A Machine learning Approach for Crime Rate Analytics: Survey

Priyanka Jadhav[1], Amruta Ladgaonkar[2], Kaivalya Patil[3], Shivani Dive[4], Prof. Nakul Sharma[5]

[1234](Students, STES'S SINHGAD ACADEMY OF ENGINEERING, KONDHWA, PUNE, India)
[5](Professor, STES'S SINHGAD ACADEMY OF ENGINEERING, KONDHWA, PUNE, India)

Abstract:

Criminal activities are present in every region of the world affecting the quality of life along with the social and economic development of the country. Thus, there is an urgent need for security representatives and agencies to fight and reduce crime. Many governments are trying to use advanced techniques to tackle such issues.

Crime prediction practices historical data and after examining data, predicts the upcoming crime concerning location, time, day, season and year. The main aim of this paper is to describe a summary of the computational techniques that has been implemented toward crime data analysis and prediction. The survey of different techniques is performed besides with an abstract view of the proposed system that we are going to implement. The proposed system helps to predict and solve crimes at a much faster rate using machine learning techniques to reduce the crime rate. The system helps to prevent crimes or to improve the investigation efforts.

Keywords: Crime Rate Analytics, Machine eLearning Approach, Crime Prevention, crime types, prediction model, Data Collection, Dependent Variable Analysis, Word2Vector Formation, and Algorithm Analysis.

I. INTRODUCTION

Nowadays crimes are increasing drastically worldwide which is harmful not only to some individuals or individuals but also to a community, society or the state (Jazeem et al., 2015).It directly affects the financial growth of a nation by placing the financial burden on government due to the need for additional police forces, courts for solving the crime issues. (Hitesh Kumar et al., 2018)figure out that there is an 8% rise in the murder rate, an increase of 46 victims, with 629 homicides recorded in the 12 months to June, excluding the 35-people killed in the London and Manchester terrorist attacks. Such incidents can be avoided if preventive care is taken in advance. If it is possible to predict the crime occurrence, location, date and time then crime rates can be significantly reduced that helps save lives of the most valuable thing. Crime forecasting is a way of trying to mine out and decreasing the upcoming crimes by forecasting the future crime that will occur.

For crime detection in a particular region, the underlying pattern of crime and its relation to a region or area helps us for the same. According to previous studies, it is clear that in every city there have few streets or locations which have a higher concentration of criminal activities compared to others. Crimes can be predicted as the criminals are mostly active and operate in their comfort zones. Once successful they try to repeat the crime under a similar environment. The crime sequence and the patterns (Benjamin et al., 2017)which several criminals follow when committing a crime make it easy for analyzing the crime. The occurrence of the crime depends on several factors such as intelligence of a criminal, security of a location, date and time, etc. Usually, Criminals choose a similar location and time for attempting the next crime. Although it may not be true for all the cases, the possibility of repetitions is high, as per studies, and this makes the crimes predictable.

Instead of the availability of the advanced techniques, there are some challenges in the crime prediction techniques mentioned below:

A. Challenges of Crime Prediction Techniques

- ❖ Provide an accurate prediction for the location of the criminal.
- ❖ Collecting and managing large volumes of accurate data.
- ❖ Provide good performance by combining prior knowledge.
- ❖ Fast evolution of crime site data.
- ❖ Maintain effective crime analysis resource.

Thus, there is an urgent need for security agents and agencies to fight for and reduce the crime from society. The paper encloses the overview of the available techniques for crime prediction.

II.　　　LITERATURE REVIEW

(Suhong Kim et al. 2018)provides machine-learning-based crime prediction model for Vancouver. Vancouver crime data for the last 15 years is analyzed using two different data-processing approaches. Machine Learning predictive models KNN and boosted decision tree were used to obtain crime-prediction accuracy between 39\% to 44\%

(McClendon et al.,2015) implemented the Linear Regression, Additive Regression, and Decision Stump algorithms using the same finite set of features, on the communities and crime un normalized dataset to conduct a comparative study between the violent crime patterns from this particular dataset and actual crime statistical data for the state of Mississippi that has been provided by neighborhoodscout.com

(Kiran et al., 2018) proposed a framework for crime prediction based on the naïve Bayes classifier. The naive Bayes classifier is compared with the KNN classifier. The proposed techniques are implemented in Anaconda and a simulations result shows that naïve Bayes has high accuracy, less execution time.

(Xinlei et al., 2016) build the Vector Motion Model and propose a new algorithm named as TPML-WMA (Transition Probability Matrix Learning and Weighted Moving Average algorithm) to predict a future robbery distribution and figure out how it transfers. The data from the city in china from 2001 to 2011 is used to evaluate the TPML-WMA algorithm on brigandage prediction and discuss the performance of algorithms under different initial conditions.

(Jazeem et al., 2015).Proposeda hybrid approach to crime event prediction Using Deep Learning, to predict crime by focusing patterns and trends from various contributing factors.

(Morimoto et al., 2019) present a concept of information entropy for crime prediction and establishes a model flexible enough to apply under different circumstances.This system helps security organizations to address or react to crime occurrence proactively and helps local policy-makers to prevent or manage crime risks

(Adewale et al.,2017) provides a framework for visualizing the crime networks and analyzing them by various machine learning algorithms using the Google Maps and various R packages. Machine learning algorithms were used to extract the knowledge out of these large datasets and discover the hidden relationships among the data to get the crime patterns for crime analysts to analyze these crime networks by the means of various interactive visualizations for crime prediction.

The detection of criminal'sfrom any crime committed within the Redeemer's University is given by (Hitesh Kumar et al., 2018). For crime detection the Past data for both crimes and criminals were collected from Directorate of Students and Services Development (DSSD) and pre-processed to get clean and accurate data. The Iterative Dichotomiser 3 (ID3) decision tree algorithm was used to analyze and train the data. The trained model is used to develop a system that showed the hidden relationships between the crime-related data, in form of decision trees for crime prediction.

A.　Techniques Used in Crime Prediction

a.　Naive Bayesian(Kiran et al.,2018) (Mehmet et al.,2016)

Naive Bayesian is a good classification model. It gives the probability distribution to get an optimal result. It is based on probability. This is applied for calculating the posterior from the prior and the likelihood as it is easy to calculate from a probability model. This technique is used when the dimensionality of input is high. The equation used for calculating the posterior likelihood

$$P(A|B)=P(B|A)P(A)) / (P(B))$$

That is,

$$Posterior = ((Likelihood)(Propositionpriorprobability)) / (Evidencepriorprobability)$$

Steps of performing this algorithm are:

(i)　　　Convert the data set into a frequency table
(ii)　　 Create Likelihood table by finding the probabilities
(iii)　　Naive Bayesian equation is used for calculating the posterior probability for each class.

The factors related to geographic report are calculated first. For each factor, by using a discrete distance decayfunction(Renjie et al., 2010) which derives from the crime prediction theory Journey to Crime, create a geographic profile which is a probability distribution of being the next crime site on known geographical locations.

b. K-mean Clustering(M. Aasha et al.,2017)

(M. Aasha et al.,2017) Presented crime analysis for six cities of Tamilnadu, India by using clustering practices k-means, DBSCAN and Agglomerative clustering for grouping similar patterns to recognize crimes. The DBSCAN clustering performs better for grouping similar patterns in crime identification for six cities of Tamilnadu, India. The author used KNN to extract and predict future crimes that will occur in the future in six cities of Tamilnadu.

This algorithm finds groups in the given data substituted by k. It assigns each point to one group on the basis of features. An advantage of k means algorithm is fast, robust, and easy to understand and gives the best result. Steps for K-mean Clustering:

1. Input dataset, clustering variables and maximum number of clusters needed.

2. Initialize the cluster centroid.

3. Calculate the Euclidean Distance.

$$\text{Euclidean Distance} = \sqrt{(X_H - H_1)2 + (X_W - W_1)2}$$

Where,

- ❖ **X_H:** Observation value of variable height.

- ❖ **H_1:** Centroid of Cluster 1 for variable height.

- ❖ **X_W:** Observation Value of variable weight.

- ❖ **W_1:** Centroid value of cluster 1 for variable weight

Continue the steps until all observations are assigned and required clusters are found.

c. Random Forest Algorithm

It is a supervised classification algorithm with a decision tree model that performs both classification and regression. It is operated by constructing decision trees at training time and the result comes as individual trees. The advantages of this algorithm are it uses classification and the regression task both, also handle the missing values and it won't overfit when there is the number of trees in the forest and it also classifies categorical values. Steps for Random Forest Algorithm:

1. K features are randomly selected from total m features, where k<<m

2. Node d is calculated.

3. Nodes are splited into its branch nodes.

4. Steps from 1 to 3 are repeated still number of node become 1

5. Steps from 1 to 4 are repeated until a forest is created.

III. PROPOSED SYSTEM

The proposed system helps to predict and solve crimes at a much faster rate using machine learning techniques to reduce the crime rate. The system helps to prevent crimes or to improve the investigation efforts. Figure shows the system architecture of the proposed system.
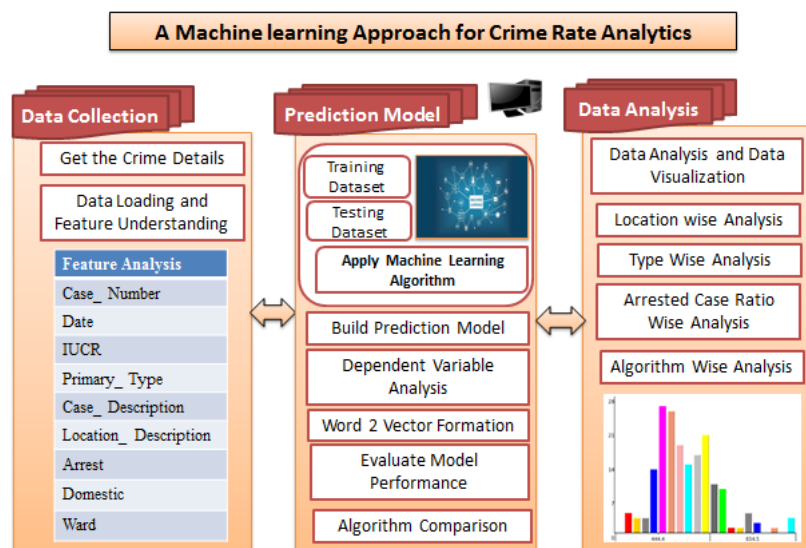

Figure 1: System Architecture

Criminal activities take place all over the world and law enforcement agencies have to deal with them effectively and efficiently. If enforcement agencies have a prior assumption of the class of the crime, it would give them tactical advantages and help resolve cases faster. The proposed system helps to detect the criminal activity in a geographic area to understand the underlying pattern of the crime the area suffers from.

1. Data Uploading and feature understanding

The data downloaded from kaggle is preprocessed first so that we can extract important features that are quite natural for predicting the crime. Such as few streets or locations, date and time, areas have a higher concentration of criminal activities compared to others. This data is used as input to the system to predict and solve crimes at a much faster rate.

Dataset Description :

This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. In order to protect the privacy of crime victims, addresses are shown at the block level only and specific locations are not identified. This data includes unverified reports supplied to the Police Department. The preliminary crime classifications may be changed at a later date based upon additional investigation and there is always the possibility of mechanical or human error. Therefore, the Chicago Police Department does not guarantee (either expressed or implied) the accuracy, completeness, timeliness, or correct sequencing of the information and the information should not be used for comparison purposes over time.

Link: - https://www.kaggle.com/chicago/chicago-crime?select=crime

This is a BigQuery Dataset. There are no files to download, but you can query it through Kernels using the BigQuery API.

The proposed system builds a prediction model for predicting the arrests for a given crime. The prediction accuracy is calculated based on the classification report generated. Here we have used a total of 100000 records for experimentation.

- Total Records→100000
- Training Records→75000
- Testing Records→25000

Attributes: 11

1) ID

2) case_number

3) Date

4) Block

5) IUCR

6) Primary_type

7) Description

8) Location_description

9) Domestic

10) Ward

Class = Arrest (Decision Class)

Different studies and researches have shown that significant concentration of crime happens at micro level of a region. Also, an overall study of criminal activity in a geographic area helps to understand the underlying pattern of the crime the area suffers from.

2. Dependent Variable Analysis

Dependent variables or Predicted variables are the ones that help to get the factors that mostly dependent on crime-related variables. For example, the Code for the type of crime (iucr) has nothing to do with the crime rate prediction or it is least bother for the prediction. So here by using the dataset, we achieve the terms or the factor that is mostly affecting to make crime prediction.
The analyzed data is visualized for the word to vector formation and on this fine-tuned data we can apply the algorithm to get the final result.

3. Analytics :
Exploratory Data Analysis is an initial process of analysis, in which you can summarize characteristics of data to can predict numerous crimes and predicting the type of crime, probable location, which may happen in the future depending upon various conditions.

4. Built Prediction Model
The system builds a prediction model by using a random forest technique. It is one of the ensembles learning technique which consists of several decision trees rather than a single decision tree for classification. While classifying all the trees in the random forest gives a class to an unknown example and the class having maximum votes will be assigned to the unknown example.
The techniques perform dependent variable analysis and word-formation vector to predict important aspects of crime detection and prevention.
The data analysis helps to identify useful features for building a predictive model, Such as Predicting arrests for a given type of crime in a given location and predicting the number of crimes in particular on a given day and time.

IV .CONCLUSION

Preventing a crime from occurring is better than investigating what or how the crime had occurred. Criminal activities are present in every region of the world. Accurate real-time crime predictions help to reduce the crime rate but remain a challenging problem as crime occurrences depend on many complex factors like crime location, Time, Date, Crime Type, Ward, Domestic, etc. A large amount of work is already being done toward this area with different approaches. But still, some work can be done to improve the prediction model. A survey of the existing system is conducted so that Crime prediction can be improved by the use of efficient data collection and data mining strategies.

Here an attempt is made using machine learning algorithms to fight against crimes and saving humanity. The primary objective of this work is to create a prediction model that can accurately predict crime for a particular location using the Random Forest machine learning technique.

REFERENCES

[1] JazeemAzeez ,D. John Aravindhar ,"Hybrid Approach to Crime Prediction using Deep learning ",2015 IEEE.

[2] Hitesh Kumar Reddy ToppiReddya,, BhavnaSainia, GinikaMahajana ,"Crime Prediction & Monitoring Framework Based on Spatial Analysis ",International Conference on Computational Intelligence and Data Science (ICCIDS 2018).

[3] H. Benjamin Fredrick David1and A. Suruliandi2 ,"SURVEY ON CRIME ANALYSIS AND PREDICTION USING DATA MINING TECHNIQUES", ICTACT JOURNAL ON SOFT COMPUTING, APRIL 2017, VOLUME: 07, ISSUE: 03.

[4] SuhongKim ; Param Joshi ; Parminder Singh Kalsi ; PooyaTaheri , "Crime Analysis Through Machine Learning", 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON).

[5] Lawrence McClendon and NatarajanMeghanathan*,"USING MACHINE LEARNING ALGORITHMS TO ANALYZE CRIME DATA",An International Journal (MLAIJ) Vol.2, No.1, March 2015.

[6] J. Kiran; K Kaishveen. ,"Prediction Analysis of Crime in India Using a Hybrid Clustering Approach", 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I- SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference

[7] XinleiWei ;Jinyao Yan ; Zheng Chen ; Tuo Shi ," Analysis of Crime Rate Distribution Based on TPML-WMA", 2016 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC).

[8] ShusukeMorimoto ; Hajime Kawamukai ,"Prediction of Crime Occurrence using Information Propagation Model and Gaussian Process", 2019 14th Asia Joint Conference on Information Security (AsiaJCIS).

[9] AdewaleOpeoluwa Ogunde1, *, Gabriel Opeyemi Ogunleye2, OluwalekeOreoluwa1 ,"A Decision Tree Algorithm Based System for Predicting Crime in the University ",Machine Learning Research 2017; 2(1): 26-34.

[10] Mehmet Sait, and Mustafa Gök. "Criminal prediction using Naive Bayes theory."Springer28.9 (2016): 2581-2592.

[11]   RenjieLiao, XueyaoWang,Lun Li and ZengchangQinh, "A Novel Serial Crime Prediction Model Based on Bayesian Learning Theory"Ninth International Conference on Machine Learning and Cybernetics, Qingdao, pp. 1757-1762, IEEE 2010.

[12]   A.Anisha.,C.Dhanashree, and A.Arpita ,. Application for analysis and prediction of crime data using data mining.International journal of advanced computational engineering and networking,2017

[11]   RenjieLiao, XueyaoWang,Lun Li and ZengchangQinh, "A Novel Serial Crime Prediction Model Based on Bayesian Learning Theory"Ninth International Conference on Machine Learning and Cybernetics, Qingdao, pp. 1757-1762, IEEE 2010.

[12]   A.Anisha.,C.Dhanashree, and A.Arpita ,. Application for analysis and prediction of crime data using data mining.International journal of advanced computational engineering and networking,2017