similar to one another than the data that are in the other clusters. K-means algorithm: it partitions the data into k number of clusters in which each data observed is assigned to the nearest centroid. The user gives the specified k number of centroids. Each cluster must have a centroid. This process will be repeated until all the data is assigned to a cluster

2.1 Materials and Methods

This research involves design and development of machine learning model which will predict crime rate of the study area. As stated by (Deprez *et al.*, 2017). Machine learning, or automated learning, is a subcategory of Artificial Intelligence, where the model tries to make a machine to learn from past experiences. In other words, this means turning experience into expertise and knowledge.

2.2 Area of the Study

Jimeta is located within and is the headquarters of Yola North Local Government Area (LGA) of Adamawa State, Nigeria. It is situated between latitudes $9^{o}15$"00N and $9^{0}18$'00"N and logitute$12^{0}26$'00"E and $12^{0}30$' 00E with an area of 30.25sqkg and is situated on an altitude 185.9m above sea level. The area is bounded to the North by Girei LGA, to the South by Yola South LGA, to the East by Fufore LGA and to the West by Demsa LGA. The Administrative capitalof the state comprise of 11 wards that include: Ajiya, Alkalawa,Doubeli, Gwadabawa, Karewa, Limawa, Jambutu, Luggere, Nassarawo, Runde and Yelwa wards.

2.3 Data Source

State Security Services as well as Jimeta Division

2.4 Method of Data Collection

Dataset was collected from documents and records of law enforcement agencies within area of study for the period of five (5) years (2020 – 2024)

2.4.1 Justification of Method of Data Collection

Documents and records used because the criminal data are secondary data which were already recorded at the crime departments.

2.5 Used Programming Tools

Javascript/MySQL was used for the development of the model

2.5.1 Justification of the Programming Tools

The software programming language will be used because of its simplicity and consistency, access to great libraries and frameworks for machine learning (ML), flexibility and platform independence.

2.6 The Architectural Model The sample of Architectural model is described in figure 2.1

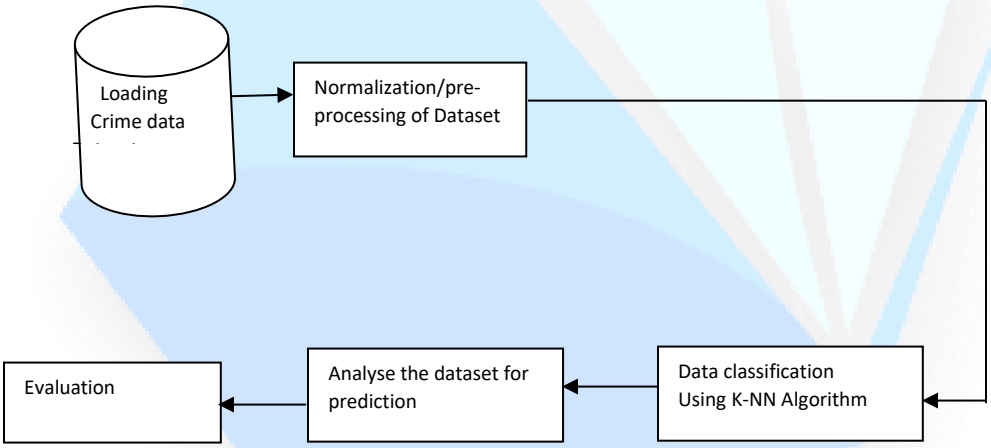

Figure 2.1: Architectural Model

2.6.1 Crime Data

Generally, crimes are classified in to three groups: infractions, felonies, and misdemeanors based on the severity, punishment, and seriousness of crimes. Infractions are minimal crimes such as tailgating, parking overtime, and speeding. Meanwhile, Felonies are considered as most severe crimes followed by misdemeanors which are considered less severe crimes (Goel & Singh, 2018) as shown in Table 2.1 below

Table 2.1: Crimes, Crimes Category and their respective descriptions

Crimes Category	Crimes	Description
Felony	Homicide	The killing of one human being by another.
	Rape	Unlawful sexual intercourse.
	Robbery	The attempt to take anything of value from the custody, control, or care of a person by forcing or threatening by force by putting a victim in fear.
Misdemeanor	Theft	The act of stilling.
	Trespass	Unlawful intrusion
Infraction	Public Nuisance	An action or omission constituting an instruction to the peace and comfort of a section of the public

K- Nearest Neighbor, popular as **K-Nearest Neighbor (KNN)**, is an algorithm that helps to assess the properties of a new variable with the help of the properties of existing variables. **KNN** is applicable in classification as well as regression predictive problems. **KNN** is a simple non-parametric test. It does not involve any internal modelling and does not require data points to have certain properties. It simply takes the voting of the majority of variables and accordingly treats new variables (Sharma & Chetta, 2018). The crime data input should be from the dataset sample as in Table 2.2 below

With the following attributes: Age, Gender, Crimes, Location/Area as well as the Year

Table 2.2: Dataset sample

Age	Gender	Homicide	Public Nuisance	Theft	Rape	Arm Robbery	Trespass	Location	Year
27	1	0	1	0	0	0	1	29	2020
30	1	0	1	0	0	0	0	21	2021
22	0	0	0	0	1	1	0	12	2020
24	1	0	0	1	0	1	0	26	2023
28	1	0	1	0	0	0	1	8	2020
20	0	1	0	0	0	1	0	3	2023
35	0	1	0	1	0	0	0	6	2022
20	1	0	0	0	0	1	1	13	2021
28	0	0	1	0	1	0	0	25	2020
42	1	1	0	1	0	0	1	16	2021

From Table 2.2, the attribute Gender have two fields 1s and 0s which represent males and female respectively while for the crimes 1s and 0s represents crime occurrence and non-occurrence respectively likewise location contained numerals which represents various areas/locations.

Table 2.3: Locations/Areas with their respective descriptions

Locations ID	Descriptions
29	Bachure
21	Damsawo
12	Jambutu
26	Luggere
8	Zango
3	Nasawo
10	Wuro Jabbe
6	Yelwa
13	Shinko
25	Karewa
16	Doubeli

2.6.2 Data Normalization/Pre-processing

Normalization is a data preparation technique that is frequently used in machine learning. The process of transforming the columns in a dataset to the same scale is referred to as normalization. Normalization is an essential step in data pre-processing in any machine learning application and model fitting which improve the model accuracy while data pre-processing is an integral step in Machine Learning as the quality of data and the useful information that can be derived from which can directly affects the ability of model to learn. Normalization is a data preparation technique that is frequently used in machine learning. The process of transforming the columns in a dataset to the same scale is referred to as normalization.

2.6.3 K-NN Algorithm for Data Classification

K Nearest Neigbors Algorithm is non-parametric, supervised learning classifiers which use proximity to make classification or predictions about the grouping of an individual data point. In KNN Classifier, a new data point is classified based on its proximity to the K nearest neighbors in the training set. The proximity is measured using a distance metric, such as Euclidean distance or Manhattan distance. In this research we used Euclidean Distance Formula as: $Duv = \sqrt{(U1 - V1)\text{^}2 + (\text{U2} - \text{V2})\text{^}2 + \cdots + (Uq - Vq)\text{^}2}$

Where; *U*- Represent the value of testing instances

V- Represent the value of training instances

D1 =√(46-27)^2+(1-1)^2+(0-0)^2+(1-1)^2+(1-0)^2+(0-0)^2+((0-0)^2(1-1)^2+(10-29)^2

=√273

= 26.9

D2 =√(46-30)^2+(1-1)^2+(0-0)^2+(1-1)^2+(1-0)^2+(0-0)^2+(0-0)^2+(1-0)^2+(10-21)^2

=√379

=19.5

D3 =√(46-22)^2+(1-0)^2+(0-0)^2+(1-0)^2+(1-0)^2+(0-1)^2(0-1)^2+(1-0)^2+(10-12)^2
=√586
=24.2
D4 =√(46-24)^2+(1-1)^2+(0-0)^2+(1-0)^2+(1-1)^2+(0-0)^2+0-1)^2+(1-0)^2+(10-26)^2
=√743
27.3
D5 =√(46-28)^2+(1-1)^2+(0-0)^2+(1-1)^2+(1-0)^2+(0-0)^2+(0-0)^2+(1-1)^2+(10-8)^2
=√329
=18.1
D6 =√(46-20)^2+(1-0)^2+(0-1)^2+(1-0)^2+(1-0)^2+(0-0)^2+(0-1)^2+(1-0)^2+(10-3)^2
=√731
=27.0
D7 =√(46-35)^2+(1-0)^2+(0-1)^2+(1-0)^2+(1-1)^2+(0-0)^2+(0-0)^2+(1-0)^2+(10-6)^2
=√161
=12.7
D8 =√(46-20)^2+(1-1)^2+(0-0)^2+(1-0)^2+(1-0)^2+(0-0)^2+(0-10^2+(1-1)^2+(10-13)^2
=√688
=26.2
D9 =√(46-28)^2+(1-0)^2+(0-0)^2+(1-1)^2+(1-0)^2+(0-1)^2+(0-0)^2+(1-0)^2+(10-25)^2
=√553
=23.5
D10 =√(46-42)^2+(1-1)^2+(0-1)^2+(1-0)^2+(1-1)^2+(0-0)^2+(0-0)^2+(1-1)2+(10-16)^2
=√54
=7.3
The Euclidean distances ranks results shows in Table 3.1

2.6.4 KNN Data Analysis

K NN Algorithm is analysing the data by finding the K nearest points in the training dataset and uses their class to predict the class or value of a new data point. In this research we achieved this by assigning value for K as 1, 3 and check the ranking to find out the class of crime of the new instance which can either be high or low. Table 2.1

Step 1: Determine the value of K while k is the number of nearest neighbors.

Step 2: Calculate the distance between testing instances and all training instances (Sample) using Euclidean distance as shown in Table 2.1.

Step 3: We assigned value for K as 1, 3 which determined the ranking class of crime rate of the new instance as high or low

If K is 1, the class of crime rate is high as shown by distance ranking in Table 2.1 the model predicted that in the year 2021 location number 16 (Doubeli) was having a High crime rate and

crimes occurred were house trespass, theft and homicide by male of 42 years of age If K is 3, the class of crime rate is low in the year 2020 the location number 8 (Zango) was having a low crime rate and type of crimes mostly occurred was public nuisance and house trespass by males of 28 year of age. Table 2.1

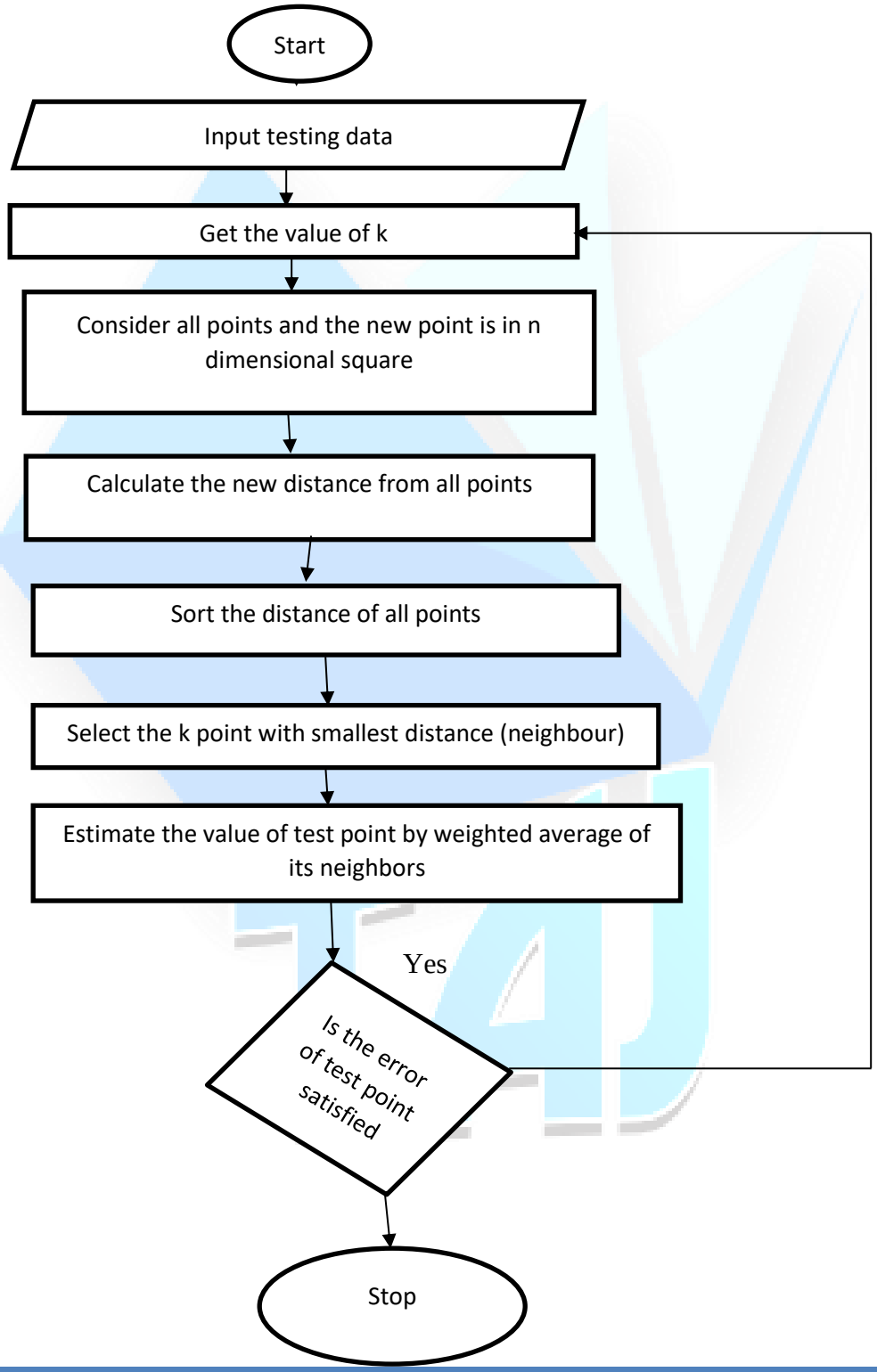

Fig 2.2: Flowchart for K NN Algorithm (www.researchgate.net/publication)

2.6.5 *Training/Testing Dataset*

Dataset is divided into two based on percentage for the training and testing of the proposed model (70% for training and 30% for testing)

During the training phase, the KNN algorithm stores the entire training dataset as a reference. When making predictions, it calculates the distance between the input data point and all the training samples, using Euclidean distance.

2.6.6 System Validation

Testing dataset/confusion matrix used for the model performance evaluation

K NN Euclidean formula was used for testing dataset to find out the shortest distance to the nearest neighbour (new instance) as shown in Table 2.1

To evaluate the performance of the model confusion matrix and test data form Table. 2.3 which shows the K nearest neighbour with shortest distance with their respective ranks.

2.6.7 System Performance Evaluation

To evaluate the performance of this classifier we used F-1 score which is basically used to compare the performance of precision and recall results.

$$\text{F-1 Score} = 2*\frac{\text{Precision}*\text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.3)$$

The computation of precision and recall was used to measure the performance of three classes of the dataset as the value of was chose to be 3 which was ranked based on their nearest neighbors.

The formula for Precision and Recall are as follows:

$$\text{Precision} = \frac{TPi}{TPi\text{-}FPi} \quad (2.4)$$

$$\text{Recall} = \frac{TPi}{TPi\text{-}FNi} \quad (2.5)$$

Where TPi is the nuber of true positive for class i, FPi is the number of false negative for class I and FNi in the number of false negative for class i.

The performance measure can be derived from the confusion matrix which represented by predicted and actual classification in form of n*n where n is a number of classes. The results of prediction ranking based on k Value is shown in Table 2.2.

2.8 The Comparison of Precision and Recall Results

The performance results of the model based on the confusion matrix (precision and recall) as well as F-1 score are shown in Table 2.3 and Table 2.4 respectively

3.1 Results

Table 3.1: Training and Testing Dataset with Distances and Ranks Results

Case	Age	Gender	Homicide	Public Nuisance	Theft	Rape	Arm Robbery	Trespass	Location	Year	CrimeRate	Distance	Rank
1	27	1	0	1	0	0	0	1	29	2020	Low	26.9	8
2	30	1	0	1	0	0	0	0	21	2021	Low	19.5	4
3	22	0	0	0	0	1	1	0	12	2020	High	24.2	6
4	24	1	0	0	1	0	1	0	26	2023	High	27.3	10
5	28	1	0	1	0	0	0	1	8	2020	Low	18.1	3
6	20	0	1	0	1	0	1	0	3	2023	High	27.0	9
7	35	0	1	0	1	0	0	0	6	2022	High	12.7	2
8	20	1	0	0	0	0	0	1	13	2021	Low	26.2	7
9	28	0	0	1	0	1	0	0	25	2020	High	23.5	5
1	42	1	1	0	1	1	0	1	16	2021	High	7.3	1
			New Instance										
	46	1	0	1	1	0	0	1	10	2024			

Table 3.2: Prediction Ranking Results based on K value

Classes	**Distances**	**K=3**	**Areas**	**Total**
Class A	7.1	1	8	16.1
Class B	12.7	2	6	20.7
Class C	18.1	3	16	37.1
Total	37.9	6	30	73.9

Table 3.3: Comparison of Precision and Recall Results

Predicted Area	**Precision**	**Recall**
Doubeli	0.19	0.44
Yelwa	0.33	0.10
Zango	0.53	0.23

Table 3.4: F-1 Score Comparison Results

Predicted Area	F-1 Score
Doubeli	0.27
Yelwa	0.16
Zango	0.32

The results of this research was obtained by using Machine Learning Technique called K Nearest Neighbor (KNN) Algorithm as shown in Table 3.1

Table 3.5: The Model Results

Rank	Distance	Location/Area	Offender's Age	Offender's Gender	Crime Committed	Crime Rate	Year
1	7.3	Doubeli	42	Male	Homicide, Theft and House Trespass	High	2021
2	12.7	Yelwa	35	Male	Theft and Rape	High	2022
3	18.1	Zango	28	Male	Public Nuisance and House Trespass	Low	2020
4	19.5	Demsawo	30	Female	Public Nuisance	Low	2023

Table 3.6: Future Prediction sample results

	Wuro Jabbe	46	Male	Homicide, Theft and House Trespass	Predicted Crime Rate High	Year of prediction 2024

4. Discussion of the Results

The result of this research was obtained based on K Nearest Neighbor (KNN) Algorithm which shows that, between the year 2020 - 2023 the locations within the area of study which have high/low crime rates based on their distance ranking and the different types of crimes were committed with the ages and gender of the offenders. For the future prediction, it also shows an

area that may have high crime rate in the year 2024 and types of crimes that may occur within the area as shown in Table 3.5.

The model was validated by using new instance to calculate various distances of the training dataset which was ranked based on the shortest distance or nearest neighbor as presented in Table 4.1 while the system performance was evaluated using confusion matrix which comprises of precision, recall and f-1 score as represented in Table 3.3 and 3.4 respectively.

The model results shows that in the year 2020 the location/area known as Zango ranked as No 3 has low Crime rate and the most crimes occurred were Public Nuisance and House Trespass followed by Doubeli in the 2021 with crime rate which ranked as No 1 with the shortest distance of 7.3 and Yelwa in the year 2022 with also high crime rate which ranked as No 2 with the shortest distance of 12.7 as shown in Table 3.5.

For future Prediction, Wuro Jabbe was predicted and it will high crime rate and the type of crime to be committed are Homicide, Theft and House Trespass in the Year 2024 as shown in Table 3.6. This kind of information will assist the law enforcement agencies to take necessary action against future crime occurrence. In respect to the prior studies the following gaps were found (Mahmud, Nuha, & Sattar, 2021), The accuracy of their model only considered single area, (Mahendra, Nani, Balu, Avinash, & Aditya 2020) Their work only considered crime rate based on robbery and (Anisha, Dhanashree, Arpita, & Divya 2016) did not considered types of crime, and its was taken care of in this research by considering different areas as well as different types of crimes which were occurred most within the area of the study.

5. Conclusion

Crime is a deliberate act that can cause physical or psychological harm, as well as property damage or loss, and can lead to punishment by a state or other authority according to the severity of the crime. This research was conducted with the aim of developing a machine learning model using K Nearest Neigbors (KNN) algorithm which predicts the crime rate of different areas/locations within Jimeta, Yola North Local Government Area of Adamawa State which may assist the law enforcement agencies to know the Areas/locations which have high/low crime rate in order to take necessary actions by relevant authorities before future crimes occurrence. For achieving this secondary data was collected from state security services (SSS) as well as Jimeta Police Division of Yola North Local Government Area, Adamawa State which formed the sample dataset used for the Training/Testing of K NN Algorithm model which performed the predictions.

The model was validated by using new instance to calculate various distances of the training dataset which was ranked based on the shortest distance or nearest neighbor while the system performance was evaluated using confusion matrix which comprises of precision, recall and f-

1 score. In this research, the application of machine learning algorithm K- Nearest Neighbor (KNN) was employed and performed the crime rate predictions of different areas within the area of the study. Conclusively, the research work considered certain type of crimes with other attribute such as: Age and Gender of perpetrators, locations/areas where crime was committed as well as years. The results show the areas/locations which have high/low crime rate within the area of study. It also indicates areas which may have high/low future crimes occurrence and help the relevant authority within the area of the study to take necessary actions at the appropriate time.

Therefore, this study is focused on crime rate prediction of areas/location within Jimeta Metropolis, Yola North Local Government Area of Adamawa State. For the period of five (5) years, from 2020-2024. Based on Machine learning technique (K-Nearest Neighbors Algorithm.) The applicability of learning algorithm (technique) in this research is limited to the development of the model and the functionality of the model is limited to predict the crime rate of the study area. Conclusively, I recommend that as the number and different types of crime is increasing in our societies on daily bases it become necessary to developed and used this kind of model to assist the law enforcement agencies to take necessary action against future crime occurrence which will assist to save the peoples life, properties and promote economic development of the society. For further research more input variables such as Time of crime occurrence can be added as an attribute and any other machine learning algorithm than K NN Algorithm can be used to enhance the performance of the model.

References:

Alpaydin, A E. (2014). *Introduction to machine learning. adaptive computation and machine learning*. third edition Themit press.

Allott, A., N., Thomas, D., A., Clarke, D., C., Edge, I., D., & Bernard, T., J. (2023)"crime". *Encyclopedia Britannica*, https://www.britannica.com/topic/crime-law.

Anisha, A., Dhanashree, C., Arpita, A., & Divya, C. (2016). Application for analysis and prediction of crime data. *International Journal of Advanced Computational Engineering and Networking, 4* (5).

Awad, M., & Khanna, R. (2015). *Efficient learning machines*: *theories, concepts, and applications for engineers and system designers*. 1st Edition Apress Berkeley CA https://doi.org/10.1007/978-1-4302-5990-9

Balaji, S., & Srivatsa, S. (2012). Customer segmentation for decision support using clustering and association rule based approaches. *International journal of computer science & engineering technology (ijcset).* Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.403.6935&rep=rep1&type=pdf

Bekkerman, R., Bilenko, M., & Langford, J. (2012). *Scaling up machine learning*. 1st edition. Cambridge university press new york, ny 10013-2473 usa

Bhuvaneswari, E., & Sarma, V. (2013). The study and analysis of classification algorithm for animal kingdom dataset. *Information engineering*, **2**(1), 6-13.

Dakalbab, F., Abu Talib, M., Elmutasim, O., Bou Nassif, A., Abbas, S. & Nasir, Q. (2022) Artificial intelligence & crime prediction: *a systematic literature review. Soc Sci Humanit* 6:100342

Deprez, P., Shevchenko, P., & Wuthrich, M. (2017). Machine learning techniques for mortality modeling. *european actuarial journal*, **7**, 337–520. Retrieved december 10, 2019 from https://researchers.mq.edu.au/en/publications/machine-learning-techniques for-mortality-modeling

Goel, A. & Singh, B. (2018). White collar crimes: *A study in the context of classification, causation and preventive measures*. Contemporary Social Science.27:84–92.

Hetal, B., & Amit, G. (2012). A comparative study of training algorithms for supervised machine learning. *international journal of soft computing and engineering (ijsce)*, **2**(4), 74-81.

James, G., witten, D., Hastie, T., & Tibshirani, R. (2017). *Introduction to statistical learning: with applications in r.* 8th edition. New york, ny, usa .

Mahendra, C., Nani, G., Balu, G., Avinash, A., & Aditya, Y. (2020). Crime Rate Prediction. *Journal of Engineering Science, 11* (5), 692-697.

Mahmud, S., Nuha, M., & Sattar, A. (2021). *Crime Rate Prediction Using Machine Learning and Data Mining*. Department of Computer Science and Engineering, Daffodil International University, Bangladesh

Munasinghe, M., Perera, H., Udeshini, S., & Weerasinghe, R. (2015). *Machine learning based criminal short listing using modus operandi features* https://doi.org/10.1109/icter.2015.7377669

Neil, S., Nandish, B. & Manan,S (2021) Crime Forecasting: *A Machine Learning and Computer Vision approach to crime prediction and prevention* 4(1) DOI:10.1186/s42492-021-00075-z

Umamaheswaran, R., Radha, R. & Preethi A. (2020). Crime Rate Prediction using KNN *International Research Journal of Engineering and Technology (IRJET)* 7(5) www.irjet.net

Sharma, P. & Chetty, P. (2018). *How to use K-Nearest Neighbor (KNN) algorithm on a dataset?*. [online] Project Guru. Available at: https://www.projectguru.in/k-nearest-neighbor-knn-algorithm/ [Accessed 28 Jul. 2023].