

EMPIRICAL STUDY ON CRIME PREDICTION AND POLICING SYSTEM

Riya Srivastava¹, Shikha Nigam², Kritika Srivastava³, Mahima Verma⁴, Mr. Ashutosh Kumar Rao⁵

*¹⁻⁵Department of Computer Science and Engineering
Institute of Technology and Management Gorakhpur
Uttar Pradesh, India*

Abstract

This review paper deals with the increasing rate of crime and its adverse impact on society by introducing predictive (ML) machine learning algorithms to foresee future crime hotspots based on previous crime data. It empowers law enforcement agencies to better police the area. It helps in utilizing available resources more efficiently. Therefore, the method of predicting crime is most important, and it identifies future crimes and reduces the number of crimes. In this paper, we have designed a model to predict future crime incidents and also predict what kind of crime may occur. First, we analyse how certain crime features like given a date, time, and location of the crime. Second, we discuss several analytics techniques we used to find meaning in our data, classification models like naïve Bayes, Logistic Regression, and Decision Tree. Finally, we select the apt model for predicting crime type and seriousness, and location of the crime.

Keywords- Machine Learning, Naïve Bayes, Logistic Regression, Decision Tree, Crime Prediction, Predictive Policing.

1. INTRODUCTION

Daily, crime incidences are steadily increasing. Because the crime is neither planned nor random, criminals manage to carry out their crimes thanks to new technologies and high-tech tactics. The basic concept of anticipating crimes before it occurs is known as crime forecasting. To predict a crime beforehand happens, you'll need some tools. In today's world, cops have a wide range of equipment to aid them with specialized jobs. Law enforcement agencies have amassed a huge amount of crime data in the past situation. For academics and analysts, it will present both opportunities and challenges. The authorities can use the links between crime features to detect the source of the crime.

A systematic technique for discovering and evaluating patterns in crime episodes is required for crime analysis.

Throughout the world, any society that has flourished has one thing in common i.e., less crime rate. Society needs to be crime-free to flourish.

To reduce the crime rate an efficient policing system is needed that is equipped with advanced technologies so that they tackle ongoing criminal activities. (Explaining how crime affects the society badly) crime or criminal offense is a harmful act not only to some individual but also to a community, society, or the state.

The main elements of crimes are as follows:

- Crime's work. One hidden assumption is the definition of law.
- The offenses thus defined are prosecuted not by private persons but by the agents of the State/Government.
- The outcome of a conviction is a punishment, not compensation (if any) to the injured party.

Hinder developments in society. It leads to the death of 6 people and the defacement of their property. It increases the costs of living for individuals within the society.

Crime can be reduced by implementing two ways.

- 1- To make the society aware of crime in their surroundings, so that they could protect themselves before some crime happens to them.
- 2- By equipping the police with a better policing system. It will be very helpful if they could predict the crime beforehand by analyzing the crime pattern in the area.

This will also decrease the huge load on the judicial system and courts which leads to a long delay in prosecution. Courts won't be overloaded with cases. It aims to propose an improved technology for criminal prediction to provide the police with rich sources of information and intelligence that can be used to predict and prevent crime incidents. The main goals are to provide accurate results quickly which can be helpful for the government officials to identify criminals and crime-prone areas. This paper, therefore, aims to apply Naive Bayes techniques that have been developed for Crime-hotspot Prediction.

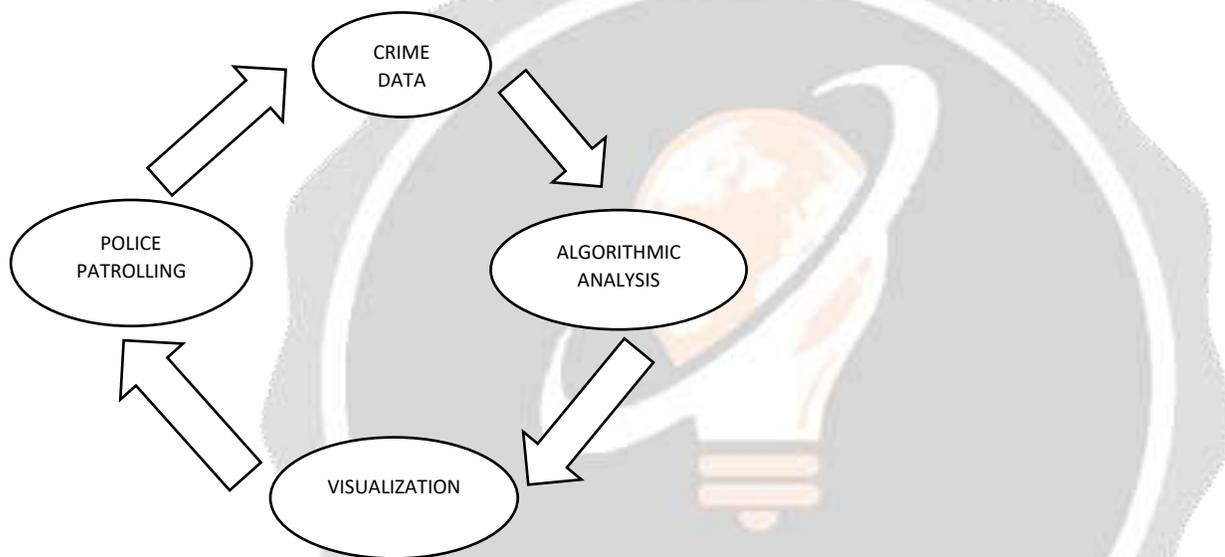


Fig 1 Framework for predicting crime

2. LITERATURE REVIEW

Tayla proposed an approach for crime detection and criminal identification in India by using data mining [1]. It is divided into six modules, which include: data extraction, Google Map representations, data pre-processing, WEKA implementation, classification, and clustering.

K-means clustering was used by them for analyzing crime detection. It generates two crime clusters based on similar crime attributes. Then WEKA was used for crime validation of their results, while KNN classification was used for criminal identification and prediction. The crime-analysis related work covers visualization methods [2, 3, 4], statistical methods [5, 6, 7], supervised [8, 9], and unsupervised [9, 10] learning techniques. The main purpose of this work is to identify and sort the text into natural images. It uses a text-classification engine to filter out series with low classification confidence scores. [11]

Authorized entities only provide limited access to incident-level data. For example, the date and location of events and criminal information, due to the potential for misuse of published information and breach of personal fact. Such as the Atlanta Police Department hides criminal-related information that has to be kept confidential for security issues.

Ajay Bazil Issac, C.P. Chaithanya, N. Manohar, C.P. Chaithanya, C.P. Chaithanya, C.P. Chaithanya, C.P. Chaithany proposed a text detection technique for identifying places in a photograph where there is text. Text detection and categorization in natural images are critical for various types of computer vision applications, such as optical character recognition, distinguishing between machine and human inputs, and spam detection.

In a Chinese town, Chen et alia proposed an auto-regressive coordinated shift toward a normal model for here and now crime expectations. For forecasting future events, the model used a well-known time series analysis technique.

In Salinas, Shingleton JS et al. introduced a method that used a multivariate analysis technique to predict three types of crimes: cruelty, arson, and homicide. They employed a conventional statistical procedure in California, including negative binomial, Poisson regression, and regression models. The results of the experiment performed for all three models shows that they perform similarly. Poisson regression models and Negative binomial regression, on the other hand, need the assumption that data traces the Poisson distribution. If the data cannot be fitted using a Poisson distribution, common minimum squares will suffice.

Shiju Sathyadevan, Deva M.S, suggested that the rate of crime is steadily growing. Because crime is neither random nor systematic, it is impossible to foresee. Modern technologies and high-tech approaches also aid criminals in committing their crimes.

Policymakers and law enforcement organizations all over the world are struggling to discover effective measures to regulate criminal networks, according to PAC Duijn, V Kashirin. Both network resilience and network topology are known to affect the success of disruption methods. Due to the secrecy with which these illegal criminal networks operate, data-driven understanding of the success of various criminal network disruption techniques is scarce.

K. Zakhir Hussain proposed it. Crime analysis is a branch of criminology, which is a work that implies investigating and detecting crimes as well as their connections to criminals. Because of the complexity of relationships between a large numbers of datasets, criminology is an ideal topic for applying data mining techniques. The initial step in generating further analysis is to identify the characteristics of crime. The knowledge gathered through data mining is a very effective tool for detecting violent criminal behavior. The goal is to apply data mining and a simulation model to try to capture years of human experience in computer models.

According to Nishat Sharma, criminal activities influence the socioeconomic growth and quality of life in every region of the world. As a result, several governments are concerned about it, and they are employing various technological technologies to address the problem.

Kianmehr and Alhajj et al. [9] developed a computational system for use in, St. Louis, Missouri, and, Columbus, Ohio that used SVM with the k-means for clustering to forecast criminal hot spots. Due to the lack of negative instances in a variety of dataset categories contributes to this problem when using the k-means for clustering and dataset division into small sets.

Based on the varied crime rates, these divided sets are identified as either a non-hotspot class or a hotspot class with this system. A dataset is categorized as a positive category (hot spot) if the criminal offense rate records are higher than a pre-defined rate, or as a negative category if the rate is lower (non-hot spot). Their framework can be used for criminal hot spot forecasting and other types of forecasting tasks.

SVM was employed by Wang et alia to forecast criminal lapsing. Training a logistic regression and a multilayer neural network was used to analyze the SVM. Even though the SVM and multilayer impartial system both outperformed the strategic relapse plan, they both outdid it. As a result, the authors combined the three models' expectations and found the best execution.

"Crime Analysis and Prediction Using K-Means Clustering Technique" by Wasim A. Ali, Husam Alalloush, and Manasa K.N proposes a framework for forecasting the likelihood of a crime occurring in a city by analyzing the dataset of crime incidents and visualizing the findings on a Google map for better comprehension. It foresees the areas where the most offense will occur. This study is carried out using the K-means clustering method, which groups related objects into small clusters. This technology might be used by officers to predict criminal cases and take appropriate action to capture perpetrators.

Sivanagaleela B [2] proposed a crime analysis methodological approach for detecting crime regions called "Crime Analysis and Prediction" utilizing Fuzzy C-Mean Clusters. This technique is mainly concerned with determining where the crime might occur. The fuzzy C-Means algorithm is used to cluster the crime-related data for total knowable crimes such as abduction, burglary, murder, theft, robbery, cheating, crimes against women, etc. in the existing Naive Bayes classification system. Based on the clustering technique, this concept is effective for identifying criminal hotspots. Because crime patterns are not static, they are also detected, which aids investigators in solving the crime rate. Because it takes less time, the crime-prone zones are discovered using the fuzzy clustering technique.

"Crime Analysis and Prediction" Dr. Sarvanaguru RA and Alkesh Bharati. K [3] proposed a method that focuses on predicting crime using the attributes present in the dataset using machine learning. The information was gathered from official sites. We can forecast the crime type that might occur in a given location using a machine learning system with Python as the core. The goal is to develop a predictive model. The training and testing dataset would be used to validate the training. Depending on the accuracy, a better algorithm can be used to build the model. For crime prediction, the (KNN) K-Nearest Neighbor classification and other algorithms will be employed.

"Crime Prediction Using Machine Learning," proposed by Rahul Shah [4], and "Crime Prediction Using Machine Learning," proposed by Akash Kumar and Aniket Verma. Many experts have worked on crime rate prediction and analysis using different methods such as k-means clustering, KNN, Naive Bayes, Fuzzy c algorithm, and others. Gandhali Shinde developed a scheme "Crime Prediction Using K-Nearest Neighboring Algorithm" [5], and many experts have worked on crime rate prediction and analysis using different methods such as k-means clustering, KNN, Naive Bayes. The Naive Bayes Algorithm was determined to be one of the best strategies for predicting future crimes among all the methods tested. As a result, they developed a system that combines Bernoulli NB with linear regression.

Implementation of the "prediction-led policing business process" (Perry et alia 2013). PRECOBS are a type of software that focuses on "closest prediction." This implies that only house burglaries that may happen as a result of the initial event are expected. Only a small amount of data, largely from police records, is required for the calibration of the software as well as for predictions. As a response, collecting data is a common aspect of police activity. The approach to prediction was understandable to the police officers who used the software, and they saw it as a useful addition to their arsenal.

Countering criminal activity has been a primary concern for all governments around the world. Many studies have been conducted to identify signs and evaluate crime patterns in an attempt to prevent them from happening again..

A study has been done to see whether there is a link between criminal activity and socioeconomic characteristics.

Some researchers examined mobile network infrastructure and concluded that it may be utilized to anticipate crime hot locations in London.

Sojavee et. al. used a several techniques, with the KNN algorithm outperforming the others, with a prediction accuracy of 89.50 percent.

Wang et alia [16] suggested an algorithm for detecting criminal trends perpetrated by an individual or a group of individuals.

Sadhana and Sangareddy [13] suggested that social networks may be misused as a prospective indicator of criminal behavior. They used Twitter data with sentiment analysis to forecast crime in real-time.

Toppireddy, Saini, and Mahajan use geo-location information to visualize and predict the next likely crime location, whereas two prior studies used document attributes as the data for analyzing and predicting crime..

Hakim et. al. used TF-IDF to create a lexicon to classify Indonesian news [8]. They acquired 7500 news items, each of which is divided into 500 categories.

Wongso et al. [9] attempt to combine various categorization algorithms. They mix (SVD) Singular Value Decomposition and three forms of Nave Bayes with TF-IDF implementation in their research. According to their findings, combining Multinomial Nave Bayes and TF-IDF yields 98.2 percent accuracy, which is the greatest among other combinations.

The majority of prior works on crime analysis have only focused on the most accurate and efficient method of crime prediction. Our research will be focused on determining the best accurate data mining strategy for acquiring criminal data from various categories on the internet. As a result, future academics and police officers will have access to more relevant and up-to-date crime data. In comparison to earlier studies, this study presents a comparison of the approaches used to classify Indonesian news.

There are numerous crime data mining approaches, including regression, classification, and clustering, according to studies. Depending on the kind of data provided, the training set, and the testing set, each method is employed in different contexts.

This section outlines some of the successful crime prediction systems developed by other researchers.

Wang et alia created the (HOT)Hotspot Optimization Tool, a crime-hotspot mapping tool that applies spatial data mining to the field of hotspot mapping. It can capture the differences between two classes in a spatial dataset [3]. HOT was able to precisely map the crime hotspots in a north-eastern city in the United States as part of the study.

Table1. Machine Learning Algorithms

S No.	Name of the method system	Performance Metric	Optimization
1	Decision Tree	Binary tree	Greedy search over a tree
2	Naïve Bayes	Bayes theorem	Bayesian
3	Logistic Regression	h (weighted sum)	Iterative, system of equations

3. CONCLUSION

This research work was done to compare different algorithms of machine learning and identify the most efficient algorithms. Logistic regression demonstrated optimal performance for all test cases compared to Naive Bayes and Decision Trees. It exhibits better F-measurement, sensitivity, precision, and specificity than Naïve Bayes and Decision Tree algorithms. Logistic regression and Naïve Bayes being a supervised machine learning algorithm gives better results. Crime is a common social problem that affects the quality of life and economic development of society. With the rise in crimes, law enforcement agencies are demanding advanced systems and new approaches to improve crime analysis and better protect their communities. With the increase in crime, laws enforcement agencies are continuing to demand advanced systems and new approaches to improve crime analytics and better protect their communities. This work proposes a crime event foresee strategy by comparing different models.

4. REFERENCE

1. D. K. Tayal, A. Jain, S. Arora, S. Agarwal, T. Gupta, and N. Tyagi, "Crime detection and criminal identification in India using data mining techniques," *Ai & Society*, vol. 30, pp. 117-127, 2014.
2. Brunson C, Corcoran J, Higgs G (2007) Visualizing space and time in crime patterns: a comparison of methods. *Comput Environ Urban Syst* 31:52–75
3. Vural MS, Go'k M, Yetgin Z (2013) Generating incident-level artificial data using GIS-based crime simulation. In: International conference on IEEE electronics, computer, and computation (ICECCO' 2013). Ankara, pp 239–242
4. Xiang Y, Chau M, Atabakhsh H, Chen H (2005) Visualizing criminal relationships: comparison of a hyperbolic tree and a hierarchical list. *Decis Support Syst* 41:69–83
5. Tollenaar N, Van Der Heijden PGM (2013) which method predicts recidivism best? a comparison of statistical, machine learning, and data mining predictive models. *J R Stat Soc Ser A Stat Soc* 176:565–584
6. Enzmann D, Podana Z (2010) Official crime statistics and survey data: comparing trends of youth violence between 2000 and 2006 in cities of the Czech Republic, Germany, Poland, Russia, and Slovenia. *Eur J Crime Policy Res* 16:191–205
7. Holst A, Bjurling B (2013) A Bayesian parametric statistical anomaly detection method for finding trends and patterns in criminal behavior. In: Intelligence and security informatics conference (EISIC) 12–14 Aug. 2013. Uppsala Sweden IEEE, pp 83–88
8. Jain LC, Seera M, Lim CP, Balasubramaniam P (2014) A review of online learning in supervised neural networks. *Neural Comput Appl* 25:491–509
9. Corsini P, Lazzarini B, Marcelloni F (2006) combining supervised and unsupervised learning for data clustering. *Neural Comput Appl* 15:289–297
10. Mackenzie MD (1995) CDUL: class directed unsupervised learning. *Neural Comput Appl* 3:2–16
11. Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017.
12. Chen P, Yuan H, Shu X, et al (2008) Forecasting crime using the ARIMA model. In: Proceedings of the 5th international conference on fuzzy systems and knowledge discovery. pp 627–630.
13. Shingleton JS et al (2012) Crime trend prediction using regression models for salinas, California. Naval Postgraduate School, Monterey
14. Kianmehr K, Alhajj R et al (2006) Crime hot-spots prediction using SVM. In: Proceedings of the international conference on computer systems and applications. pp 952–959
15. Wang P, Mathieu R, Ke J, Cai HJ et al (2010) Predicting criminal recidivism with SVM. In: Proceedings of the international conference on management and service science. pp 1–91
16. Braithwaite, J.: *Crime Shame and Reintegration*. Cambridge University Press, Cambridge (1989)
17. Ehrlich, I.: On the relation between education and crime. In: *Education, Income, and Human Behavior*, pp. 313–338. NBER (1975)
18. Freeman, R.B.: *The Economics of Crime*. *Handbook of Labor Economics*, vol. 3, pp. 3529–3571 (1999)
19. Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., Pentland, A.: *Once upon a crime: towards crime prediction from demographics and mobile data*. In: Proceedings of the 16th International Conference on Multimodal Interaction, pp. 427–434. ACM (2014)
20. Patterson, E.B. Poverty, income inequality, and community crime rates. *Criminology* 29(4), 755–776 (1991).
21. Shojaei, S., Mustapha, A., Sidi, F., Jabar, M.A.: A study on classification learning algorithms to predict crime status. *Int. J. Digital Content Technol. Appl.* 7(9), 361 (2013)
22. Wang, X., Gerber, M.S., Brown, D.E.: Automatic crime prediction using events extracted from Twitter posts. In: International Conference on Social Computing, Behavioral-cultural Modeling, and Prediction, pp. 231–238. Springer (2012)