Enhanced Crime Hotspot Prediction and Visualization for Women's Safety through Deep Learning

C. Maneesh[1], M. Murugesh[1], S. Muthukumar[1], R. Harish [1], T. Arachelvi[2]
UG Students[1], Associate Professor[2]
Computer Science and Engineering[1], Electronics and Communication Engineering[2], Tamilnadu College of Engineering, Coimbatore, India

***Abstract*—The identification of crime hotspots—geographic places with a notably greater prevalence of criminal activity than nearby locations—is essential to preserving public safety and security. Hotspots where crimes like sexual harassment, assault, domestic violence, stalking, and human trafficking disproportionately affect women are especially worrisome. Better resource allocation, strategic policing, and community awareness are made possible by knowing and anticipating these hotspots. The goal of this project is to create a predictive system that analyzes extensive crime datasets, including variables like the type of crime, frequency, time of occurrence, and exact geographic location, using a Deep Explainable Decision Tree model. The technology can predict regions with a higher probability of crimes against women by analyzing and learning from past crime trends. Additionally, the explainable nature of the model improves interpretability and trust, enabling stakeholders to comprehend the elements that contribute to the formation of hotspots. By integrating Google Maps, the anticipated hotspots are shown, providing a user-friendly and interactive platform for citizens, urban planners, and authorities to plan preventive measures, monitor risk areas in real time, and ultimately help create a safer environment for women.**

***Keywords*—Crime Hotspot Prediction, Women's Safety, Deep Explainable Decision Tree, Crime Data Analysis, Google Maps Visualization**

I. INTRODUCTION

In many parts of the world, there is growing concern about the increase in crimes against women, including sexual harassment, assault, domestic abuse, stalking, and human trafficking. Enhancing public safety and safeguarding vulnerable people require identifying and addressing places where these crimes are more likely to occur. Certain geographic areas that see a noticeably higher frequency of criminal activity than other places are known as crime hotspots. Authorities can take proactive steps to avoid crimes and more efficiently allocate resources by anticipating certain hotspots.In order to anticipate crime hotspots that specifically target women's safety, a machine learning method utilizing a Deep Explainable Decision Tree model is proposed in this study. The system looks for trends that point to increased risks by examining past crime data, such as the kind of crime, how often it occurs, when it occurs, and where it occurs. Incorporating Google Maps into visualization offers a user-friendly and interactive method of showcasing anticipated hotspots, assisting law enforcement, urban planners, and the general public in making well-informed choices. In addition to mapping and forecasting crime-prone areas, this effort aims to use data-driven insights and contemporary technologies to help create safer surroundings for women.

1.1 CRIME HOTSPOT PREDICTION

A crucial tool in contemporary public safety management is crime hotspot prediction, which identifies regions with a high concentration of criminal activity. Authorities can predict possible danger areas before crimes happen by using machine learning models, spatial analytic techniques, and historical crime data. By anticipating crime hotspots, law enforcement organizations can better allocate resources, organize focused patrols, and launch community safety initiatives. Hotspot prediction in the context of this study focuses on pinpointing areas where women are more susceptible to crimes such as assault and harassment. Sophisticated prediction models aid in capturing temporal patterns, such as elevated risk at night or during particular events, in addition to incidence frequency. In both urban and rural areas, proactive interventions based on hotspot prediction can dramatically lower crime rates and improve women's perceptions of safety.

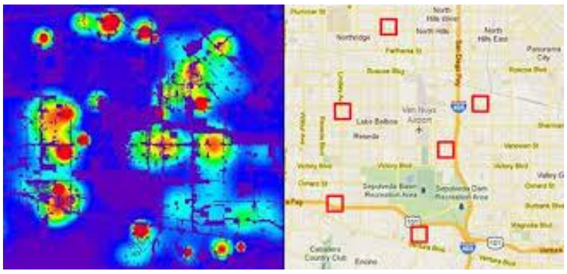

FIGURE 1. CRIME HOTSPOT PREDICTION

2. WOMEN'S SAFETY

Given the disproportionate incidence of crimes against women in public, private, and online settings, women's safety continues to be a significant global concern. Predictive, preventive, and educational tactics that empower women and discourage future perpetrators are necessary to ensure women's protection; reactionary measures alone are not enough. There are serious physical, emotional, and psychological repercussions for the kinds of crimes that target women, such as sexual harassment, assault, domestic abuse, stalking, and human trafficking. Technology-driven solutions like crime hotspot prediction, which offers vital information on areas where women are more likely to face threats, are used in this initiative to improve women's safety. Additionally, visualizing these areas can raise public awareness, assist women in making well-informed decisions regarding their travel routes and destinations, and persuade authorities to create safer urban environments by enhancing lighting, surveillance, and accessibility for public transportation.

FIGURE 2. WOMEN’S SAFETY

3. DEEP EXPLAINABLE DECISION TREE

A complex machine learning model called the Deep Explainable Decision Tree (DEDT) blends the interpretability and transparency of decision trees with the predictive ability of deep learning approaches. DEDT organizes its predictions in a way that is easy for humans to understand, in contrast to conventional deep learning models that function as opaque black boxes that make it challenging to comprehend how they make decisions. A logical choice based on input features such crime type, place, time, and frequency is represented by each branch and node in the tree. When it comes to crime hotspot prediction, DEDT enables authorities and technical users to comprehend the rationale for a specific location's designation as a high-risk area. In order to improve intervention tactics based on understandable insights rather than depending on opaque algorithms, this transparency is essential for maintaining public trust. Additionally, DEDT models can be updated as new crime data becomes available, increasing forecast accuracy over time.

4. CRIME DATA ANALYSIS

The methodical study of datasets containing information about criminal activity in order to identify patterns, correlations, and trends that are not immediately apparent is known as crime data analysis. Data points including the kind of crime, location coordinates, time stamps, victim demographics, and incident context are gathered for this research. To obtain significant insights, sophisticated statistical methodologies, visualization strategies, and machine learning algorithms are used. The prediction model in this project is constructed using the study of crime data. The technology determines hotspots (such as empty streets, dimly lit places), peak crime periods (such as weekends and evenings), and the most frequent offenses against women by examining historical data. Knowing these trends makes it easier to modify law enforcement's response, influence public policy, and advise the public about safety precautions. Ensuring data veracity and safeguarding victim privacy are two ethical factors that are emphasized in the proper processing of crime data.

5. GOOGLE MAPS VISUALIZATION

An effective method for interactively presenting spatial data is Google Maps visualization, which helps a broad audience understand complex datasets. This project allows users to zoom in, explore communities, and visually grasp danger levels by plotting forecasted crime hotspots right onto Google

Maps. In addition to making crime prediction data more easily accessible, this type of visualization helps the public, authorities, and urban planners make decisions in real time. A lady planning her trip, for example, can immediately consult the map to choose safer routes or steer clear of high-risk areas.The visual output allows law enforcement to more strategically assign patrol units. Additionally, Google Maps offers capabilities like location-based alerts, real-time updates, and integration with navigation apps that can be used to provide dynamic safety messages and raise user situational awareness.

FIGURE 3. GOOGLE MAP VISUALIZATION

II. LITERATURE REVIEW

2.1 MARKED POINT PROCESS HOTSPOT MAPS FOR HOMICIDE AND GUN CRIME PREDICTION IN CHICAGO

Mohle, George et al. Crime hotspot maps, as suggested in this study, are a popular and effective way to allocate police resources and show spatial crime patterns. However, only one crime type is frequently used to construct hotspot maps across a single timeline. Particularly for low frequency crimes like homicide, risk estimations suffer from a huge variance when using short-term hotspot maps that use several weeks of crime data. Near-repeat effects and new hotspot trends are not considered in long-term hotspot maps that use data spanning several years. In this study, we demonstrate how a marked point process technique can be used to extend point process models of crime to incorporate leading indicator crime types, while capturing both short-term and long-term patterns of risk. Accurate hotspot maps that can be utilized for predictive policing of gun-related crime are produced by methodically combining years' worth of data and a wide variety of crime types. We use the methodology on a sizable open source data collection that the Chicago Police Department has made publicly accessible online. With an application to predictive policing, we present a model in this work for predicting homicide and precursory gun offenses.

2.2 NATURAL LANGUAGE PROCESSING AND E-GOVERNMENT: EXTRACTING REUSABLE CRIME REPORT INFORMATION

This study makes a proposal by Gondy A. Leroy et al. It must be possible to report crimes around-the-clock. There are a number of alternative reporting options, from in-person reporting to internet submissions, even though 911 and tip-lines are the most well-known. Crime victims and witnesses can report occurrences to police at any time, from any location, using internet-based crime reporting systems. However, witnesses' memory recall is not well supported by the current email and text-based systems, which results in reports that are less accurate and contain less information. Additionally, these solutions do not make it easier to integrate and reuse the supplied data with other information systems.In order to gather pertinent crime information from witness accounts and ask follow-up questions based on that information, we are creating an anonymous online crime reporting system. We promote memory recall by using investigative interviews and natural language processing techniques, and we facilitate knowledge reuse by mapping the data straight to a database. We present the assessment of the system's Suspect Description Module (SDM). 70% (recall) of the information from witness narratives is accurately captured by our interface. The design and development techniques employed for this module will be applied to other modules.In the United States, millions of crimes are perpetrated annually. According to the Federal Bureau of Investigation, there were 1.38 million violent crimes and 10.3 million property crimes in 2003. But according to the Department of Justice, only a third of all property crimes and half of all violent crimes were reported in the same year. Police departments, law enforcement officials, and policy makers can better regulate violence and distribute resources (i.e., policies, budgets, legislation, and program evaluation) when more accurate information is available. This is why it is crucial for both residents and authorities to report crimes to the police.

2.3 NATURAL LANGUAGE PROCESSING BASED ON SEMANTIC INFERENTIALISM FOR EXTRACTING CRIME INFORMATION FROM TEXT

Pinheiro Vládia et al. This article outlines the architecture for web-based information extraction systems that are based on natural language processing (NLP) and specifically designed for the investigation of crime-related data. The NLP module, which is based on the Semantic Inferential Model, is the architecture's primary feature. By using the design to contribute to WikiCrimes, a collaborative web-based crime registration system, we show that the architecture is feasible. Textual narratives are one of the primary information sources that managers and analysts of public safety employ. These reports include details about crimes that have been committed as well as profiles of individuals that intelligence services keep an eye on. In order to generate knowledge and take practical steps to enhance public safety, these reports highlight the traits, quirks, and connections between the events and individuals. They also enable one to identify patterns. However, it takes a lot of effort and time to read the large amount of information in natural language. As a result, a Natural Language Processing (NLP) system that aids in the comprehension of these documents is extremely beneficial. Understanding meaning is limited to explicit textual input in the majority of NLP techniques now in use. In this research, we present a framework for information extraction based on the Semantic Inferential Model, or SIM, a novel model for natural language text expression and semantic analysis. This concept seeks to construct an NLP system with an additional text comprehension layer that provides reasoning on the text's explicit and implicit information. For instance, additional information can be deduced from the description of a crime, including the likely locations of crimes, the use of violence and weapons, involvement in gangs, the criminal's profile, etc. This information is frequently not made clear in the text and is only discernible through a network of premises and inferences from the sentences that are employed.

2.4 DEEP LEARNING FOR REAL-TIME CRIME FORECASTING AND ITS TERNARIZATION

According to Bao Wang et al., real-time crime predictions is crucial in this approach. It is challenging to make an exact prediction about the time and location of the next crime, though. Such a complicated system has no known physical model that can reasonably approximate it. There is a weak signal of interest and a lack of historical crime data in both space and time. We begin this task by providing an accurate depiction of crime data. In order to forecast the distribution of crime in Los Angeles at the hourly scale in neighborhood-sized parcels, we next modify the spatial temporal residual network on the well-represented data. These tests, together with comparisons to a number of other prediction methods currently in use, show how accurate the suggested model is. In order to solve the resource consumption problem for its practical implementation, we lastly introduce a ternarization technique. It is a significant scientific and practical challenge to forecast crime in micro-geographic regions at hourly or even finer temporal scales. Finding out where and when crimes are most likely to happen opens up new avenues for crime prevention. However, it is very difficult to forecast crime accurately at fine spatial temporal scales. Crime is influenced by a variety of intricate elements, many of which are impossible to quantify.

2.5 INTELLIGENT CRIME ANOMALY DETECTION IN SMART CITIES USING DEEP LEARNING

According to Sharmila Chackravarthy et al., this approach makes it clear that protecting any home depends on the prompt and precise identification of criminal activities. Crime detection system integration aims to increase security in light of smart cities' explosive growth. To accomplish this, traditional video surveillance has always been heavily relied upon. This frequently results in a backlog of video data that needs to be watched over by a supervisor. Error rates rise in large urban areas as a result of the supervisory officers' ever heavier workload. Workload reduction strategies have been put into place. Although they have a number of drawbacks, auto regressive models are currently utilized to more accurately predict criminal activity. We suggest a method for analyzing video stream data that combines neural networks with a Hybrid Deep Learning algorithm. The workload for the supervisory officials will be lessened as a result of our system's ability to swiftly recognize and evaluate illicit activities. An effective and flexible criminal detection system will be possible if it is integrated into the infrastructure of smart cities. It is now challenging to police and keep an eye on places with a high crime likelihood due to the recent population boom in urban areas. Insecurity and criminal activity have increased in certain regions as a result of this lack of control. There is a chance to develop innovative solutions to these issues as smart city infrastructure develops.

III. EXISTING SYSTEM

Crime and violations are intended to be managed because they pose a threat to justice. Computationally, accurate crime prediction and trends for the future can help improve urban safety. Early and precise crime prediction and forecasting are hampered by humans' limited capacity to process complicated information from huge data. Numerous computational opportunities and challenges arise from the precise calculation of the crime rate, types, and hot locations based on historical patterns. Even with extensive research, improved predictive algorithms are still required to guide police patrols toward criminal activity. Prior research on learning models for crime forecasting and prediction accuracy is weak. In order to better fit the crime data, this study used a variety of machine learning algorithms, including logistic regression, support vector machine (SVM), Naïve Bayes, k-nearest neighbors (KNN), decision trees, multilayer perceptrons (MLP), random forests, and extreme gradient boosting (XGBoost). Additionally, time series analysis was done using the autoregressive integrated moving average (ARIMA) model and long-short term memory (LSTM). Additionally, the key regions for both cities were used to further identify the outcomes of the crime predictions. All things considered, these findings are helpful in guiding police tactics and practice since they offer early detection of crime, hotspots with greater crime rates, and future trends with better predictive accuracy than with other approaches.

IV. PROPOSED SYSTEM

The goal of the suggested system is to create a sophisticated, data-driven platform that can identify and display crime hotspots that are especially relevant to the safety of women. The system examines past crime data using a Deep Explainable Decision Tree (DEDT) model to pinpoint regions where crimes including sexual harassment, assault, domestic abuse, stalking, and human trafficking are most likely to occur. To identify trends and generate precise forecasts, the model takes into account a number of variables, such as the type of crime, its frequency, the coordinates of the crime scene, and temporal elements like the time of day and the day of the week. The system incorporates Google Maps to visually represent anticipated hotspots, enabling users to interactively explore risk zones in order to improve accessibility and usability. In addition to enabling law enforcement to strategically allocate resources and carry out surveillance, this also enables citizens—particularly women—to be informed about locations to avoid or exercise caution in. By using an explainable model, decision-making is transparent and stakeholders can comprehend the considerations that go into each forecast. Furthermore, the system's dynamic design enables real-time updates and retraining in response to newly available crime data, thereby increasing its accuracy and dependability. This proactive strategy promotes safer surroundings and well-informed decision-making for personal awareness and public safety planning.

V. MODULE DESCRIPTION

1.Data Collection Module

This module is in charge of compiling crime-related datasets from dependable sources, including open crime databases, police records, and public safety reports. It gathers crucial data such as the type of offense, the time, date, and location (latitude and longitude). This data's quality and accuracy are essential for creating a predictive model that works.

2. Data Pre-processing Module

The raw criminal data is cleaned, formatted, and arranged by the data pre-processing module. It manages missing values, eliminates duplicate entries, normalizes the data for improved model training, and, if required, transforms categorical data into numerical formats. The dataset is kept in the best possible condition for analysis and machine learning thanks to this module.

3. Feature Extraction and Selection Module

Finding and removing significant features from the dataset that aid in crime prediction is the main goal of this module. Key characteristics are chosen, including crime type, frequency, time patterns, and geographic regions. To increase the prediction model's effectiveness and precision, feature selection techniques are employed to choose only the most pertinent attributes.

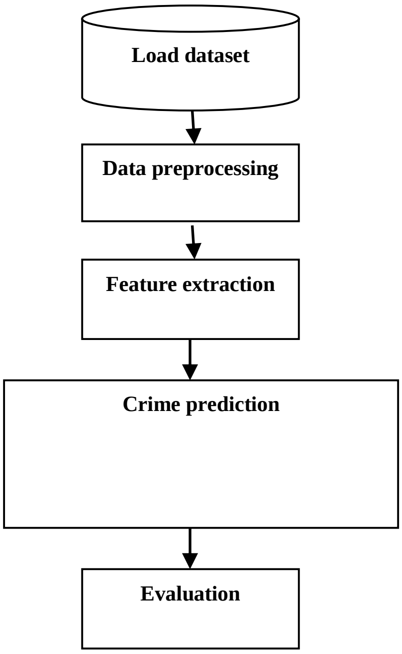

FIGURE 4. SYSTEM FLOW DIAGRAM

4. Crime Hotspot Prediction Module
This module examines the processed data and predicts crime hotspots using a Deep Explainable Decision Tree (DEDT) model. It makes forecasts about regions that are expected to see greater rates of crimes against women by identifying trends from past crime incidents. Users can better comprehend the reasoning behind each prediction thanks to the model's explainable nature.

5. Visualization Module (Google Maps Integration)
This module plots the anticipated hotspot sites on Google Maps in an interactive manner. Users can zoom in to various locations, see the crime-prone zones visually, and have a better grasp of how crimes are distributed spatially. Both the public and authorities will find the system easier to use and more useful as a result of this visualization.

6. User Interface Module
Users, law enforcement, and researchers can interact with the system through the user interface (UI) module's straightforward and user-friendly front end. It shows maps, gives users access to comprehensive hotspot data, and could have other features like notifications for nearby police stations or route safety monitoring. The user interface is made to be accessible and easy to use.

7. Model Update and Retraining Module
This module allows the system to retrain the Deep Explainable Decision Tree model and update its dataset as fresh crime data becomes available. Frequent updates make the system dynamic and future-proof by ensuring that predictions remain accurate and take into account the most recent trends in crime.

VI. RESULT ANALYSIS

The effectiveness of employing machine learning, namely the Deep Explainable Decision Tree model, in identifying areas with a higher likelihood of crimes against women is demonstrated by the result analysis of the suggested crime hotspot prediction system. Several key hotspot areas were accurately identified by the algorithm after it was trained and tested using historical crime data. Plotting the anticipated hotspots on Google Maps gave the results an easy-to-understand visual representation, which facilitated result interpretation. The model's reliability was confirmed by a significant correlation between the projected hotspots and actual recorded crime areas. Additionally, the decision tree's explainable structure made it possible to comprehend the variables affecting each prediction, including crime type, geographic patterns, and time of occurrence. The analysis provided useful information for law enforcement and the public at large by highlighting the fact that crimes against women were more common in specific locations on weekends and in the late evenings. The system may greatly aid in proactive safety measures, well-informed decision-making, and resource planning to improve women's safety in susceptible areas, according to the results analysis conducted overall.

VII. CONCLUSION

In conclusion, by precisely pinpointing high-risk regions that are likely to experience criminal activity, the suggested crime hotspot prediction system provides a creative and practical way to improve women's safety. The system's implementation of the Deep Explainable Decision Tree model results in trustworthy forecasts as well as decision-making transparency, which increases user trust. Both authorities and the general public may easily visualize expected hotspots because to the system's interaction with Google Maps. As new crime data becomes available, the model's dynamic retraining and real-time updates guarantee its continuous applicability. All things considered, this system is a useful instrument for people, law enforcement, and urban planners, assisting in the decrease of crime, enhancement of public safety, and creation of safer surroundings for women in high-risk locations.

VIII. FUTURE WORK

Future research could improve the suggested crime hotspot prediction system by adding more data sources that could affect the frequency and distribution of crimes, like the weather, public gatherings, or socioeconomic variables. Prediction accuracy and responsiveness could be further increased by enlarging the dataset to incorporate real-time data from citizen reports, social media feeds, and surveillance systems. Furthermore, including more advanced machine learning models—like deep learning neural networks—may improve the model's capacity to identify minute patterns in huge datasets. Predictive analytics for particular crime categories might potentially be added to the system, providing specialized solutions to stop crimes like domestic abuse and human trafficking. Notifications, safety alarms, or automatic suggestions for safer routes based on real-time risk assessments might also be included to the user interface as part of an optimization. To encourage more community involvement and participation in crime prevention initiatives, a mobile application that provides users with on-the-go access to crime hotspot data and safety advice might be created.

REFERENCES

[1] G. Mohler, "Homicide and gun crime prediction in Chicago using marked point process hotspot maps," Int. J. Forecasting, vol. 30, no. 3, pp. 491–497, July 2022.

[2] A. Iriberri and G. Leroy, IEEE Int. Conf. Inf. Reuse Integr., Las Vegas, IL, USA, Aug. 2020, pp. 221–226., "Natural language processing and e-government: Extracting reusable crime report information."

[3] "Natural language processing based on semantic inferentialism for extracting crime information from text," by V. Pinheiro, V. Furtado, T. Pequeno, and D. Nogueira, in Proceedings of the IEEE Int. Conf. Intell. Secur. Informat., Vancouver, BC, Canada, May 2021, pp. 19–24.

[4] "Deep learning for real-time crime forecasting and its ternarization," by B. Wang, P. Yin, A. L. Bertozzi, P. J. Brantingham, S. J. Osher, and J. Xin Chin. Ann. Math., B, vol. 40, no. 6, pp. 949–966, Nov. 2020.

[5] S. Chackravarthy, S. Schmitt, and L. Yang, "Deep learning-based intelligent crime anomaly detection in smart cities," in Proceedings of IEEE 4th International Conference on Collaboration Internet Computing (CIC), Philadelphia, PA, USA, October 2021, pp. 399–404.

[6] H.-W. and H.-B. Kang, "Deep learning for predicting the occurrence of crimes from multimodal data," PLoS ONE, vol. 12, no. 4, April 2019, art. no. e0176244.

[7] "Proposing a hybrid approach mobile apps with big data analysis to report and prevent crimes," by A. Fidow, M. Hassan, M. Imran, X. Cheng, C. Petridis, and C. Sule, in Social Media Strategy in Policing (Security Informatics and Law Enforcement), B. Akhgar, P. S. Bayeri, and G. Leventakis, Eds. Cham, Switzerland: Springer, 2019, pp. 177–195.

[8] Brantingham, M., Mohler, G. O., and P. J. "Does predictive policing result in biased arrests? A randomized controlled trial's findings, Statist. Public Policy, vol. 5, no. 1, pp. 1–6, January 2020.

[9] "A study on performance evaluation of machine learning algorithms for crime dataset," by A. Nasridinov and Y.-H. Park Dec. 2021, Adv. Sci. Technol. Lett., vol. 90, pp. 90–92

[10] "Forecasting crime with deep learning," by A. Stec and D. Klabjan [Online] 2022, arXiv:1806.01486. http://arxiv.org/abs/1806.01486 is accessible.