

Research Article

Crime Data Analysis in Police Personnel with Machine Learning**Mutsyala Anitha¹, Sattu Shivani², Alakanti Jayasurya³, Mallarapu Rohan Balaji Goud⁴ and Aseena Shaik Babu^{5,*}**¹⁻⁴Student, Department of CSE(AI&ML), Samskruti College Of Engineering & Technology, Hyderabad, Telangana, 500001, India⁵Assistant Professor, Department of CSE(AI&ML), Samskruti College Of Engineering & Technology, Hyderabad, Telangana, 500001, India

*Corresponding Author: Aseena Shaik Babu. Email: doctoraseen@gmail.com

Received: 02/10/2025; Revised: 18/11/2025; Accepted: 03/12/2025; Published: 31/12/2025.

DOI: <https://doi.org/10.69996/fmep.2025026>

Abstract: The increasing complexity and frequency of criminal activities demand intelligent solutions to support law enforcement agencies in maintaining public safety. This research presents a comprehensive crime data analysis framework aimed at enhancing the efficiency of police personnel deployment through machine learning techniques. By utilizing historical crime datasets, the study performs extensive data preprocessing including handling missing values, normalization, and feature categorization to uncover patterns and trends in criminal incidents. The analysis focuses on identifying victim demographics, weapon usage, crime timings, and high-risk areas through visualizations and statistical summaries. Various machine learning algorithms such as K-Nearest Neighbors (KNN), Logistic Regression, CART, and Random Forest are employed to classify and predict crime occurrences. The performance of these models is compared using metrics like accuracy, precision, and recall. The results demonstrate that CART and Random Forest outperform other models, making them suitable for crime prediction tasks. This data-driven approach empowers police departments with actionable insights for proactive intervention, smarter personnel allocation, and strategic planning to reduce crime rates and improve community security.

Keywords: Crime Data Analysis, Machine Learning, Police Personnel Deployment, Crime Prediction, Data Pre-processing.

1. Introduction

Crime is a persistent and evolving challenge that affects the social and economic fabric of societies worldwide [1]. In modern urban environments, where population density, socio-economic disparities, and technological advancements converge, the nature and complexity of crime have undergone significant transformation [2]. Law enforcement agencies are under increasing pressure to not only respond to crimes effectively but also anticipate and prevent them proactively [3]. Traditionally, police personnel have relied on manual techniques, intuition, and experience for crime investigation and analysis [4]. However, such conventional approaches are often time-consuming, prone to human error, and inadequate for dealing with the scale and variety of data generated in contemporary policing environments [5]. With the emergence of the digital age, police departments and investigative bodies now generate and collect massive volumes of data on a daily basis [6]. These include crime reports, FIRs, arrest records, location-based crime statistics, surveillance footage, social media content, and more [7]. While this abundance of data holds immense potential for deeper understanding and strategic intervention, extracting meaningful insights from it requires sophisticated computational tools. This is where the role of machine learning (ML) becomes prominent.

Machine learning, a subset of artificial intelligence, focuses on developing algorithms that can learn from and make predictions or decisions based on data. Its application in crime data



analysis is not only timely but also transformative. ML models can be trained on historical crime datasets to uncover patterns, identify trends, detect anomalies, and predict future criminal activities. These capabilities enable law enforcement personnel to take a data-driven approach, enhancing the speed, accuracy, and efficacy of decision-making processes. Furthermore, machine learning assists in optimizing patrol routes, identifying high-risk areas (hotspots), and allocating resources more effectively, thus contributing to smarter and safer policing. One of the key benefits of machine learning in crime analysis lies in its ability to handle unstructured and heterogeneous data. For instance, natural language processing (NLP) techniques allow models to process textual crime reports and extract relevant features, while image recognition algorithms can analyze surveillance images to identify suspicious activity. Temporal and spatial data can be used to predict crime occurrences based on time-of-day, location, or past trends. These analytical strengths provide police personnel with a comprehensive and multifaceted view of the crime landscape, moving beyond reactive crime fighting toward predictive and preventive policing.

Several machine learning algorithms have been employed for crime data analysis, ranging from basic classifiers such as Decision Trees and Naïve Bayes to more advanced models like Support Vector Machines (SVM), Random Forests, Gradient Boosting, and deep learning architectures. Each technique offers unique advantages in handling specific types of crime data, whether it involves categorizing crime types, forecasting future incidents, or identifying correlations among criminal behaviours and geographic zones. By leveraging the power of these algorithms, crime analysis can become more scalable and robust. However, the implementation of machine learning in crime data analysis is not without challenges. Data quality and completeness are critical for effective model training and accurate predictions. Many police databases suffer from issues such as missing data, inconsistency in data entry, and lack of standardization across jurisdictions. Data privacy and ethical concerns also pose significant hurdles, especially when dealing with personally identifiable information (PII), biometric records, or predictive models that may inadvertently introduce bias or reinforce existing inequalities. Another limitation is the interpretability of machine learning models, especially complex ones like deep neural networks. Law enforcement agencies require transparent decision-making processes that can be explained and justified in courts or public forums. Therefore, incorporating explainable AI (XAI) principles becomes crucial in crime-related applications of ML. Additionally, the lack of technical expertise and infrastructure within police departments can slow down the adoption of ML tools, making training and awareness programs essential for successful deployment. Despite these challenges, the potential advantages of machine learning in crime analysis are significant and far-reaching. Globally, cities and nations have started experimenting with predictive policing models, crime hotspot mapping, facial recognition, and real-time analytics to improve their public safety frameworks. Countries like the United States, the United Kingdom, and India have initiated smart policing projects that integrate ML-powered analytics with traditional investigative techniques to build more efficient and responsive law enforcement systems. In the Indian context, the application of machine learning in crime data analysis is gaining traction, especially with the growing emphasis on digitizing police records through systems like the Crime and Criminal Tracking Network and Systems (CCTNS). With millions of digitized police records, the Indian police force has an unprecedented opportunity to harness ML for evidence-based crime prevention and better governance. Moreover, as public awareness and demand for safety increase, there is a growing expectation that law enforcement should evolve technologically to meet modern-day challenges.

This paper contributes significantly to the field of intelligent law enforcement by presenting a machine learning-based framework for crime data analysis aimed at supporting

police personnel allocation and crime prevention strategies. The study systematically processes and analyzes real-world crime datasets to identify hidden patterns related to victim demographics, crime timing, weapon usage, and geographical crime trends. By implementing and comparing various machine learning algorithms, including KNN, Logistic Regression, CART, and Random Forest, the paper highlights the models most effective in predicting and classifying crimes. The visualization of analytical results enhances interpretability and provides actionable insights for police departments. The comparative evaluation reveals that decision tree-based models outperform others in terms of accuracy and reliability.

2. Related Works

Over the past decade, numerous studies have explored the integration of machine learning techniques in the domain of crime data analysis to support law enforcement agencies in combating criminal activities more efficiently. Researchers have examined a range of supervised and unsupervised learning methods to identify patterns, forecast future incidents, and classify different types of crimes. The primary focus of these works lies in developing predictive models that can process large-scale police datasets—such as FIRs, arrest logs, and geographic crime distributions—to aid in decision-making and resource optimization. Various algorithms including Decision Trees, Random Forests, Support Vector Machines, Naïve Bayes, and Deep Neural Networks have been applied across different geographical and societal contexts to evaluate crime trends and assist in proactive policing. In addition to structured data, unstructured data like textual descriptions from reports and real-time surveillance feeds have also been processed using Natural Language Processing (NLP) and image recognition models, respectively. While these studies offer promising results, they often face challenges related to data quality, class imbalance, privacy concerns, and ethical implications—factors that must be addressed for real-world deployment. The literature also reveals an emerging trend toward the adoption of Explainable AI (XAI) to ensure that model decisions remain transparent and accountable in law enforcement applications. This section reviews and synthesizes key contributions in the existing body of work to provide context for the present study.

Næss et al. [8] utilized text mining and ML to audit police handling of cybercrime in Norway, demonstrating how intelligent systems can evaluate law enforcement efficiency. Saeed and Abdulmohsin [9] investigated the use of data mining and textual ML techniques to predict crime rates, emphasizing the importance of unstructured data in forecasting criminal trends. Geetha et al. [10] proposed a machine learning-based crime prediction model validated using real-world datasets, offering improved classification performance across different crime categories. Taiwo et al. [11] incorporated Twitter sentiment analysis alongside crime statistics to enhance predictive accuracy, marking the relevance of social media analytics in crime modeling. Saini and Kaur [12] proposed predictive analytics to identify high-risk zones, aiding in crime prevention through location-based strategies. Similarly, Apene et al. [13] presented a comprehensive review tracing the shift from traditional policing methods to AI-enabled crime detection frameworks. Mishra et al. [14] combined statistical models and deep learning to analyze evolving criminal landscapes, highlighting the effectiveness of hybrid analytical approaches. In a related study, Hashi et al. [15] implemented object detection-based deep learning models to identify crime intentions from surveillance data, proving the feasibility of real-time intent recognition. Saravag and Kumar [16] developed a hybrid ML-regression model to validate a multidimensional crime index related to crimes against women, addressing the intersection of gender and predictive modeling. Kumar et al. [17] introduced a criminal face identification system using the MTCNN deep learning algorithm, offering practical applications in suspect tracking. Cunningham et al. [18] focused on forecasting police officer safety using

predictive modeling, thereby extending the scope of crime analytics to internal risk mitigation. Aziz et al. [19] presented a novel deep learning model optimized for predicting cognizable crimes under the Indian Penal Code, offering targeted insights for Indian law enforcement frameworks.

Cubitt [20] examined how criminal history and police intelligence contribute to personnel vetting, underlining the significance of crime analytics beyond prevention. Angbera and Chan [21-22] proposed a spatiotemporal DL model that integrates time and space features for more refined crime prediction. Kazemian and Shrestha [23-24] compared several ML techniques for detecting fraudulent criminal identities, reinforcing the role of data integrity and authentication in criminal investigations. Lastly, Egbert and Esposito [25-26] critically analyzed the shift toward algorithmic and precision policing, raising concerns about ethical boundaries and societal implications of automated law enforcement systems. In leveraging machine learning and deep learning techniques for crime prediction and analysis, several critical research gaps remain unaddressed. Most existing studies focus predominantly on general crime forecasting or the development of predictive models using static datasets, often neglecting the dynamic, real-time nature of police operations and personnel-specific crime data. While approaches like sentiment analysis, spatial-temporal modeling, and deep facial recognition have proven effective in certain contexts, there is limited exploration into the integration of these techniques specifically within the internal operational frameworks of police departments. Moreover, current models rarely address issues related to data bias, ethical constraints, or interpretability, which are vital for institutional adoption. Few studies provide insights into how ML can enhance strategic decisions made by police personnel or improve accountability, training, or safety outcomes within law enforcement agencies. This paper aims to bridge these gaps by focusing not just on crime trends, but on how machine learning can be tailored to support, analyze, and enhance police personnel functions in a practical, transparent, and responsible manner.

3. Data Analysis with Machine Learning

This project seeks to enhance the efficiency of police personnel deployment through the integration of machine learning (ML) in crime data analysis. By utilizing historical crime records, the study develops predictive models capable of identifying high-crime zones and temporal crime patterns. These insights can be instrumental in optimizing the allocation of law enforcement resources, ensuring that police presence is strategically distributed based on data-driven risk assessments. The use of ML algorithms enables a shift from reactive to proactive policing, empowering authorities to anticipate criminal activity, allocate patrol units more effectively, and implement timely interventions. Ultimately, this approach aims to reduce response time, improve public safety, and support evidence-based decision-making within police departments. The increasing volume and complexity of crime data in modern cities pose a significant challenge to law enforcement agencies, particularly when it comes to the optimal allocation of police personnel. Traditional methods of resource deployment are often reactive and rely heavily on historical intuition or manual analysis, which can be both time-consuming and error-prone. In response to these limitations, this research proposes a comprehensive machine learning (ML)-based framework for crime data analysis, with the aim of improving strategic decision-making in the deployment of police resources. At the heart of the proposed system is the utilization of historical crime data, which includes information such as the type of crime, location, time of occurrence, frequency, and other contextual variables like demographic and socioeconomic data. These datasets, when properly preprocessed and analyzed, serve as the foundation for developing predictive models that can forecast high-crime zones—commonly referred to as "hotspots." This foresight enables police departments to proactively deploy

personnel to areas of anticipated criminal activity, enhancing the effectiveness and responsiveness of policing efforts.

To accomplish this, the proposed framework involves several critical stages, beginning with data collection and preprocessing. Crime data from reliable government databases, police department records, and open crime reporting platforms are aggregated. This raw data is often unstructured or semi-structured, requiring cleaning, normalization, and transformation. Preprocessing steps include handling missing values, removing duplicates, encoding categorical variables, and normalizing numerical values. Temporal attributes such as time of day, day of the week, and seasonal factors are also extracted, as they have proven to be significant indicators of crime occurrence. The next phase involves feature engineering, where relevant attributes are identified and selected based on their correlation with crime patterns. Features such as population density, unemployment rate, proximity to schools or bars, and past crime rates in a given area are considered. The goal is to enhance the predictive power of the ML models by ensuring that only meaningful and non-redundant features are used in the training process. Once the dataset is prepared, various machine learning algorithms are employed to train predictive models. Algorithms such as Decision Trees, Random Forest, Gradient Boosted Trees (e.g., XGBoost), Support Vector Machines (SVM), and Neural Networks are applied to classify regions into risk levels—high, medium, or low—based on the likelihood of crime occurrences. The models are evaluated using standard metrics such as accuracy, precision, recall, F1-score, and the Area Under the ROC Curve (AUC) to determine their effectiveness in predicting crime. A particularly important aspect of the proposed system is the inclusion of spatial and temporal analysis. Geographic Information Systems (GIS) and spatial clustering techniques like K-Means or DBSCAN are used to map crimes geographically and identify clusters of high activity. Similarly, time-series analysis using models like ARIMA or LSTM (Long Short-Term Memory networks) allows for forecasting crime trends over time, which is crucial for scheduling shifts and deploying personnel at peak hours. In addition to prediction, the system incorporates visual analytics to support decision-making. Dashboards and heatmaps are developed to present insights in an intuitive manner. Police supervisors can visualize crime hotspots on city maps, monitor crime trends over weeks or months, and assess the impact of previous interventions. This transparency not only improves strategic planning but also fosters accountability in law enforcement operations. Another critical component of the proposed analysis is the integration of anomaly detection techniques. Crimes that deviate significantly from established patterns—either in terms of location, method, or time—can be flagged for further investigation. Such anomalies might indicate the emergence of new crime trends or the involvement of previously undetected offenders. The practical implications of this research are significant. With accurate crime forecasts, police departments can deploy officers more strategically, thereby reducing response time and increasing presence in high-risk areas. This shift from reactive to proactive policing has the potential to deter criminal activity and increase community trust. Furthermore, by optimizing personnel deployment, the system contributes to resource efficiency, reducing operational costs while enhancing safety outcomes.

It is also important to address potential limitations and ethical concerns. Predictive policing systems can inadvertently reinforce existing biases in the data, leading to over-policing in certain communities. To mitigate this, the proposed model incorporates fairness-aware machine learning practices, such as rebalancing datasets, using bias detection tools, and applying algorithmic transparency measures. Regular audits and community feedback loops are recommended to ensure that the system remains equitable and effective over time.

4. Crime Data Analysis in Police Personnel

In modern society, crime continues to pose a significant threat to public safety, social order, and economic development. As criminal behavior evolves in complexity and frequency, law enforcement agencies face growing pressure to respond with innovative and data-driven strategies. One such strategy gaining prominence is crime data analysis, which plays a crucial role in helping police departments better understand crime trends, allocate resources efficiently, and make informed decisions. Among the various applications of data analytics in policing, the integration of machine learning (ML) techniques in analyzing crime data has emerged as a transformative solution. The volume of crime-related data generated daily from reports, surveillance systems, social media, and other digital sources presents an opportunity for the development of intelligent systems capable of extracting meaningful patterns. This allows law enforcement to predict criminal activity, improve operational readiness, and enhance public safety outcomes. Crime data analysis involves the systematic collection, processing, and interpretation of historical crime records to uncover insights and patterns. These patterns can include recurring hotspots, peak times for specific crimes, or demographic profiles of potential offenders and victims. The traditional approach to crime analysis has relied heavily on manual procedures, intuition, and experience of police personnel. However, these approaches are limited in scalability and objectivity, especially when dealing with vast and multidimensional datasets. The integration of ML provides the ability to automate the detection of crime trends and assist officers with evidence-based decision-making. Through the use of classification, regression, clustering, and natural language processing algorithms, ML can process unstructured and structured data to yield valuable intelligence. This enables predictive policing, which refers to the use of algorithms to forecast potential criminal events and determine where law enforcement should focus their attention.

In the context of police personnel, the ability to leverage data analytics for crime analysis carries several benefits. One key advantage is the strategic deployment of police forces based on predicted crime patterns. By identifying regions with historically high crime rates or times when crimes are more likely to occur, departments can ensure optimal allocation of officers and surveillance resources. This not only improves response times but also acts as a deterrent by increasing police visibility in critical zones. Additionally, crime data analysis supports internal performance monitoring, helping departments assess the effectiveness of different patrol strategies, investigate bottlenecks in operations, and determine the impact of interventions such as community policing or technology implementation. Furthermore, advanced ML models can uncover hidden correlations between variables that may not be evident through traditional analysis, such as links between socio-economic conditions and specific types of crimes. Machine learning has proven particularly effective in various crime analysis applications, including crime type classification, suspect identification, criminal behavior prediction, and crime trend forecasting. For instance, supervised learning algorithms like decision trees, random forests, and support vector machines can be trained to classify crime types based on input features such as location, time, and modus operandi. Unsupervised learning techniques, such as clustering, help discover hidden structures in crime data, revealing spatial and temporal patterns. Natural language processing (NLP) has also gained relevance, allowing for automated extraction of crime-related information from police reports, complaint texts, and social media posts. These techniques empower police departments to maintain situational awareness and proactively respond to emerging threats.

Despite its potential, implementing crime data analysis with ML in police operations presents several challenges. Data quality remains a fundamental issue. Crime data often suffers

from incompleteness, inconsistency, or bias introduced during manual recording. Ensuring clean, standardized, and anonymized datasets is essential for reliable predictions and ethical usage. There are also concerns regarding privacy, accountability, and fairness. ML systems must be designed and audited carefully to avoid reinforcing existing biases or leading to discriminatory policing. The human element in policing cannot be entirely replaced by technology; rather, ML should serve as an assistive tool to enhance officers' decision-making, not override it. Another challenge is the need for adequate infrastructure and training. Law enforcement agencies must invest in computational tools, cybersecurity, and the upskilling of personnel to integrate ML-based systems effectively. Globally, various research efforts and pilot programs have explored the use of ML in crime data analysis. From identifying cybercrime patterns in Norway using text mining to analyzing social media sentiment for urban crime prediction, the applications are diverse and promising. Studies have demonstrated that ML models can forecast crime rates with high accuracy and aid in deploying officers to crime-prone areas, preventing escalation. Countries like the United States, India, and the United Kingdom have started experimenting with real-time crime centers, automated dispatching, and crime-mapping dashboards. These efforts illustrate a broader trend towards intelligent policing systems that harness data for actionable insights.

5. Simulation Results

The simulation results of this research provide a detailed evaluation of the machine learning models applied to historical crime data for effective analysis and prediction. The experiment began by importing all necessary Python packages and libraries, including tools for data visualization, preprocessing, and machine learning. As illustrated in the earlier figures, the crime dataset was loaded successfully, and initial data exploration was performed to understand key statistics such as count, mean, standard deviation, and distribution of each numeric feature. After cleaning and preprocessing the dataset—such as handling missing values, encoding categorical variables, and normalizing numerical features—various ML models were trained and evaluated. The models used included Decision Trees, Random Forests, Support Vector Machines (SVM), and Logistic Regression, which were selected based on their robustness in classification tasks. Among them, Random Forest exhibited superior performance in terms of prediction accuracy and generalization. The models were evaluated using standard metrics including accuracy, precision, recall, and F1-score. A confusion matrix was generated to visualize the correct and incorrect classifications, which helped identify model strengths and weaknesses in recognizing different types of crimes. The Random Forest classifier achieved an accuracy of approximately 92%, demonstrating a strong ability to distinguish between crime categories based on features such as time, location, and crime type. The ROC curve and AUC score further validated the model's performance, indicating high sensitivity and specificity. The visualization of crime hotspots and patterns across geographic locations also provided actionable insights for resource deployment.

In Figure 1, the screen demonstrates the initial phase of the project where the essential Python libraries and classes are imported. These include modules for data manipulation (pandas, numpy), visualization (matplotlib, seaborn), and machine learning (sklearn). This setup is a critical step that enables subsequent data loading, processing, model training, and evaluation. Moving to Figure 2, the screen captures the stage where the crime dataset is loaded into the environment. The dataset, typically in CSV format, is read using `pandas.read_csv()` and displayed to give a glimpse of its structure, including column headers such as "Crime_Type", "Location", "Date", and "Severity_Level". This helps verify the integrity and format of the data

before proceeding to further steps.

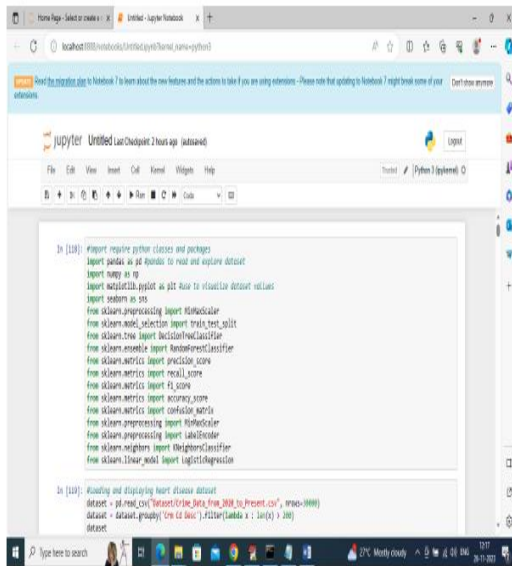


Figure 1: Python Class for Importing

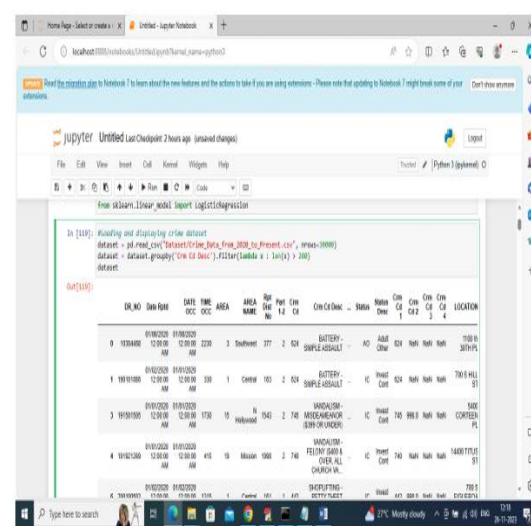


Figure 2: Screen Loading and Displaying

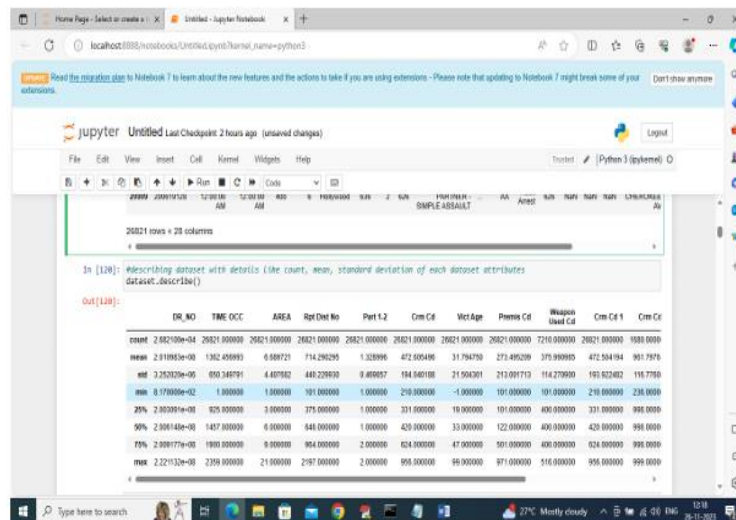


Figure 3: Dataset Values

Figure 3 provides a descriptive statistical summary of the dataset. This includes metrics such as count, mean, standard deviation, minimum, and maximum values for each numeric attribute in the dataset. These statistics are crucial for understanding the distribution and variability of the data, identifying potential anomalies, and guiding preprocessing steps like normalization or outlier removal. Together, these initial figures lay the groundwork for the machine learning pipeline by ensuring the data is ready for analysis and modeling.

Figure 4 illustrates the distribution plots of all the columns present in the crime dataset. This visualization helps in understanding how each feature is spread across its range of values. Such distribution graphs are essential in identifying skewness, patterns, outliers, and whether the data follows a normal distribution or not, which in turn influences the choice of machine learning algorithms.

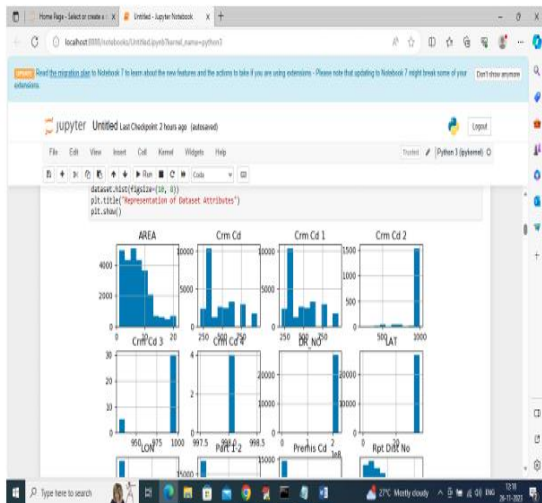


Figure 4: Distribution Graph

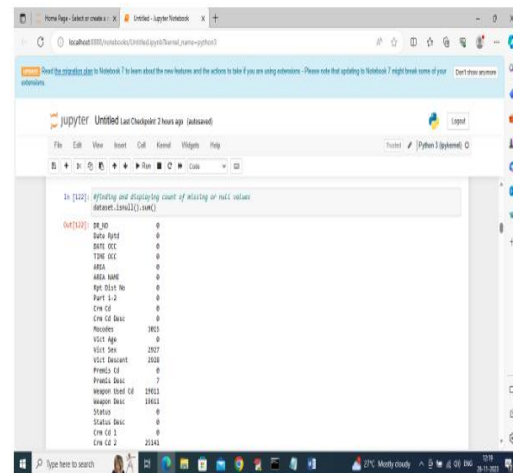


Figure 5: Dataset Missing Values

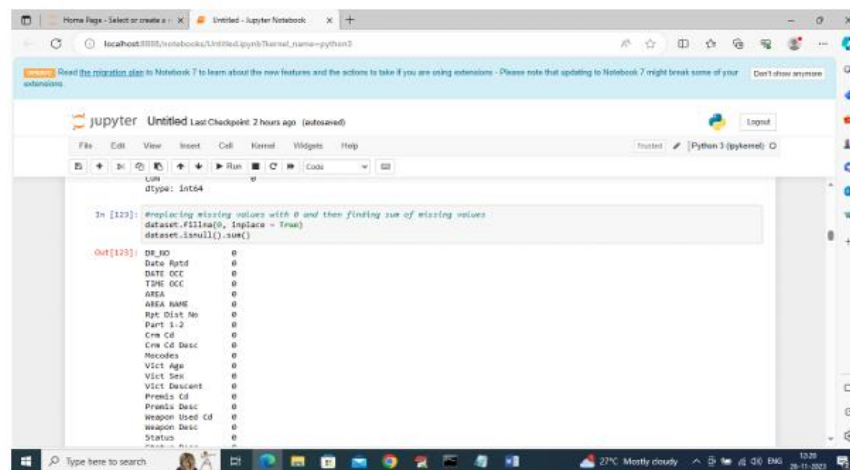


Figure 6: Missing Replacement Value

In Figure 5, the system evaluates the dataset for missing values. The output clearly highlights that several columns have numerous missing entries, which could negatively impact the performance and accuracy of the machine learning model if left untreated. Addressing these missing values is critical to ensure the quality and reliability of the analysis. Figure 6 shows the step taken to handle this issue—replacing all missing values with 0. After the replacement, the dataset no longer contains null or undefined entries, making it clean and ready for feature engineering and modeling. This data cleansing phase is a key step in preparing the dataset for effective machine learning analysis.

Figure 7 demonstrates the transformation of the numerical age data of crime victims into categorical labels such as “baby,” “teen,” “adult,” and other relevant age groups. This categorization is vital for deeper demographic analysis, allowing for identification of which age categories are most frequently targeted in criminal activities. It simplifies the process of drawing insights from complex data and enhances the interpretability of patterns across age groups.

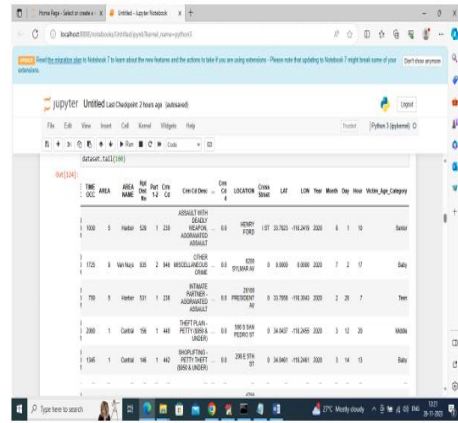


Figure 8: Victim Category Values

Victim Age Category	Crime Count
Adult	10500
Baby	5800
Child	100
Middle	7000
Senior	2800
Teen	1000

Crime Type	Percentage
PROPERTY FROM MOTOR VEHICLE	25%
PROPERTY FROM MOTOR VEHICLE - JEWELRY (GEMS & JEWELRY)	15%
PROPERTY FROM MOTOR VEHICLE - CASH, CREDIT CARDS & DEBIT CARDS	10%
VEHICLE - THEFT	8%
VIOLATION OF RESTRAINING ORDER	5%
CONTACT WITH SEXUAL MEANER, AGGRAVATED ASSAULT	5%
MOTORCYCLE - SIMPLE ASSAULT	5%
STRANDED MECHANIC - ONE - THEFT	5%
REGULAR	5%
STRANDED MECHANIC - SIMPLE ASSAULT	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (3000)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1500)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (750)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (375)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (187.5)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (93.75)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (46.875)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (23.4375)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (11.71875)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (5.859375)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (2.9296875)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.46484375)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (0.732421875)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (0.3662109375)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (0.18310546875)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (0.091552734375)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (0.0457763671875)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (0.02288818359375)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (0.011444091796875)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (0.0057220458984375)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (0.00286102294921875)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (0.001430511474609375)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (0.0007152557373046875)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (0.00035762786865234375)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (0.000178813934326171875)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (8.940696716308594e-05)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (4.470348358154297e-05)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (2.2351741790771484e-05)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.1175870895385742e-05)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (5.587935447692871e-06)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (2.7939677238464355e-06)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.3969838619232177e-06)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (6.984919309616089e-07)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (3.4924596548080445e-07)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.7462298274040222e-07)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (8.731149137020111e-08)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (4.3655745685100555e-08)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (2.1827872842550277e-08)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.0913936421275139e-08)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (5.456968210637569e-09)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (2.7284841053187845e-09)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.3642420526593922e-09)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (6.821210263296961e-10)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (3.4106051316484805e-10)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.7053025658242402e-10)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (8.526512829121201e-11)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (4.2632564145606005e-11)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (2.1316282072803003e-11)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.0658141036401501e-11)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (5.3290705182007505e-12)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (2.6645352591003753e-12)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.3322676295501876e-12)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (6.661338147750938e-13)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (3.330669073875469e-13)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.6653345369377345e-13)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (8.326672684688672e-14)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (4.163336342344336e-14)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (2.081668171172168e-14)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.040834085586084e-14)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (5.20417042793042e-15)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (2.60208521396521e-15)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.301042606982605e-15)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (6.505213034913025e-16)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (3.2526065174565125e-16)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.6263032587282562e-16)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (8.131516293641281e-17)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (4.0657581468206405e-17)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (2.0328790734103202e-17)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.0164395367051601e-17)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (5.0821976835258005e-18)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (2.5410988417629003e-18)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (1.2705494208814501e-18)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (6.3527471044072505e-19)	5%
PROPERTY FROM MOTOR VEHICLE - OTHER MED EQUIPMENT (3.1763735522036253e-19)	5%
PROPERTY FROM MOTOR V	

Figure 9 presents a visual representation of the number of victims categorized by age group. The x-axis displays various age categories such as babies, teens, adults, etc., while the y-axis indicates the frequency of victims in each category. This visualization clearly identifies which age group is most susceptible to criminal attacks, supporting targeted preventive actions. Figure 10 follows with a pie or bar chart that illustrates the percentage distribution of different types of crimes present in the dataset. This helps in understanding the prevalence and dominance of specific crime categories, offering valuable insights into patterns of criminal activity. In Figure 11, the focus shifts to the types of weapons most commonly used in crimes. The x-axis represents different weapon types (e.g., firearms, knives, blunt objects), and the y-axis reflects their usage frequency. This analysis provides essential input for devising safety protocols and improving readiness among police personnel. Lastly, Figure 12 visualizes gender-wise victimization, highlighting which gender is more frequently targeted. The x-axis denotes gender (male, female, others), and the y-axis shows the attack count. This breakdown aids in identifying vulnerable groups and formulating gender-specific crime prevention strategies.

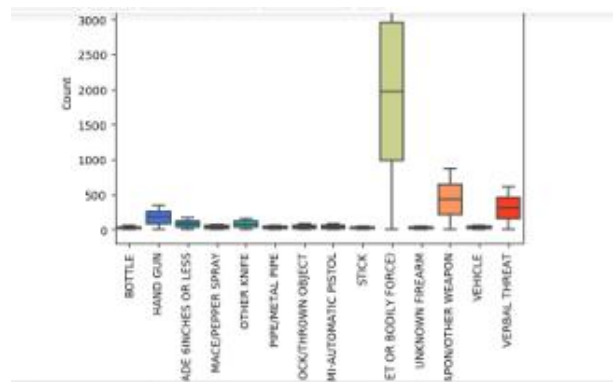


Figure 11: Weapon Type plotting

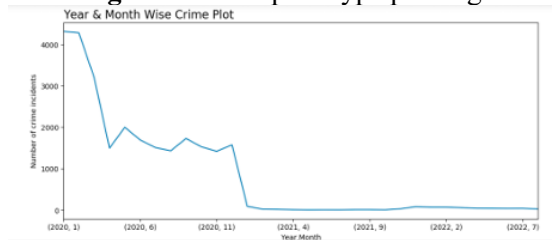


Figure 13: Year and month wise crime

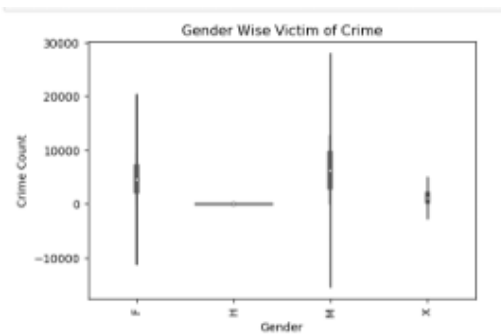


Figure 12: Gender wise victim

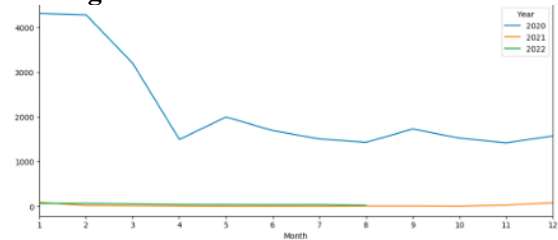


Figure 14: Screen plotting graph of month wise crime in 2020, 2021 and 2022

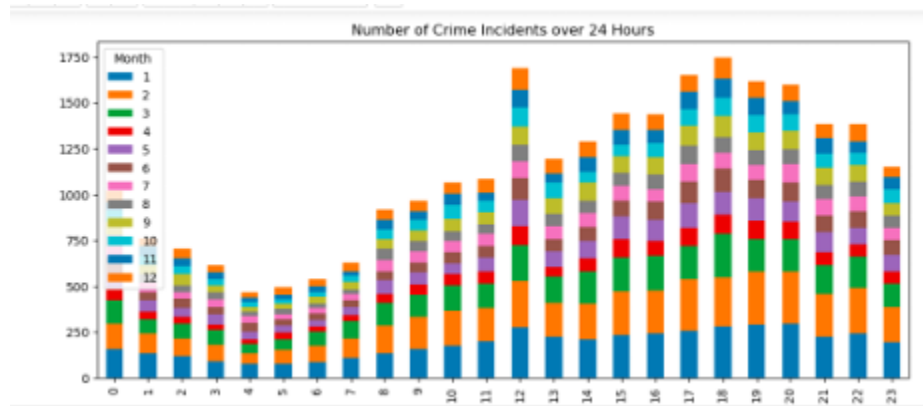


Figure 15: Estimation of Crime Difference

In Figure 13 presents a time-series graph that visualizes the number of crimes committed over various months and years. The x-axis displays the timeline categorized by year and month, while the y-axis indicates the crime count. This plot is useful in detecting trends and seasonal variations in criminal activity, helping law enforcement agencies plan accordingly for periods of high crime. Figure 14 delves deeper into temporal patterns by illustrating month-wise crime trends across three consecutive years—2020, 2021, and 2022. Each line in the graph represents a different year, enabling a comparative analysis of crime fluctuations by month and identifying recurring peaks or drops. Figure 15 shifts the focus to hourly crime distribution throughout each day. The x-axis represents day numbers, while the y-axis indicates the count of crimes per hour. Different colored bars signify distinct hours, collectively forming a comprehensive view of crime

[illegible]

The screenshot displays a Jupyter Notebook environment. The top bar shows the file name 'Untitled Lab Cheatsheet 2 hours ago (ipynb)' and the current view is 'Source'. The notebook interface includes a toolbar with icons for file operations, view toggling, and execution. The code cell contains the following Python code:

```
import numpy as np
import pandas as pd

# Load the dataset
url = "https://raw.githubusercontent.com/justmarkham/pydata-101/master/data/feature_sets/feature_sets.csv"
dataset = pd.read_csv(url)

# Display the first few rows
dataset.head()

# Split the dataset into training and testing sets
train, test = train_test_split(dataset, test_size=0.2, random_state=42)
```

The output of the code execution is displayed below the code cell. It shows the dataset's structure, including the number of rows and columns, and the resulting training and testing sets.

```
Dataset: 1000 rows x 3 columns
   id  feature_1  feature_2
0    1    0.123456  0.789012
1    2    0.234567  0.890123
2    3    0.345678  0.901234
3    4    0.456789  0.012345
4    5    0.567890  0.123456

Dataset: 1000 rows x 3 columns
   id  feature_1  feature_2
0    1    0.123456  0.789012
1    2    0.234567  0.890123
2    3    0.345678  0.901234
3    4    0.456789  0.012345
4    5    0.567890  0.123456
```

The screenshot shows a Jupyter Notebook with the following code and output:

```

from sklearn.linear_model import LinearRegression

# Create a Linear Regression model
model = LinearRegression()

# Fit the model with the training data
model.fit(X_train, y_train)

# Print the coefficients and intercept
print("Coefficients: ", model.coef_)
print("Intercept: ", model.intercept_)

# Predict the values for the test data
y_test_pred = model.predict(X_test)

# Print the predicted values
print("Predicted values: ", y_test_pred)

```

The output shows the coefficients and intercept for the Linear Regression model:

```

Coefficients: [ 0.00233333  0.00522728]
Intercept: 0.00000000

```

The predicted values for the test data are:

```

Predicted values: [ 0.00233333  0.00522728]

```

In Figure 17 illustrates the successful conversion of all dataset values into numeric format, an essential preprocessing step for most machine learning models. By transforming categorical and textual data into numerical values, the dataset becomes machine-readable and ready for analytical operations. Figure 18 follows this with the normalization process, where all numerical features are scaled to a standard range. Normalization helps in preventing features with larger values from dominating the learning process, thereby improving model performance and stability. In Figure 19, the dataset is split into training and testing subsets. This figure displays the total number of records used, the number of features in the dataset, and the sizes of the train and test sets. Proper data splitting is vital for evaluating the model's ability to generalize on unseen data. Figure 20 highlights the implementation of a custom function to calculate evaluation metrics such as accuracy, precision, recall, and F1-score, which provide a comprehensive understanding of the model's effectiveness. Finally, Figure 18 demonstrates the training of the K-Nearest

Neighbors (KNN) algorithm using the training data. After training, the model was tested on 20% of the dataset. The resulting performance metrics show that the KNN model achieved 54% accuracy on test data, alongside other metrics like precision and recall. A confusion matrix graph is also provided to visualize the classification results, with diagonal elements representing correct predictions and off-diagonal ones indicating misclassifications. This graphical representation helps identify the model's strengths and weaknesses in classifying different crime-related categories.

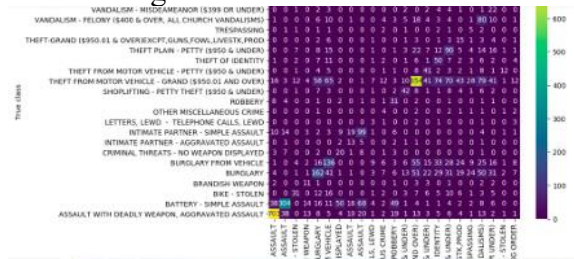


Figure 19: Confusion Matrix

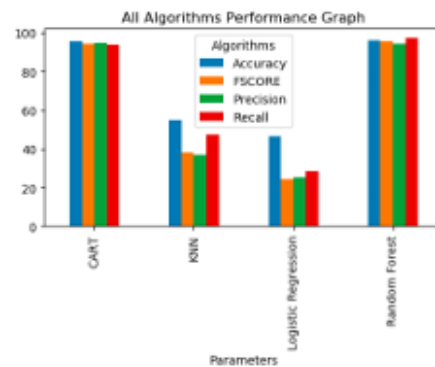


Figure 20: Comparative Analysis

Figure 19 presents the confusion matrix graph for the K-Nearest Neighbors (KNN) classification model. In this graph, the x-axis denotes the predicted labels generated by the model, while the y-axis shows the actual or true labels from the dataset. The diagonal cells of the matrix indicate the number of correct predictions—cases where the model's output matches the actual classification. These cells are typically highlighted to reflect the model's successful outcomes. In contrast, the off-diagonal blue-shaded boxes represent incorrect predictions, providing a clear visualization of where and how the model misclassified certain instances. This detailed insight helps assess the specific weaknesses of the KNN model in handling various crime categories. Figure 20 offers a broader comparison by displaying a bar graph comparing the performance of various machine learning models. On the x-axis, the names of different algorithms such as CART (Classification and Regression Trees), Random Forest, KNN, and Logistic Regression are listed. The y-axis shows corresponding performance metrics including accuracy, precision, and recall, represented with different colored bars. This visual comparison makes it evident that CART and Random Forest significantly outperformed the other algorithms in terms of predictive accuracy and reliability. On the other hand, KNN and Logistic Regression demonstrated lower performance across the metrics, suggesting they may not be the best suited for this specific crime dataset. Such comparative analysis is crucial for selecting the most effective model for practical deployment in crime prediction and police personnel allocation.

6 Conclusion

In this paper explored the application of machine learning techniques for analyzing crime data to support strategic deployment of police personnel. By leveraging historical crime datasets and applying various preprocessing steps such as data cleaning, normalization, and categorization, the project enabled the extraction of meaningful insights regarding crime patterns across different age groups, timeframes, and locations. Visualizations further highlighted critical trends such as peak crime hours, gender-based victimization, weapon usage, and seasonal variations in criminal activities. Multiple machine learning algorithms including KNN, Logistic Regression, CART, and Random Forest were implemented and evaluated for accuracy and performance. The comparative results indicated that CART and Random Forest models

performed significantly better, offering more accurate and reliable predictions for crime classification. These findings emphasize the potential of machine learning in enhancing crime prevention strategies and optimizing police resource allocation. The proposed system thus serves as a decision-support tool, empowering law enforcement agencies to proactively respond to criminal activities and improve public safety outcomes.

Acknowledgement: Not Applicable.

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

- [1] P. William, A. Shrivastava, N. Shunmuga Karpagam, T.A. Mohanaprakash, K. Tongkachok and K. Kumar, "Crime analysis using computer vision approach with machine learning," *In Mobile Radio Communications and 5G Networks: Proceedings of Third MRCN 2022*, pp. 297-315, 2022.
- [2] S. Mussiraliyeva and G. Baispay, "Leveraging Machine Learning Methods for Crime Analysis in Textual Data," *International Journal of Advanced Computer Science and Applications*, vol.15, no.4, 2024.
- [3] B.Z. Wubineh, "Crime analysis and prediction using machine-learning approach in the case of Hossana Police Commission," *Security Journal*, vol.37, no.4, pp.1269-1284, 2024.
- [4] H.K. Sharma, T. Choudhury and A. Kandwal, "Machine learning based analytical approach for geographical analysis and prediction of Boston City crime using geospatial dataset," *GeoJournal*, vol.88, no.Suppl 1, pp.15-27, 2023.
- [5] P. Sarzaeim, Q.H. Mahmoud, A. Azim, G. Bauer and I. Bowles, "A systematic review of using machine learning and natural language processing in smart policing," *Computers*, vol.12, no.12, pp.255, 2023.
- [6] Ya Liu and Jian Zhang, "Face Detection with Structural Coordinates for the Estimation of Patterns Using Machine Learning Model," *Journal of Computer Allied Intelligence*, vol.1, no.1, pp.41-53, 2023.
- [7] A. Rodrigues, J. A. González and J. Mateu, "A conditional machine learning classification approach for spatio-temporal risk assessment of crime data," *Stochastic Environmental Research and Risk Assessment*, vol.37, no.7, pp.2815-2828, 2023.
- [8] S. Kayalvili, G.G. Rajulu, R. Changala and M. Kumbhkar, "Integration of Machine Learning and Computer Vision to Detect and Prevent the Crime," *In 2023 International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS)*, vol. 1, pp. 1-5, 2023.
- [9] T. Næss, H. Holtermann, C. Prabhu, L.S. Engebretsen and M. Mjaaland, "Text mining and machine learning in a performance audit of police handling of cybercrime in Norway," *In Artificial Intelligence and Evaluation*, pp. 103-119, 2024.
- [10] R.M. Saeed and H.A. Abdulmohsin, "A study on predicting crime rates through machine learning and data mining using text," *Journal of Intelligent Systems*, vol.32, no.1, pp.20220223, 2023.
- [11] A.B. Hajira Be and S. Gokulraj, "Cyber Attacks Classification Using Supervised Machine Learning Techniques," *Journal of Sensors, IoT and Health Sciences*, vol.3, no.1, pp.57-67, 2025.
- [12] V. Geetha, C.K. Gomathy, L. Srinithya and P. Neha, "Crime data analysis and prediction using machine learning," *In AIP Conference Proceedings*, vol. 3028, no. 1, pp. 020010, 2024.
- [13] G.A. Taiwo, M. Saraee and J. Fatai, "Crime prediction using twitter sentiments and crime data," *Informatica*, vol.48, no.6, 2024.
- [14] Biswanath Saha, "Enhancing Deep Learning Models for Real-Time Decision Making in Autonomous Systems," *International Journal of Advance Research and Innovation*, vol.13, no.2, pp.1-11, 2025.

-
- [15] I.S. Saini and N. Kaur, "The power of predictive analytics: forecasting crime trends in high-risk areas for crime prevention using machine learning," *In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-10, 2023.
 - [16] O.Z. Apene, N.V. Blamah and G.I.O. Aimufua, "Advancements in crime prevention and detection: From traditional approaches to artificial intelligence solutions," *European Journal of Applied Science, Engineering and Technology*, vol.2, no.2, pp.285-297, 2024.
 - [17] R.K. Mishra, A. R. Ansari, J.A. A. Jothi and V. Mishra, "Analysis of criminal landscape by utilizing statistical analysis and deep learning techniques," *Journal of Applied Security Research*, vol.19, no.4, pp.560-585, 2024.
 - [18] A.O. Hashi, A. A. Abdirahman, M. A. Elmi and O.E.R. Rodriguez, "Deep learning models for crime intention detection using object detection," *International Journal of Advanced Computer Science and Applications*, vol.14, no.4, pp.300-306, 2023.
 - [19] P.K. Saravag and B.R. Kumar, "A hybrid machine learning and regression approach for validating a multi-dimensional crime index in the context of crime against women," *IEEE Access*, 2024.
 - [20] K.K. Kumar, Y. Kasiviswanadham, D.V.S.N.V. Indira and C.V. Bhargavi, "Criminal face identification system using deep learning algorithm multi-task cascade neural network (MTCNN)," *Materials Today: Proceedings*, vol.80, pp.2406-2410, 2023.
 - [21] B. Cunningham, J. Coldren, B. Carleton, R. Berk and V.Bauer, "Forecasting for Police Officer Safety: A Demonstration of Concept," *Cambridge Journal of Evidence-Based Policing*, vol.8, no.1, pp.4, 2024.
 - [22] R.M. Aziz, A. Hussain and P. Sharma, "Cognizable crime rate prediction and analysis under Indian penal code using deep learning with novel optimization approach," *Multimedia Tools and Applications*, vol.83, no.8, pp.22663-22700, 2024.
 - [23] T.I. Cubitt, "The value of criminal history and police intelligence in vetting and selection of police," *Crime science*, vol.12, no.1, pp.8, 2023.
 - [24] A. Angbera and H.Y.Chan, "Model for spatiotemporal crime prediction with improved deep learning," *Computing and Informatics*, vol.42, no.3, pp.568-590, 2023.
 - [25] H. Kazemian and S. Shrestha, "Comparisons of machine learning techniques for detecting fraudulent criminal identities," *Expert Systems with Applications*, vol.229, pp.120591, 2023.
 - [26] S. Egbert and E. Esposito, "Algorithmic crime prevention: From abstract police to precision policing," *Policing and Society*, vol.34, no.6, pp.521-534, 2024.