

## **A Study Of Crime Analysis : A Systematic Review**

**Aby Rose Varghese<sup>1</sup>, Dr. Angeline Prasanna G<sup>2</sup>**

<sup>1</sup>*Research Scholar, Karpagam Academy of Higher Education*

<sup>2</sup>*Associate Professor, Karpagam Academy of Higher Education, Coimbatore*

### **ABSTRACT**

Nowadays, certain crimes occur within a society, and criminality is increasing on a daily basis. There has also been a significant increase in criminal activity. Society has been negatively impacted by crime. Controlling crime is essential for government support, social stability, and growth. Regulation implementation organizations are searching for a framework that will allow them to successfully objective crime structure. The intelligent crime data analysis provides sufficient knowledge about the elements of illegal activities, running through examples of criminal behavior while aiming to be advantageous in catching where, when running through examples of criminal behavior while aiming to be advantageous in and why infractions may occur. It is in need of data storage, analysis, and algorithm which can manage a large amount of data and produce high accuracy. This paper summarizes the many machine learning approaches utilized during criminal analysis that should be eliminated in order to achieve the best result.

**Keywords: Crime Data Analysis; Machine Learning; Methodology; Law enforcement**

### **Introduction**

One of the difficult issues in our general public is crime. It is the most overwhelming part of our general public. It is likewise prevalent in the public eye. Along these lines, the counteraction of crime is one of the significant undertakings. The crime investigation should be in precise manner. As analysis makes it significant in the detection and prevention of crime. The analysis distinguishes the investigating patterns and help in the detection of patterns in crime. A crime is only an activity that establishes an offense. It is a punishable offense. The distinguishing proof and analysis of hidden crime is truly a challenging task for the police officials. Additionally there is voluminous information of the crime is accessible. Here we analyze a few methodologies that should help in investigation. Along these lines, the technique should assist with settling the crime. The machine learning approaches can all the more likely assistance in prediction and analysis of the crime. Crime is one of the greatest and using issues in our general public and its counteraction is a significant task. Every day there are enormous quantities of violations submitted often times. This requires monitoring every one of the wrongdoing and keeping a dataset for same which might be utilized for future reference. In these papers we compare various machine learning methods used in crime analysis

### **1. Machine Learning Algorithms to Analyze Crime Data**

-Lawrence McCendon and Natarajan[1]

In this paper they have done a comparative study between the violent crime pattern from the communities and crime normalized dataset. Here they used the Linear Regression, Additive Regression and Decision stump algorithm on the communities and crime dataset. Five matrices are used to evaluate the effectiveness and efficiency of the effectiveness and efficiency of the algorithms like Correlation coefficient, Mean absolute error, Root Mean Square Error, Root relative squared, Relative Absolute Error etc. For comparative evaluation of the crime statistics these matrices can be used. This research objective is to present the effectiveness of the algorithms that can be in determining patterns of criminal activities.

Linear Regression, Additive Regression, and	MODEL(BEST RESULT)	CORRELATION COEFFICIENT	MEAN ABSOLUTE	ROOT MEAN SQUARE	RELATIVE ABSOLUTE ERROR	ROOT RELATIVE SQUARED
MURDER	Linear Regression Model	.99	3.0	6.4	26%	11%
MURDER PER 100 K POPULATION	Addictive Regression Model	0.88	2.6	4.4	41%	48%
Result for Rape	Linear Regression Model	0.98	7.6	17.3	22%	16%
RAPE PER 100 K POPULATION	Addictive Regression Model	.83	13.2	19.2	52%	56%
Result for Robberies	Linear Regression Model	0.99	53.4	109	15%%	5%
Robberies PER 100 K POPULATION	Linear Regression Model	.95	47	69.6	31	30
Result for Assaults	Linear Regression Model	0.99	107	224	25%	11%
Assault PER 100 K POPULATION	Linear Regression Model	93	92.8	163	30	37

Table 1

Based on the error values the overall accuracy of the algorithms is evaluated Among three algorithm the greatest correlation coefficient value also generated lowest error values. The least accurate is decision stump model for each of the nine goal feature. The addictive regression model is the most Prominent for all of the crime per 100K population features except for the Robberies Per Pop, Assault Per pop and violent crime per pop. In this paper a comparison has been done between the model produced by WEKA and the crime statistical data provided by neighborhoodscout.com WEKA provides attribute information for the features in predicting the crime data based on the training set input for the three algorithms the linear regression algorithm to be very effective and accurate.

## 2. Crime Prediction Using Machine Learning

Prof. Shivaprasad More, Sakshi Mench, Saloni Kuge , Hafsa Bagwan[2]

For the detection of crime patterns from inferences a model is designed in this paper. The data are collected from the crime scene and this paper demonstrates the prediction of the perpetrator. This paper analyses the way for the prediction of perpetrator age and gender. Major aspects of crime prediction are given in this paper. One is convict gender and other is convict age. Different parameters used here are the year, month and weapon used in the unsolved crimes. The number of unsolved crimes are identified in the analysis part and prediction part involves the description of the convict's age, sex and relationship with the victim. The prediction is done using multi liner regression, K-Neighbor's classifier and neural networks. The dataset used here contains almost 63000 values.

The dataset is from the Kaggle website. The dataset contain entries from 1980 to 2014. The number of unsolved crimes, the weapons used in the crimes, the month when the maximum crime took place, the place of occurrence of the crime, the state where the crime rate is high etc. are analyzed. Multilinear regression and the other is K- neighbors classifier are used for prediction

### a) MultiLinear Regression

Find the relationship between the dependent variable with the given set of independent variables this algorithm gives the mathematical approach. Here perpetrator's age is a dependent variable, and piece of evidence collected from the crime scene are the independent variable. Perpetrator's age can be predicted based on input features such as state, year, month, place and crime solved etc.

### b)K-Neighbors Classification

The target variable has more than two classes this algorithm can be used. In this dataset, the target variable isn't anything, however its culprit sex and it has named male, female and obscure. The target variable relationship has 27 unique values such as wife, friend, nephew etc. To classify these target variables K-Neighbors classier is used.

Perpetrator sex and relationship are the target variable. The dataset is divided into training and 30% for testing. This model gives 85% accuracy. The previous model gives 65% accuracy. The regression and classifier give almost 80% of accuracy. Using this model any type of crime can be predicted and dataset of different countries can be used for analysis

Model	Accuracy
K-Neighbors Classification	85%
MultiLinear Regression	65%

Table 2

## 3.Crime Prediction and Analysis Using Machine learning-Alkesh Bharatil, Dr Saravanaguru RA[3]

Various types of crimes are analyzed in this paper. Depending upon various conditions the type of crime which may happen in future can be predicted. Using Chicago crime dataset, they will be using the technique of machine learning and data science for crime prediction .The crime data is taken from the official portal of Chicago police for crime prediction. It contains information like crime location, crime type, date and time, latitude, longitude. Data preprocessing, feature selection and scaling is done before the training of the model. The K-Nearest Neighbor (KNN) classification and various other algorithms for crime prediction and one with better accuracy will be used for training. Graphical representation is used for the visualization of dataset like, at which time the criminal rates are high or at which month the criminal activities are high. The main purpose of this project is to give an idea of

how machine learning can be used by the law enforcement agencies to detect, predict and solve crime at a much faster rate and thus reduces the crime rate. Data set is divided as training set and testing set the model is trained using different algorithm as shown in the table The accuracy is listed in the table3.

ALGORITHM	ACCURACY
K-Nearest Neighbor	0.787
Gaussian NB	0.646
Multinomial NB	0.456
Bernoulli NB	0.313
SVC	0.313
Decision Tree Classifier	0.786

Table 3

As we can see from the table the KNN algorithm with accuracy of 0.787 has the highest accuracy compared with the rest of the algorithm. It can be used for predictive modeling. SVM has the low precision .Analysis can be done on the dataset and can also be plot them on different chart like bar, pie, scatter etc.

Various analysis done are listed below

- ✧ Date and time of the crime committed
- ✧ Number of crime in the city of Chicago
- ✧ Arrested ratio.
- ✧ Crimes location.
- ✧ Major crimes committed in the city.

They produced many graphs and found interesting statistics that can be used for keeping society safe.

#### 4.”Crime Prediction Using Machine Learning”-Riya Rahul Shah[4]

Analysis and prediction of crime using machine learning models is the objective of this paper. It centers on creating a model. These models help to detect the number of crime, type of crime in particular state etc. In this project different machine learning models like K-NN, Boosted decision tree etc. are used for prediction. By analyzing the pattern of crime region wise geographical analysis can be done. To help the law enforcement agencies different visualization technique and plots are used. It helps to detect and predict higher accuracy crime. In such a way the crime rate can be reduced and improve the security in required areas. The dataset utilized is Vancouver crime dataset which is available on Kaggle[5]. Data analysis is also done to understand the dataset and the number of incidents that have happened previously based on the kinds of crime. It is seen that most crimes are theft from vehicle. Pattern of crime in every year is also noticed. It is observed that in 2003 crime rate was highest in Vancouver city. The rate is much diminished till 2017. So it is a decent sign that assuming we study more we can decrease the rate more.

model	Accuracy	Precision	recall	F1 score
K-Nearest Neighbor	0.317	.2532	.3170	.32
Logistic	0.3608	0.3608	0.3607	0.36

---

Regression				
Multinomial Naïve Bayes	0.360705	0.36076	0.3607	0.36
Random Forest classification	0.3060	0.310	0.306	0.31

Table 4

### 5. Crime Data Analysis: Review-Deepika Tyagi Dr.Sanjiv Sharma[6]

The data mining technologies utilized in criminal investigations are analyzed in this paper. The commitment of this paper is to feature the methodology used in crime data analytics. It discusses the challenges emerging during the analysis. These challenges can be eliminated to obtain desired result. CRISP-DM methodology represents cross industry standard process for data mining. It comprises of business understanding, data understanding data preparation, modeling and evaluation. Dataset are collected from various sources. Comprehend the information accessible in the crime record on which the handling task is to be performed to recognize criminal patters, type and crime zone and the hot spots of future criminal activities so that crime rate can be reduced. As the gathered data are in many formats so data preprocessing is performed to improve the quality of dataset and can improve the accuracy. The model is to be intended to perform the different activities like feature selection, clustering, analysis and prediction. The result is plotted by graph using visualization tools.

In this comparison study SVM is used for identifying digital evidence related to cybercrime. ANN provides better accuracy than logistic regression. The SVM gives high accuracy as compared to multilayer perceptron neural network (ANN). Decision tree and logistic regression are used for identifying false statement of different types of crimes. Data mining techniques like Nearest Neighbor, Decision tree, Support Vector Machine (SVM), Naive Bayes and Neural network are utilized for auto insurance fraud. Neural network perform better as compared to decision tree and SVM. There are various challenges arises in this study criminal data are available in various formats from multiple data sources and transforming the data into a desirable form so that result can be obtained, is also a challenging task. There is another challenging aspect is of storing large amount of data size, so there need a large capacity storage device. Selecting a particular model is also a challenging task. Some other challenges such as matching data mining technique and methodology and exploring proper integration methods to tackle complicated investigation problem etc.

### 6. Machine learning and Data Mining For Crime Prediction -Sakib Mahmud,Musfifika Nuha and Abdus Sattar[7]

Different clustering approaches of data Ming is used here to analyses the crime rate of Bangladesh. To train the dataset KNN algorithm is used. Algorithm is used to analyze the data and to predict the rate of crime. Finally they use the forecast rate to find out safe route. This helps the individual to aware of the crime location and the safe way to the destination. Finally, they use the forecast rate, to find out safe route. This will help individuals to be aware of the crime location and find their safe way to the destination.

This paper gives awareness about the future crime based on algorithms. From the crime dataset we find out the crime rate in different location based on age, gender, location of the criminals. The data

source and method used for the prediction include various types of crime statistics, survey, literature review and statistical model that explorate crime trends into the future algorithm models that describe the behavior of observe past values. It can be used to predict future crime trends by time series analysis of crime trends into the future. The predictive model undertaking to show a relationship between certain predictor and dependent variable. To guarantee the greater accuracy those models must identify and predict the scope and nature of a number of factors that will impact crime and victimization in the future. Crime rate prediction is much more specific and accurate in this paper. Below table listed different algorithm accuracy. By referring these three algorithms, K-nearest neighbor provides high accuracy

Algorithm	Accuracy
Linear	73.61
Naïve Bayes	69.50
KNN	76.92

Table 5

### 7. Machine Learning Algorithm for the reduction of crime rate in India -Sharaddha Ramdas Bandekar, C Vijayalakshmi[8]

To interpret data and analyze ways to reduce crime rate this paper used K-Mean clustering, density-based clustering, coweb clustering and filtered clustering techniques. Bayesian neural network, Liebenberg Marquardt and scaled algorithm are the other techniques used here for crime prediction. Statistical analysis using ANOVA is also used in their method. Data Ming and machine learning are related field involving computers and mathematics. Where in the programming is done for the system to carry out the operation. Both can be used for the detection and prevention of crime. Crime analysis includes filtering of crime patterns, prediction and detection of crime. Building a model which help in prediction. For this they use decision tree and classification, support vector machine, KNN classification, regression, K -Means clustering, density based clustering and random forest. Decision tree is depend on models of utility and result for display of algorithm. KNN stores instances connecting with training dataset in multi-dimensional space. In support vector machine using hyper plane to classify data above and below providing higher level of prediction based on clustering. Random forest groups various classifications, regression and other operations giving array of decision trees at training time. K means clustering grouping N number of observations into k number of clusters in which each of the observations is a part of cluster with closest mean. Density based clustering is an unsupervised ML algorithm and clusters are dense sports in the data space. Regression is used for prediction wherein one of the variables is observed when other set values are given. This research works on the various crime types, occurrence of crime in different location and time. It is observed that compared with the rest of the crime, murder as the dominating crime. Using the Bayesian, Levenberg and scaled algorithm on trim and test data, and it was seen that scaled algorithm gave the best result in comparison with the other two. The analysis acquired was the crime rate can be decreased to 78% with accuracy of 0.78

### CONCLUSION

To predict the crime which can occur in future can help the law enforcement officers to reduce or prevent the crime before it occurs. This paper analyses various machine learning methods to reduce the crime and detect the technique of crime and stop the crime. .From this we can summaries that in general situation KNN provides the appreciated accuracy than the other methods so that most of them



use KNN for crime prediction. Sometimes it may vary according to the type of crime. This paper centered to help researchers intending to make crime prediction a reality and implement such advanced technology in real life.

## REFERENCE

- [1] Machine Learning and Applications: An international Journal (MLAIJ) Vol.2, No.1, March 2015. Lawrence McClendon and Natarajan Meghanathan Jackson, MS, USA
- [2] Crime Prediction Using Machine Learning Approach Prof. Shivaprasad More, Sakshi Mench, Saloni Kuge, Hafsa Bagwan. International Journal of Advanced Research in Computer and Communication Engineering. Vol.10, Issue 5, May 2021. Vol. 10, Issue 5, May 2021
- [3] Crime Prediction and Analysis Using Machine Learning.-Alkesh Bharatil, Dr. Saravanaguru RA. K2. International Research Journal Of Engineering and technology Volume; 05, Issue.) 9 sep 2018
- [4] Riya Rahul Shah have published a paper entitled as “Crime prediction Using Machine Learning” 2003.
- [5] Kaggle Inc, Data of crime in Vancouver (Canada) from 2003 to 2017 [6] International Journal of Engineering Technologies and Management Research 5(2):67-74  
Deepika Tyagi, Dr. Sanjiv Sharma
- [7] Crime Rate Prediction Using Machine Learning and Data Mining: Sakib Mahmud, Musfifika Nuha and Abdus Sattar 2020
- [8] Design and Analysis of Machine Learning Algorithms for the reduction of crime rate in India: Shraddha Ramdas Bandekar, C Vijayalakshmi. January 2020. Procedia Computer Science 172(3):122127
- [9] Shojaee S, Mustapha A, Sidi F, Jabar M A. A study on classification learning algorithms to predict crime status. Int J Digital Content Technol Appl. 2013;7(9):361-369.
- [10] Suhong Kim, Param Joshi, Parminder Singh Kalsi, Pooya Taheri, “Crime Analysis Through Machine Learning”, 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), November 2018, British Columbia,
- [11] YingLung Lin, Tenge -Yang Chen, Liang-Chih Yu, “Using Machine Learning to Assist Crime Prevention” 2017 6<sup>th</sup> IIAI International Congress on Advanced Applied Informatics, July 2017.