Airo National Research Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

DATA MINING ANALYSIS AND PREDICTION OF CRIME RATE IN DIFFERENT DISTRICTS OF KARNATAKA FROM 2011-2021

Prajvalarani Sushil Kumar Bogar

Dr. Amit Singhal

Abstract

As a society, we are deeply troubled by the prevalence of crime. There are a lot of crimes perpetrated every day, and the regular occurrence of these crimes has made people anxious. As a result, stopping the crime from happening is an important responsibility. Artificial intelligence has recently proven its worth in nearly every industry, including crime prediction. But, for reference purposes in the future, it is necessary to have an accurate database of the crimes that have taken place. Law enforcement authorities can be better prepared to prevent crimes before they happen if they can anticipate the kind of crimes that might occur in the future. From a strategic standpoint, law enforcement could benefit from the ability to forecast any crime based on time, place, and other factors. But with crime rates rising at such a frightening pace, reliable crime prediction is proving to be a formidable obstacle. As a result, techniques for crime prediction and analysis are crucial for spotting and reducing criminal activity in the future. Recently, a lot of people have been trying to figure out how to use machine learning and specific inputs to make crime predictions. Decision trees, KNN, and other algorithms are utilized for crime prediction. Two types of data are being used in our job: primary and secondary. We utilize the method to ascertain the path's prediction rate after evaluating the data to figure out, for many locations, the prediction rate of various crimes. Lastly, we utilize the projected rate to determine our safe route. People will learn about the dangerous area and how to safely get to their destination with the help of this job.

Airo National Research Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

Keywords: Data mining. Crime rate, analysis, prediction, Crime Prediction, Crime bureau.

1. INTRODUCTION

Data has grown in both quantity and complexity at an unprecedented rate in recent years, opening the door to new ways of thinking about and solving social problems. A data-driven strategy is necessary for complete analysis and prediction in light of current concerns, such as the increasing prevalence of crime. The purpose of this data mining study is to analyze and estimate crime rates across districts in the Indian state of Karnataka from 2011 to 2021. By leveraging advanced analytical methods, this research seeks to unravel patterns, correlations, and hidden insights within the crime data, providing valuable insights that can inform policymakers, law enforcement agencies, and communities in their efforts to develop targeted and effective crime prevention strategies. Investigating crime patterns on a district level in Karnataka might make a big splash in the criminology world, leading to better decisions based on data and a safer society overall.

Most public and commercial organizations have seen a dramatic shift away from traditional and antiquated methods of operation as a result of advancements in computer technology and new approaches. In most Western countries, police stations, record bureaus, and related units of the crime investigation department have been computerized. This has led to the use of advanced software to analyze, detect, and predict crime. Other features, such as time series data with general packet radio service and geographic information systems, can provide information about hotspots and related details. Such software is connected to the Internet, intranet, wifi, and blue tooth technologies.

Crime investigation is becoming more complicated at a rate that matches the rate of change in technology, the nature of crimes, and society at large. As a result, crime record bureaus frequently make efforts to adapt. When it comes to doing their jobs, the Indian police have also opted for new processes and cutting-edge technology. Efficient resource utilization is made possible with the use of computer-assisted technologies that integrate information and technology. Throughout the investigation and beyond, evidence, statistics, and all information pertaining to the crime will be preserved in various formats. This includes details pertaining to

ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

the victims, the area, witnesses, the FIR, and the 23 subheadings. Even in law enforcement, data is valuable for assessing new patterns and predicting specifics about criminal activity based on past incidents. An act or failure to act that gives rise to a legally punishable offense is known as a crime. Another way of looking at it is as breaking a public law, either by doing something that is illegal or by failing to do something that is required of them, which can lead to legal consequences.

2. LITERATURE REVIEW

Dheenathayalan, K., & Savitha, K. K. (2023) Identified and analyzed previously unseen data patterns for the purpose of data classification. In order to reduce crime more quickly, it is necessary to identify patterns. A "cluster" of crimes committed in close proximity to one another is known as a "hot spot" in the criminal justice system. Law enforcement agencies can be more effective when they use the geographic plot of crime to pinpoint problem areas. Using k-means (KM) clustering and an improvement by the grasshopper optimization algorithm (GOA), this suggested approach identifies crime patterns. Geospatial data is used to plot crime hotspots by mapping crime patterns. Following this, we will go over a few of the social and environmental aspects that play a role in the perpetuation of violence against women. Data obtained from the National Crime Records Bureau (NCRB) in India is used to apply the suggested method to official crime records. The purpose of this study is to examine and catalog the various forms that violence against women in India takes. It can also be used to examine and forecast trends of any type of crime.

Prathap, B. R. (2022) People utilize social media for a variety of purposes, including personal and professional networking, content sharing (photos, videos, etc.), and idea exchange. The study found that academics may use social media to look at geographical and temporal interactions as well as aspects of individual behavior. Research employing data collected from many online social media sites, including Facebook, news feed articles, Twitter, and others, has shown that criminology has emerged as a popular field of research on a global scale. The utilization of spatiotemporal links in user-generated content can yield valuable information for the investigation of criminal activities. By collecting and visualizing data from various news sources,

ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

the study alludes to the application of text-based data science. The aforementioned findings from a variety of social media offenses as well as official crime statistics were the impetus for this study. In this chapter, we will examine 68 different crime keywords that can be used to identify the type of crime involving geographical and temporal data. The Naive Bayes classification method is employed to subdivide crimes into groups based on time and location as reported in news feeds. You can use the Mallet program to get keywords out of news feeds. The K-means approach is used to identify the hotspots in crime hotspots. This methodology has resolved the issues with the present KDE algorithm and is used to handle crime density using the KDE approach.

Vivek, M., & Prathap, B. R. (2023) started to become more popular in the late 90s and has been instrumental in bringing people together all across the world. An enormous user base has been amassed and maintained through the continual development of new social media platforms and the enhancement of existing ones. Users may now connect with others who shared their ideas and give in-depth reports of global happenings. The posts of the average person became more prominent as a result, and blogging as a whole became more popular. It was a watershed moment for journalism when these posts started to be authenticated and featured in major news pieces. Using statistical and machine learning methods, this research intends to use Twitter as a social media platform to categorize, display, and predict crime tweet data in India. It will also offer a spatio-temporal perspective on crime in the country. After applying geographical limits and the search function of the Tweepy Python package with a '#crime' query, 318 distinct crime keywords were used for substring-keyword classification of relevant tweets. You can make analytical visualizations with the Bokeh module in Python and geospatial visualizations with the gmaps module. We compare the accuracy of three models—SARIMA, Long Short-Term Memory (LSTM), and Auto-Regressive Integrated Moving Average (ARIMA)—to anticipate the number of tweets about crimes over time.

Rajarathinam, A. (2023) An increasing number of incidents of child abuse, exploitation, and trafficking have made crimes against children a major problem in India. Poverty, illiteracy, and social shame in vulnerable groups worsen the issue of child abuse, which has reportedly increased by 67% in the past decade. Many children are trafficked or subjected to sexual

ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

exploitation because they were forced to work in dangerous conditions. Victims of these crimes often suffer from psychological and physiological effects that last a lifetime. The trend of India's Total Crime Rate (TCR) for children was evaluated in this study using panel data regression. The analysis was found to be suitable for the fixed-effect model. Additionally, the study anticipated that TCR will rise next year.

Wadhwa, A., & Srimuruganandam, B. (2022) In order to identify potential dangers and areas of sensitivity caused by natural disasters as well as human-caused reservoir management and dam construction, landscapes were subjected to vulnerability assessments. Here, physical vulnerability assessment index maps were generated using a quantitative and qualitative technique. The use of cellular automata techniques allowed for the proper advancement in identifying the spatial and temporal variance in urbanization. Slope, slope profile, aspects, relative relief, curvature, soil texture, lithology, river morphometric, precipitation, LULC, mass movement, floods, geological elements, earthquakes, and anthropogenic activities like hydroelectric projects were some of the selected parameters used in the vulnerability assessment. Using prediction models (MLR for P, D r, and D B), we were able to determine the total impact of each element on urbanization change and establish guidelines for the transition. From 2002–2019, the urban pixel count increased by over 34% according to the yearly comparison of classed photos, whereas official reports confirm a 43.2% increase in urban growth up to 2019. The most important element in rising urbanization, which is directly related to population change, was found to be the distance from main roads and trains, according to individual study for each causal factor. Rising populations, changing housing needs, and more commercialization were identified as the primary drivers of urbanization. Aside from political and social instruments, other causal elements were included while assessing the proximity matrix and the growing population. The vulnerabilities were projected all the way to 2050, providing a framework for managers and decision-makers to comprehend the future demands on land and water in relation to growth.

Airo National Research Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

3. RESEARCH METHODOLOGY

- **Dataset**

Primarily derived from fieldwork, the crime dataset contains information about actual crimes. Approximately 600 rows of detailed information make up this dataset. The system input features are chosen from the dataset and include important details such as Name, Years, Months, Crime Type, Crime Areas, Victim Genders, Victim Ages, Victim Areas, and Months. Table 1 shows the target variables of the select system, which are the characteristics of the perpetrators' ages, genders, and the victims' relations.

Table 1: Shows the target variables of the select system

No.	Details		
	Name	**Types of columns**	**Descriptions**
I.	Individual ID	Value type	Individual ID in dataset
II.	Name	String	Casualty individual name
III.	Year	Numeric	Crimes happens year
IV.	Age	Numeric	The age of the person in question
V.	Gender	String	Casualties fix
VI.	Time	String	Time when the crime has happened
VII.	Victim area	String	Region where the crime has happened
VIII.	Region	String	district of the person in question
IX.	Home town	String	Home town of the victim
X.	Month	String	The year in which the crime has occurred

- **Preprocessing**

We choose to exclude unknown values from the dataset because they represent incomplete works and not actual values to be evaluated. The papers were formatted with dates and times in MM/DD/YY HH/MM. This was done since the classification system made it difficult to directly match the dates and times. It was thought that a date might be classified into three categories:

Airo National Research Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

weekends, weekdays, and unaware. The characteristics of the time intervals between dates provide the basis of this categorization (Fig. 1). Gives an outline of how the system works. The first step of the process is to retrieve information from the data collection, which stores datasets for various purposes. The main data will undergo preprocessing to make it more suitable for criminal analysis. The forecast includes four target variables.

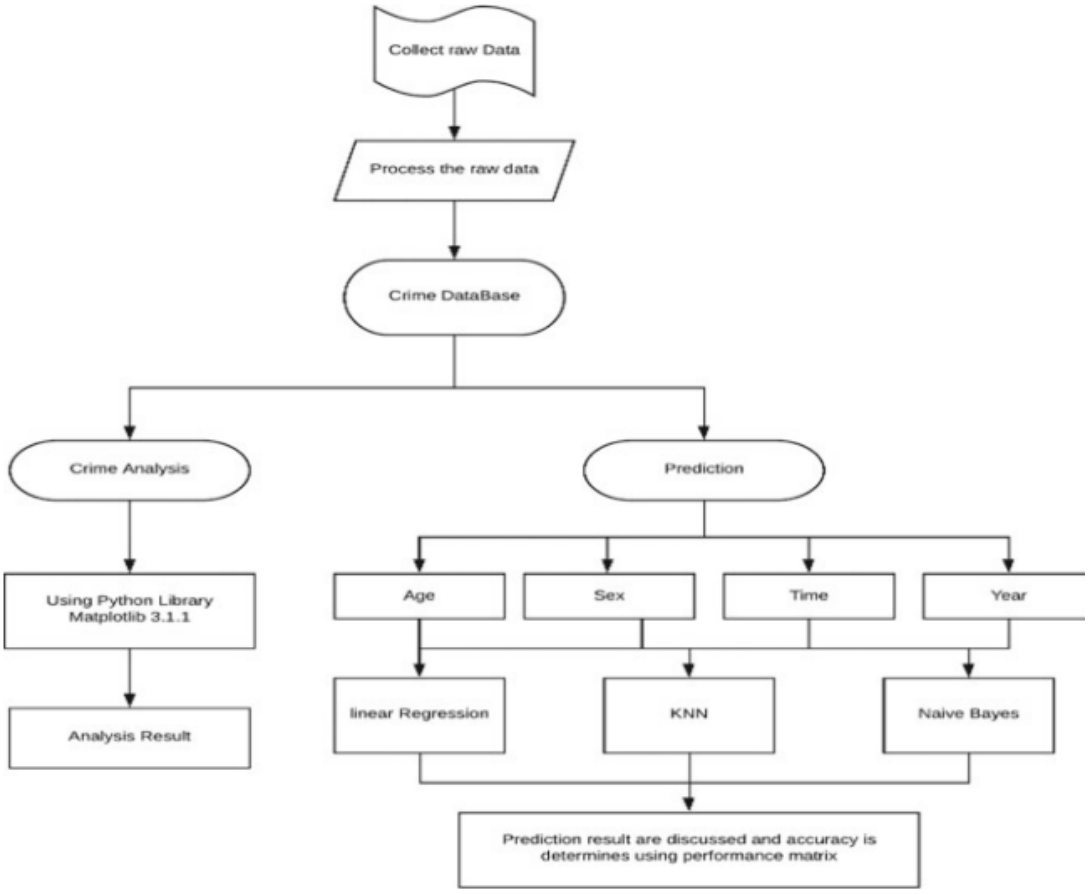


Figure 1 Diagram of the work flow

- **Crime Dataset**

Refer to Table 2. A few of the dataset's qualitative properties include the months, crime categories, victim gender, age, and area; these are included in the data set mentioned above. To utilize the mathematical models for prediction, this qualitative data needs to be categorized as either order 0 or 1.

Airo National Research Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

Table 2 Binary digit identification of males and females

Gender	Male	Female
	0	1

Table 3: presents the status of the month using binary values.

Months	D_A	D_B
Jan-April	0	1
May-August	1	0
Sept-Dec	1	0

4. DATA ANALYSIS

- **Linear Regression**

A mathematical method known as multi-linear regression can be used to determine a relationship between a set of independent variables that include values collected from the crime scene and a dependent variable, in this case the age of the victim. Based on the input criteria shown in the metadata column, this algorithm estimates the era of the victims' age values. According to the multi-linear regression,

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

Here,

Y serves as the dependent variable.

X serves as the independent variable

β reflect the regression's coefficient formula function.

Airo National Research Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

Table 4: Karnataka's male to female crime rate

Gender	Crime Rate				
	2017	**2018**	**2019**	**2020**	**2021**
Female	55	43	60	80	75
Male	60	50	72	85	80
None	0	0	0	0	0

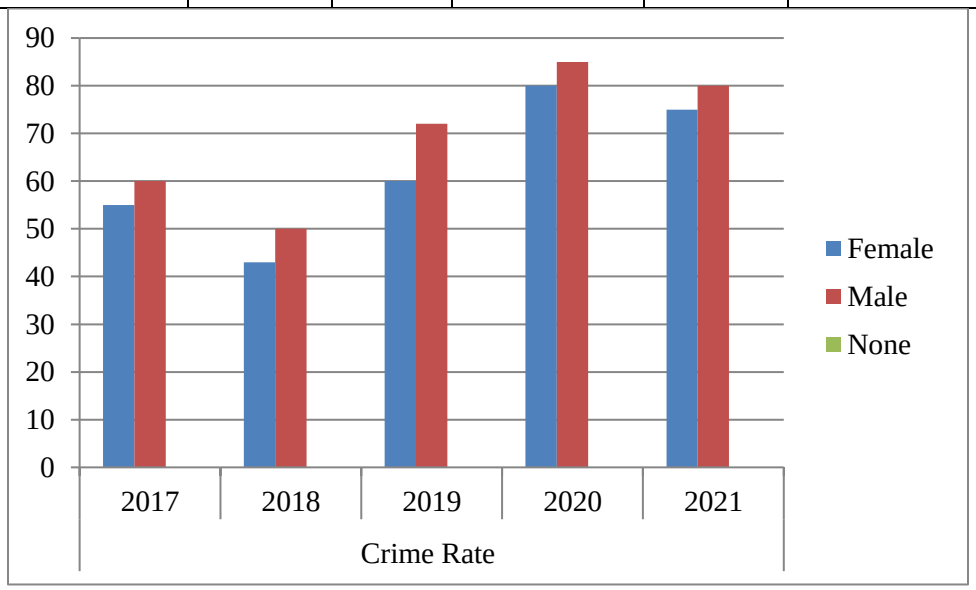


Figure 2: Karnataka's male to female crime rate

Interpretation: There are two primary groups in the data: female and male. A third group, "None," is used to identify genders that are not stated or are uncertain. Annually, the reported incidences per unit of population are used to calculate the crime rates. It seems like there's been a pattern of fluctuation in the Female category over the years. The crime rate peaked in 2017 at 55, dropped to 43 in 2018, and then shot up to 60 this year. The following year, 2020, saw a sharp increase to 80, while 2021 saw a small decline to 75.The crime rates for males tend to fluctuate throughout time. The 2017 rate was 60, the 2018 rate was 50, and the 2019 rate is 72. In the years that followed, the rising trend persisted, peaking at 85 in 2020 before declining marginally to 80 the following year.The fact that the None category always gives a crime rate of 0 indicates that no information is available for this particular gender.

Airo National Research Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

Table 5: Age in a range

Age	Range
Teenager	13-19
young	20-55
Old	56-100

Table 6: age-based crime rate

Age	Year				
	2017	2018	2019	2020	2021
Old	10	15	14	12	10
Young	50	55	60	70	80
Child	21	20	10	15	15

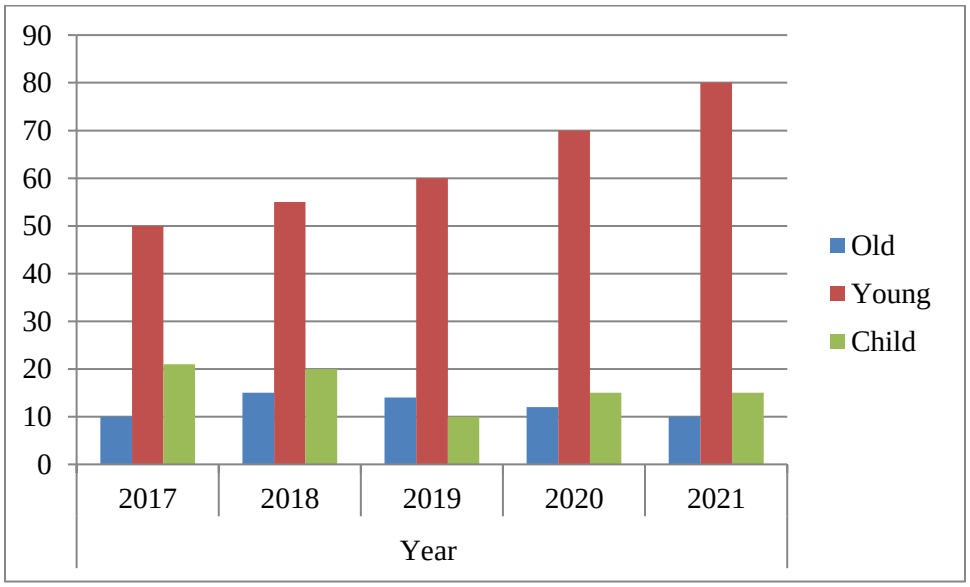


Figure 3: age-based crime rate

Interpretation: There are three separate age groups included in the table: Old, Young, and Child. The data is broken out over a five-year period, from 2017 to 2021. For each age category,

Airo National Research Journal
Peer-Reviewed
Multidisciplinary

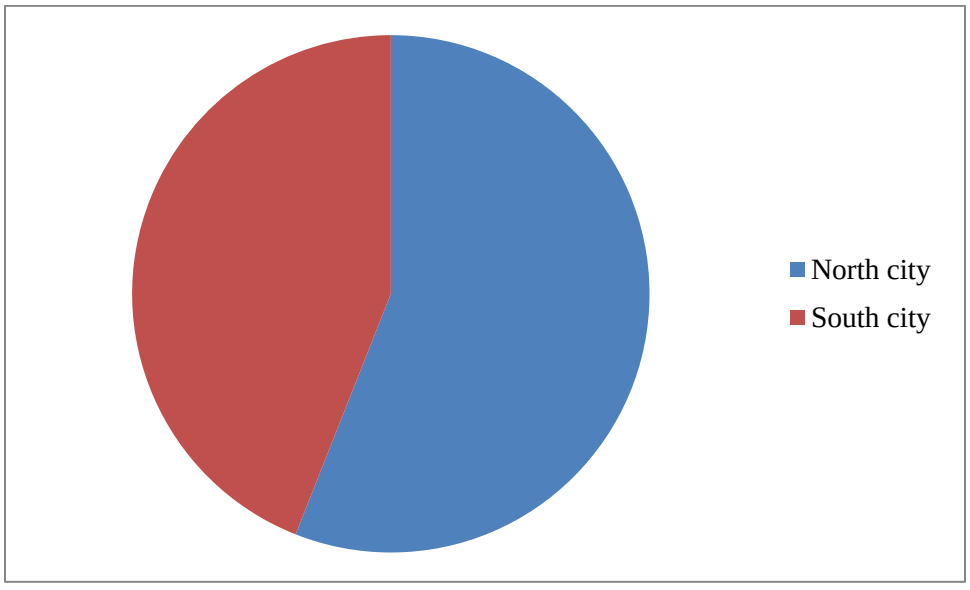
ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

the numerical values show the number of reported events per unit of population. From 2017 to 2021, the Old category's crime rates fall steadily, from 10 to 10. It appears that the criminal behaviors involving those classified as Old have stayed mostly consistent during the given time period. On the other hand, crime rates are on the rise among the Young category. Rates start at 50 in 2017 and go up every year until they reach 80 in 2021. Over the course of the five years, there appears to have been a marked increase in the number of recorded events involving those classed as Young. Contrarily, a more diversified pattern is observed in the Child group. From 2017's high of 21, the crime rate drops to 20 in 2018, then drops precipitously to 10 in 2019, before rising slightly to 15 in 2020 and 2021. Further examination into the underlying issues is warranted due to oscillations in the Child category, which may imply changes in reported occurrences involving individuals designated as Children.

Crime Rate Prediction Using Machine Learning and Data Mining

Table 7: Karnataka's rate of crime

North city	56
South city	44



ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

Figure 4: Karnataka's rate of crime

5. CONCLUSION

Notwithstanding, the use of the prediction rate region explicit displaying is made more troublesome by the way that crime is generally exceptional in many spots. Inside the extent of that work, we used the AI calculation to create and assess a gauge of crime in view old enough, orientation, year, moment, and month. Later on, we will actually want to find a more noteworthy level of accuracy by using prediction frameworks. Moreover, by using this degree of accuracy, we will actually want to distinguish and find the districts with the most elevated crime rates. For us to effectively follow through with this responsibility, we might want to utilize the CNN calculation to assess the image data and incorporate the Google Programming interface for seeing the hot zone.

REFERENCES

1. Chauhan, C., Sehgal, S.: A review: crime analysis exploitation data processing techniques and algorithms, pp. 21–25 (2017).
2. Dheenathayalan, K., & Savitha, K. K. (2023). CRIME AGAINST WOMEN IN INDIA: ANALYSIS OF SPATIAL DATA AND FACTORS INFLUENCING CRIME. *Semiconductor Optoelectronics*, *42*(1), 429-443.
3. Hass, J. (2021). Buried in Sand: Understanding Precarity in the Context of the Political and Criminal Economy in India.
4. Jayavadivel, R., Rajkumar, N., Shankar, B. P., Vetrimani, E., Viji, C., & Aarthy, G. (2022). Text Sentiment Analysis for Intelligent Transportation Systems by Using Web-Based Traffic Data: A Descriptive Study. *Recent Advances in Mathematical Research and Computer Science Vol. 10*, 11-26.
5. Kerr, J.: Vancouver police go high tech to predict and prevent crime before it happens. Vancouver Courier, July 23, 2017.
6. Lin, Y., Chen, T., Yu, L.: Using machine learning to assist crime prevention. In: 2017 sixth IIAI International Congress on Advanced Applied Science (IIAI-AAI) (2017)

ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

7. Mallouhy, R. E. (2023). *Predictive analysis of time series in various application contexts* (Doctoral dissertation, Université Bourgogne Franche-Comté).

8. Marchant, R., Haan, S., Clancey, G., Cripps, S.: Applying machine learning to criminology: semi parametric spatial demographic Bayesian regression. Security Inform. 7(1) (2018)

9. Munasinghe, M., Perera, H., Udeshini, S., Weerasinghe, R.: Machine learning based criminal short listing using modus operandi features (2015).

10. Prathap, B. R. (2022). Geospatial crime analysis and forecasting with machine learning techniques. In *Artificial intelligence and machine learning for EDGE computing* (pp. 87-102). Academic Press.

11. Rajarathinam, A. (2023). Panel data regression modelling for crime against children's data.

12. Ranjan, R., Vathsala, H., & Koolagudi, S. G. (2022). Profile generation from web sources: an information extraction system. *Social Network Analysis and Mining*, *12*, 1-12.

13. Saxena, R. K., & Kumar, B. (2019). Forensic Accounting: A Legal ICAI Approach to Investigate Scams And Fraud Cases In India.

14. Shah, H., Pandya, D., Panchal, K., & More, N. P. (2022, November). Classification of Machine and Deep learning Techniques for Financial Fraud Detection of Healthcare Industry. In *2022 International Conference on Futuristic Technologies (INCOFT)* (pp. 1-7). IEEE.

15. Vivek, M., & Prathap, B. R. (2023). Spatio-temporal Crime Analysis and Forecasting on Twitter Data Using Machine Learning Algorithms. *SN Computer Science*, *4*(4), 383.

Airo National Research Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2321-3914
Volume: 2 Issue: 3
June 2025
Impact Factor: 10.2
Subject: Computer Science

Author's Declaration

I as an author of the above research paper/article, here by, declare that the content of this paper is prepared by me and if any person having copyright issue or patent or anything otherwise related to the content, I shall always be legally responsible for any issue. For the reason of invisibility of my research paper on the website /amendments /updates, I have resubmitted my paper for publication on the same date. If any data or information given by me is not correct, I shall always be legally responsible. With my whole responsibility legally and formally have intimated the publisher (Publisher) that my paper has been checked by my guide (if any) or expert to make it sure that paper is technically right and there is no unaccepted plagiarism and hentriacontane is genuinely mine. If any issue arises related to Plagiarism/ Guide Name/ Educational Qualification /Designation /Address of my university/ college/institution/ Structure or Formatting/ Resubmission /Submission /Copyright /Patent /Submission for any higher degree or Job/Primary Data/Secondary Data Issues. I will be solely/entirely responsible for any legal issues. I have been informed that the most of the data from the website is invisible or shuffled or vanished from the database due to some technical fault or hacking and therefore the process of resubmission is there for the scholars/students who finds trouble in getting their paper on the website. At the time of resubmission of my paper I take all the legal and formal responsibilities, If I hide or do not submit the copy of my original documents (Andhra/Driving License/Any Identity Proof and Photo) in spite of demand from the publisher then my paper maybe rejected or removed from the website anytime and may not be consider for verification. I accept the fact that as the content of this paper and the resubmission legal responsibilities and reasons are only mine then the Publisher (Airo International Journal/Airo National Research Journal) is never responsible. I also declare that if publisher finds Any complication or error or anything hidden or implemented otherwise, my paper maybe removed from the website or the watermark of remark/actuality maybe mentioned on my paper. Even if anything is found illegal publisher may also take legal action against me.

Prajvalarani Sushil Kumar Bogar

Dr. Amit Singhal

*****