



[www.ijarr.org](http://www.ijarr.org)

## The Use of Machine Learning for the Prediction of Crime Type and Occurrence

<sup>1</sup> P. Ramakrishna, <sup>2</sup> Y. Kavya,

<sup>1</sup> Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.

<sup>2</sup> MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

### *Abstract—*

These days, criminal activity is a certain way to get yourself in hot water. Discontent at voting locations has been fanned by a nationwide increase in criminal activity. In order to detect and react to these illegal acts, it is essential to comprehend crime trends. This research analyzes crime data using the open-source Kaggle platform, which has been utilized to forecast the majority of crimes as of late. The project's overarching goal is to foretell the frequency, kind, and location of certain criminal acts. With the use of machine learning techniques like KNN and SVM, this study improves upon earlier attempts to identify a variety of crimes. Kaggle, crime model, analysis, prediction, accuracy, email notice are all keywords related to this topic.

### **I. INTRODUCTION**

It seems that crime is a significant indicator of imposition, which is used to a rather high degree. According to it, it's a crime when the leadership disobeys the law, disobeys official regulations, and resorts to violence. In order to analyze crime trends, it is necessary to look at both the various sorts of crimes and how they are reported. Technology may help the authorities put a stop to some of these crimes if they put more effort into it. Predicting crime kinds and trends using data and machine learning is so necessary. It uses historical crime data to make predictions about the kinds of crimes that will happen and when they will happen. Numerous investigations are carried out by researchers to ascertain regional crime trends and the correlations between them. It is now simple to categorize crime trends using some hotspot analysis. Faster problem-solving is facilitated by this. This technique makes use of time- and location-dependent data from the open-source Kaggle project. In order to assist you find crime hotspots and the exact kinds of crimes that are happening at certain times and dates, we have developed a categorization system. In order to identify and categorize different types of crimes and to forecast when they will occur, this article provides a thorough analysis of the use of machine learning, particularly KNN and SVM. Presently, the issue is that rising crime rates are a societal concern that has an impact on people's happiness and economic development. This is crucial information for determining whether one should relocate to a different location and for avoiding certain areas when visiting. To combat the rising crime rate and ensure the safety of their communities, law enforcement organizations are seeking innovative data mining techniques and sophisticated geospatial data processing. Criminals may strike anywhere, but they have a better chance of success if they choose areas with high crime rates. Consequently, variables associated with criminal activity must be investigated. To be better prepared to react to crime, it is vital to identify crime trends. The impact of diversity on crimes punishable by the Indian Penal Code or the Special List of Offenses (SLL) throughout all of India's states and union territories is the primary subject of this investigation. From 2001 to 2014, this research analyzed information from every state and union territory in India. Crime statistics in India broken down by area cover the years 2001–2014. Crime statistics against organized groups, children, and planned individuals cover the years 2001–2012. Murder statistics broken down by age and gender cover the years 2001–2012. 2010. According to the results, these are major

variables that influence a person's criminal record in India. We firmly think that studying the interconnections between misuse items will greatly assist in the prediction of future hazardous hotspots.

## II.LITERATURE SURVEY

Evaluating the frequency of certain types of crimes may be done in a number of ways. Data pertaining to studies conducted on the several purposes of crime type categorization are presented in this chapter. Characteristics include the ability to get email alerts, as well as features like the accuracy rate of predictions and population. Crimes manifest in many forms in various geographical areas. Crime rates are often compared to other variables such as unemployment, income, education level, etc. Examining the interplay of societal factors was his recommendation. The K-nearest neighbor algorithm (KNN) and the pruning technique are two machine learning approaches to prediction that Suhong Kim and Param Joshi [1] suggested. The accuracy rates for both crime pattern prediction and detection varied between 39% and 44%. Benjamin FREDERIC In order to get more insight from pre-existing huge data, David. H. [2] presented a data mining method that uses assessment and analysis. It is common practice to compare newly extracted patterns with previously collected data. The unlawful exploitation of mining rights is foretold by Shraddha S. Kavathekar [3]. Implicit algorithms are used by certain machine learning systems, such as ANNs and deep neural networks (DNNs). When fed certain amounts of data, deep neural networks perform better. Predictive models for several labels are usually constructed using all convolutional layers when DNN is used. The Tensorflow API, which is built for deep learning, mostly uses it for its output method. Since there are several types of crime and certain places have a higher incidence of some types of crime than others, these results indicate that priority is necessary for determining shock value. Artificial Neural Networks [ANN] predictive analysis in issue solutions. Patterns are the result of its numerous interrelated functions. Feature extraction from cloud-based data processing was suggested by Chandy and Abraham [4] using a random forest classifier. Request number, user ID, expiry time, arrival time, and memory need are the characteristics that have been retrieved. To understand the content of the extracted product from the user's request, the prediction function is performed after the extraction feature utilizing the information obtained in the learning phase. The Apriori approach was suggested by Rohit Patil, Muzamil Kacchi, Pranali Gavali, and Komal Pimpriya [5] to examine active patterns and make use of K-means findings. The system has to handle more data and spend more time on human verification since crime has been on the rise recently. This is why K-means clustering and other cutting-edge ML techniques have been used. Crime hotspot detection technology output derived from processing (SLR) quality data analysis was suggested in a research on crime spatial hotspot prediction [6]. The aim was to identify and quantify the position and timing of the hotspots. In order to aid in the identification of power line corrosion, Fuzhan Nasiri, Zakikhani, Kimiya, and Tarek Zayed [7] put forth a prediction model. Either a large amount of historical data is included or empirical data is the only basis for most forecasting models. This helps to eliminate any environmental conditions that might cause corrosion. An examination of data mining techniques for crime prediction was carried out by Nihli Dubey and Setu K. Chaturvedi [8]. The use of technologies that can identify the country's crime and help categorize cybercrime crimes is proposed by a computer system that uses machine learning methods for crime classification [9]. A fusion approach based on deep neural networks was suggested by Hyeon-Woo Kang and Hang-Bong Kang [10] to forecast infractions at the level of equipment where problems manifest.

## III.METHODOLOGY

### *Data collection:*

Document details on past crimes, such as where they occurred, when they occurred, the nature of the incident, the surrounding weather, and any public spaces. Characteristics that guarantee they are of equal size. Several methods are often used for crime prediction: • Decision tree • Random forest • KNN • Support vector machine (SVM) • Neural network Consistent reasoning and understanding of every algorithm Performance and scalability. Training the model (B): Divide the data into two sets: one for training the model and another for verifying its accuracy. Choose a model and tell us about it. The model's performance was checked using k-fold cross-validation to make sure it can be applied to other situations. C)Check the model's stability using methods like bootstrapping or testing it on data that hasn't been seen. As time goes on and more data becomes available or the model itself changes, the model's performance is adjusted accordingly. Review criteria should be objective, clear, and fair in order to minimize prejudice.

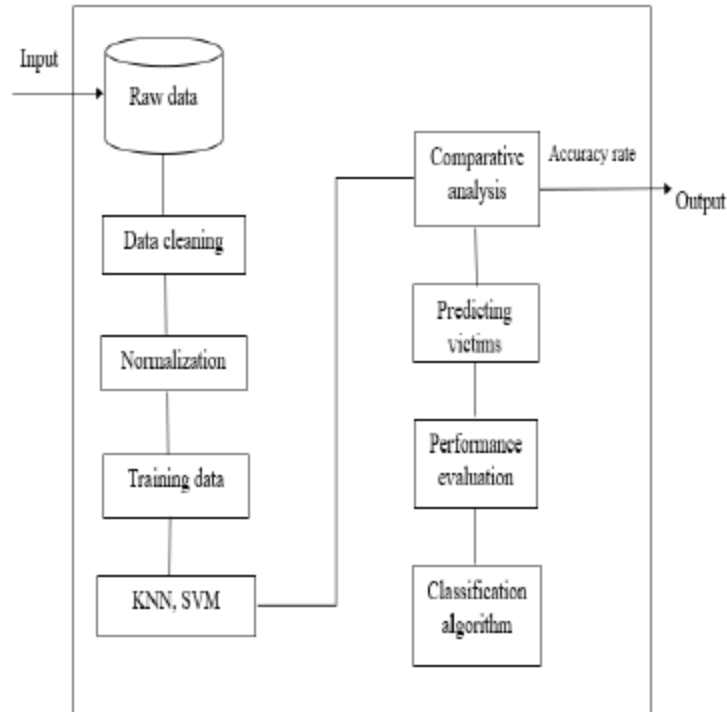


Fig.1: Flow chart of the proposed model.

### 1) Data Pre-processing :

In order to eradicate the non-conformity, preparatory information should be collected from open source. Denver, which has complete crime data going back six years, is the target of these data selections. Filters and wrappers built using machine learning techniques that seek for missing material at certain attribute values. In order to begin training predictive models, data cleansing is a must. Data is filtered by removing irrelevant material and events. Values are better evaluated with the aid of the filtering process. When you're choosing features, think about utility relationships. In order to train the prediction model, we employ the wrapping approach to assess the features' usefulness. The historical data was split into two categories: testing designs and training characteristics. Part 2: Mapping First, you need to separate apart the specifics of the crime, such as the nature of the offense and when it happened. To make labeling easier, this is then converted to an integer. Additional analysis was conducted on the noted spots and they were included into the drawing. Due to its suitability for machine learning processes, Python was selected as the programming language for the proposed project's implementation. To find out how the violation turned out, the matplotlib package is delivered. In many cases, the ensuing infractions may be visually shown on a map, which provides further forecasting assistance. Part 302 of the Indian Penal Code, which deals with murder, is Category A. • Class C (Crimes Committed at the Scene) • Class B (Offenses Against Scheduled Castes, Scheduled Tribes, and Children, IPC Sections 376 and 317) As a result, you can see which states and territories (prediction characteristics) are more likely to have this kind of crime. K-Nearest Neighbor (KNN) and other categorization methods are used. Thirdly, predicting criminal activity: Analyzing crime statistics in India is the primary goal of this research, which hopes to shed light on the root causes of crime in the country and provide solutions to the problem. Data obtained from numerous websites may be shared on Kaggle.com, data.gov.in, etc., in accordance with the topic rules. The data is prepared for use after collection by cleaning and standardizing it. See where things are with the feature extraction problem report file at the moment. Find out what kind of crime it is by making people commit crimes.

## IV. TEST AND RESULTS

Software development's completeness, security, correctness, and quality may be confirmed by testing. On their behalf, the participants undergo a rigorous study procedure that aims to unearth valuable information about the

product. Validated research. You can tell how much testing the product or procedure has undergone by looking at these factors: Make it through its design and construction without a hitch, • Accurately process a variety of inputs, In contrast, the accuracy graph reveals a steady rise in imprecision. A high score indicates that the model did a good job of identifying the crime type in the training data. As the model gains experience and knowledge, it becomes better at making predictions, which in turn produces new pictures.

**TABLE I : PERFORMANCE MEASURE OF THE KNN**

EVALUATION METRICS	CROSS VALIDATION
Accuracy	93.07%
Precision	92.53%
Recall	85.76%
F1 score	92.12%

After that, we test how well the implicit prediction worked to see whether it could outperform the old model. Educational contests facilitate the exchange of information about various course materials. Within the context of implicit cross-validation, it will assess the precision of each sector. Passing profile parameters like model name, target set, and cv helps express the segmentation result and allows us to compute the right value in Python. At last, we determined the real mean and its standard deviation. As compared to the present prediction model, the accuracy rate is much higher, reaching 93.17%.



Fig.2: Mail Notification of a Particular Location.

Number of offenses of each primary type of crime, 2001-present

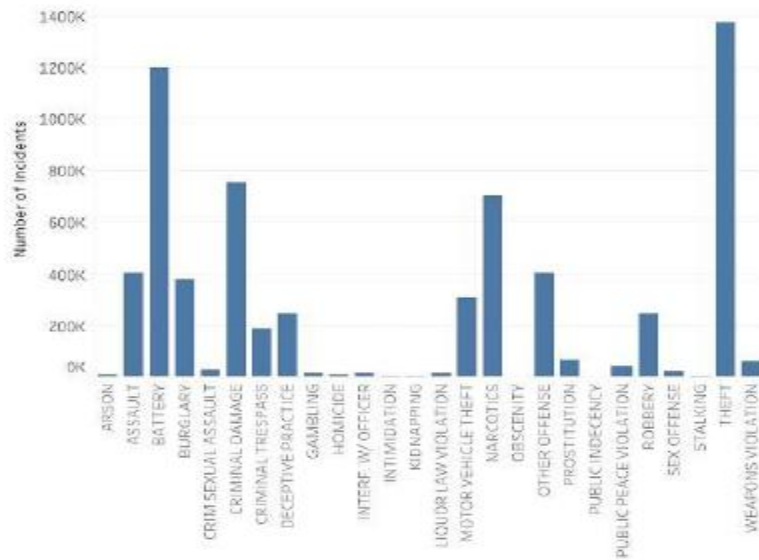


Fig. 3: Performance analysis of different crime type.

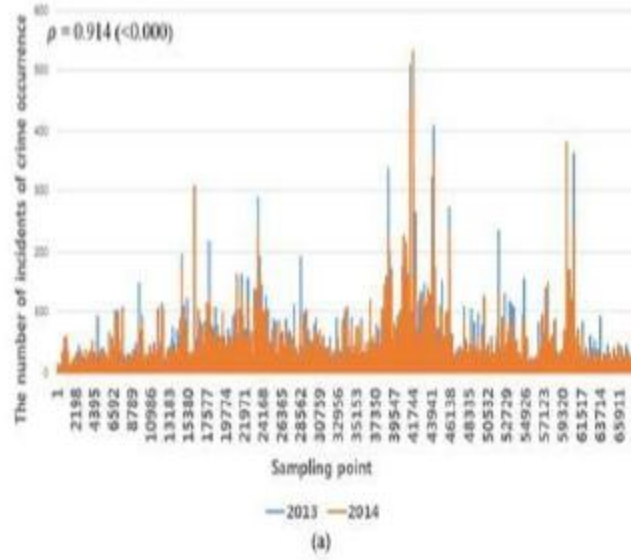


Fig.5: Comparison of the number of incident of crime occurrence in India by sampling point of 2013-2014.

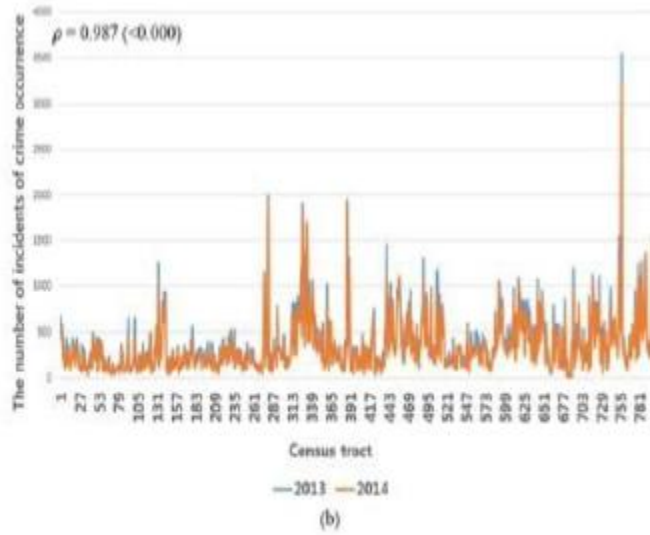


Fig. 6: Comparison of the number of incident of crime occurrence in India by census tract of 2013-2014.

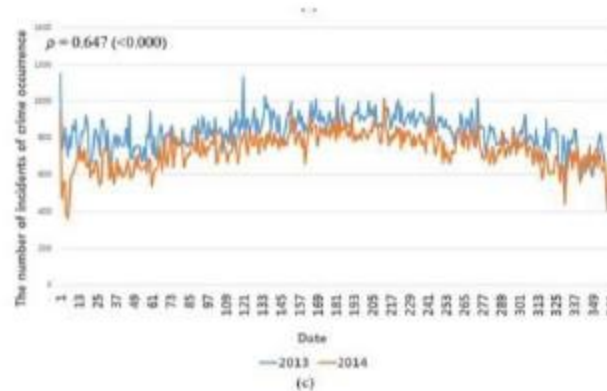


Fig.7: Comparison of the number of incident of crime occurrence in India by date of 2013-2014.

## V.CONCLUSION

Law enforcement tactics stand to benefit greatly from the use of algorithms and machine learning into crime modeling and incident prediction. We can create a more effective and data-driven strategy for crime prevention by using predictive analytics. Technology and law enforcement working together will become an effective weapon in the fight against crime, for a safer society, and for greater public confidence as these predictive applications are refined and expanded. In addition to enhancing law enforcement, criminal justice serves as a foundation for future reform efforts, community involvement, and strategic planning, all of which contribute significantly to making our society safer. One major benefit is that it shows how machine learning might be used to address societal crime issues via crime rate prediction. We build and test prediction models that might increase ownership of eight administrative public safety and law enforcement initiatives by analyzing past crime statistics, societal factors, and geographical features. It is important to note that crime prediction is a complex topic, even if this little effort is making progress in the area of machine learning-based crime prediction. Research must be continuous, and cooperation between lawmakers and law enforcement must be maintained, in order to build these models and put them into practice. It is becoming more apparent that public safety might be improved via the use of machine learning for crime prevention. We may strive to make communities safer by reducing crime, improving resource allocation, and using data-driven insights and prediction models. Based on the evidence, we think machine learning crime has a lot of potential to make intelligence and crime analysis better and faster. We will create intelligence and heuristic research into criminal trends.

## REFERENCES

1. Suhong Kim, Param Joshi, Parminder Singh Kalsi, Pooya Taheri, "Crime Detection with Machine Learning," IEEE Transactions, November 2018.
2. Benjamin Frederick David. H and A. Suruliandi, "Criminal Investigation and Criminal Investigation Using Data Mining Techniques," ICTACT Soft Computing Magazine, April 2012.
3. Study Rohit Patil, Muzamil Kacchi, Pranali Gavali and Komal Pimparia, "Crime Detection, Investigation and Research Technology", International Research Journal of Engineering and Technology, (IRJET) e-ISSN: 2395-0056, Volume: 07, Issue: 06 June 2020.
4. Umair Muneer Butt, Sukumar Letchmunan, Fadratul Hafinaz Hassan, Mubashir Ali, Anees Baqir and Hafiz Husnain Raza Sherazi, "Spatio-Temporal Analysis of Crime Hot Spot Detection and Prediction", IEEE Transactions, September 2020.
5. Nikhil Dubey and Setu K. Chaturvedi, "A study on crime prediction using data mining," Corpus ID: 7997627, published in 2014.
6. Abdulrahman Al-Ahmari, "Computational cybercrime classification using machine learning," Sustainability Journal, Volume 12, Issue 10, published May 2020.
7. Hyeon-Woo Kang and Hang-Bong Kang, "Exploring crime scenes from multiple databases using deep learning," peer-reviewed journal, published April 2017.
8. Almuhanha, M. M. Alrehili, S. H. Alsubhi and L. Syed, "Exploring communal violence in New York City using data analytics," Proc. Meeting. Artist Intel. Assessment of Information (CAIDA), April 2021.

9. Baqir, S. U. Rehman, S. Malik, F. U. Mustafa and U. Ahmad, "Performance evaluation of hierarchical clustering algorithms to identify crime hotspots," We are International. Meeting. Mathematical calculations engineer, Technology (iComET), January 2020.
10. Algefes, N. Aldossari, F. Masmoudi thiab E. Kariri, "Kev Txhaum Cai Tweet Text Discovery Methods in Saudi Arabia: From Analysis to Prediction", "Criminal Tweet Text Discovery Methods in Saudi Arabia From Analysis to Prediction", Seventh International. Meeting Data Science. Mach. to learn application form. (CDMA), March 2022.
11. S.P.C.W. Sandagiri, B. T. G. S. Kumara and B. Kuhaneswaran, "Investigating crime-related Twitter posts using artificial neural network-based approach," International Meeting Information and communication technology emerged Region (ICTer), November 2020.